

# Linguistic dynamics in heritage speakers

Insights from the RUEG group

Edited by

Shanley E. M. Allen

Mareike Keller

Artemis Alexiadou

Heike Wiese

Current Issues in Bilingualism 4



## Current Issues in Bilingualism

Editors: Andrea C. Schalley (Karlstad University, Sweden) [editor-in-chief],  
M. Carmen Parafita Couto (Leiden University, Netherlands) [editor-in-chief],  
Susana Eisenchlas (Griffith University, Australia),  
Galina Putjata (Europa-Universität Flensburg, Germany),  
Jorge Valdés Kroff (University of Florida, USA).

In this series:

1. Shim, Ji Young. OV and VO variation in code-switching.
2. Rodríguez González, Eva & Rosita L. Rivera (eds.). Language assessment in multilingual settings: Innovative practices across formal and informal environments.
3. Polinsky, Maria & Michael T. Putnam (eds.). Formal approaches to complexity in heritage language grammars.
4. Allen, Shanley E. M., Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.). Linguistic dynamics in heritage speakers: Insights from the RUEG group.

ISSN (print): 2747-9919

ISSN (electronic): 2747-9927

# Linguistic dynamics in heritage speakers

Insights from the RUEG group

Edited by

Shanley E. M. Allen

Mareike Keller

Artemis Alexiadou

Heike Wiese



Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.).  
2025. *Linguistic dynamics in heritage speakers: Insights from the RUEG group*  
(Current Issues in Bilingualism 4). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/473>

© 2025, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

ISBN: 978-3-96110-505-2 (Digital)

978-3-98554-134-8 (Hardcover)

ISSN (print): 2747-9919

ISSN (electronic): 2747-9927

DOI: 10.5281/zenodo.15056099

Source code available from [www.github.com/langsci/473](http://www.github.com/langsci/473)

Errata: [paperhive.org/documents/remote?type=langsci&id=473](http://paperhive.org/documents/remote?type=langsci&id=473)

Cover and concept of design: Ulrike Harbort

Typesetting: Lea Coy, Hannah Schleupner, Felix Kopecky, Sebastian Nordhoff

Proofreading: Alexandra Fosså, Amir Ghorbanpour, Amy Amoakuh, Benjamin Brosig, Brett Reynolds, Cem Keskin, David Carrasco Coquillat, Eliane Lorenz, Elliott Pearl, Giorgia Troiani, Jeroen van de Weijer, Justin Richard Leung, Katja Politt, Klara Kim, Laurentia Schreiber, Ludger Paschen, Mary Ann Walter, Sebastian Nordhoff, Silvie Strauß, Tom Bossuyt

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: X<sub>E</sub>La<sub>T</sub>E<sub>X</sub>

Language Science Press

Scharnweberstraße 10

10247 Berlin, Germany

<http://langsci-press.org>

[support@langsci-press.org](mailto:support@langsci-press.org)

Storage and cataloguing done by FU Berlin



# Contents

|           |  |            |
|-----------|--|------------|
| <b>1</b>  | <b>Introduction: Investigating the dynamics of language contact situations</b>   |            |
|           | Heike Wiese, Shanley E. M. Allen, Mareike Keller & Artemis Alexiadou   | <b>1</b>   |
| <b>I</b>  | <b>Methodology and application</b>   |            |
| <b>2</b>  | <b>Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups</b>                                  |            |
|           | Heike Wiese, Annika Labrenz & Albrun Roy   | <b>33</b>  |
| <b>3</b>  | <b>Mapping the mappings and then containing them all: Quality assurance, interface modeling, and epistemology in complex corpus projects</b> |            |
|           | Anna Shadrova, Martin Klotz, Rahel Gajaneh Hartz & Anke Lüdeling   | <b>69</b>  |
| <b>4</b>  | <b>Family language dynamics: Strengthening heritage speakers' linguistic resources</b>   |            |
|           | Judith Purkarthofer, Rosemarie Tracy, Sofia Grigoriadou & Johanna Tausch   | <b>111</b> |
| <b>II</b> | <b>Dynamics of grammatical structure</b>   |            |
| <b>5</b>  | <b>Inter- and intra-individual variation: How it materializes in Heritage German and why it matters</b>                                      |            |
|           | Wintai Tsehaye, Rosemarie Tracy & Johanna Tausch   | <b>141</b> |
| <b>6</b>  | <b>Null subjects in heritage Greek, Russian and Turkish</b>  |            |
|           | Onur Özsoy, Vasiliki Rizou, Maria Martynova, Natalia Gagarina, Luka Szucsich & Artemis Alexiadou   | <b>179</b> |

## *Contents*

|  |   |     |
|--|---|-----|
| 7  | <b>Dynamics of verbal aspect in heritage Greek, Russian and Turkish</b><br>Vasiliki Rizou, Onur Özsoy, Maria Martynova, Luka Szucsich, Artemis Alexiadou & Natalia Gagarina                                       | 219 |
| 8  | <b>Clause combining in narrative discourse: A contrastive analysis across heritage and majority languages</b><br>Christoph Schroeder, Kateryna Iefremenko, Kalliopi Katsika, Annika Labrenz & Shanley E. M. Allen | 251 |
| 9  | <b>Balkan Turkic as a model for understanding contact-induced change in Turkish</b><br>Cem Keskin   | 287 |
| 10   | <b>Dynamic properties of the heritage speaker lexicon</b><br>Nadine Zürn, Mareike Keller, Rosemarie Tracy & Anke Lüdeling   | 325 |
| <br><b>III Dynamics of pragmatic structure</b> |   |     |
| 11   | <b>Information packaging and word order dynamics in language contact</b><br>Oliver Bunk, Shanley E. M. Allen, Sabine Zerbian, Tatiana Pashkova, Yulia Zuban & Erica Conti   | 379 |
| 12   | <b>Intonation in heritage languages</b><br>Sabine Zerbian, Yulia Zuban, Marlene Böttcher & Oliver Bunk  | 413 |
| 13   | <b>Majority English of heritage speakers</b><br>Tatiana Pashkova, Marlene Böttcher, Kalliopi Katsika, Sabine Zerbian & Shanley E. M. Allen  | 449 |
| 14   | <b>Dynamics of discourse markers in language contact</b><br>Annika Labrenz, Kateryna Iefremenko, Kalliopi Katsika, Shanley E.M. Allen, Christoph Schroeder & Heike Wiese  | 493 |
| 15   | <b>Discourse openings and closings across languages in contact</b><br>Kalliopi Katsika, Annika Labrenz, Kateryna Iefremenko & Shanley E.M. Allen  | 527 |
| <b>Index</b>                                   |   | 577 |

# Chapter 1

## Introduction: Investigating the dynamics of language contact situations

Heike Wiese<sup>a</sup>, Shanley E. M. Allen<sup>b</sup>, Mareike Keller<sup>c</sup> & Artemis Alexiadou<sup>d,a</sup>

<sup>a</sup>Humboldt-Universität zu Berlin <sup>b</sup>University of Kaiserslautern-Landau

<sup>c</sup>University of Mannheim <sup>d</sup>Leibniz-Centre General Linguistics

In this chapter, we introduce the volume *Linguistic dynamics in heritage speakers – the culmination of six years of coordinated work in the Research Unit Emerging Grammars in Language Contact Situations: A Comparative Approach*, funded by the German Research Foundation. The work reported in this volume explores productions of heritage speakers of German, Greek, Russian, and Turkish who have either English or German as a majority language as well as productions of monolingually-raised speakers of each of those languages. The research is based on a corpus of narrative data in communicative situations differing in formality (formal, informal) and mode (spoken, written). This chapter motivates the overall research program, presents relevant terminology, highlights the main findings across the studies reported in the volume, briefly introduces each chapter, and summarizes the contributions of our research to the field.

### 1 Introduction

#### 1.1 Heritage speakers in focus

Variation is a core characteristic of language use, and contact between different linguistic systems is always part of this. Accordingly, multilingualism is the normal condition for human language. Worldwide, most speakers today are multilingual (e.g., Grosjean 1982, 2010, Romaine 1989, Myers-Scotton 2006), and many



countries recognise a range of national languages (e.g., Switzerland has four, Namibia has 13, and India recognises 22 “regional languages”).

In contrast to this multilingual reality, it is often monolingualism that is treated as the norm. Such a monolingual bias is particularly common in societies in Europe and their former colonies such as the US or Australia (e.g., Gogolin 2002, Canagarajah 2013, Cook & Wei 2016, Ortega et al. 2016). This is a heritage of European nation-state building in the 19<sup>th</sup> century. At that time, an ideological link of “one country, one nation, one language” provided the basis for an imagined homogenous speech community as the basis of a nation.

Such a monolingual bias has also influenced linguistic research, in particular in the Global North. For instance, in second language acquisition, learners have traditionally been judged against monolingual language use, and structural approaches have, for a long time, focused on monolingual speakers as the primary source for native grammars (cf. criticism in Brutt-Griffler & Samimy 2001, Bonfiglio 2010, Cook 2016, Ortega et al. 2016). However, to fathom the reality of human languages in use, we should not restrict ourselves to the special case of monolinguals, but instead include multilingual speakers, who represent the normal case of linguistic competence (Wiese et al. 2022).

One example of speakers who routinely use two or more languages in their daily lives is represented by the case of heritage speakers, that is, speakers who grew up in a bi- or multilingual home with a minority language that is part of their family’s linguistic heritage, typically as a result of immigration in an earlier generation. As minority languages, such heritage languages are spoken in the family, but they are not the language that is dominant in the larger society at the national or regional level. The dominant languages are typically part of speakers’ repertoires as well. In monolingually-biased societies, heritage speakers hence bring back some multilingual normalcy: they do not restrict themselves to one societally-recognised language, but make use of resources from different languages, and they do so from an early age.

The heritage language is typically first acquired through exposure to oral language and implicit learning and starts as a typically developing and dominant first language (L1), but often loses its dominant status after early childhood. The societal majority language is acquired as an additional L1 or an early L2; it typically starts as a less-used language, but often becomes dominant when formal education begins (kindergarten, school), which is also the time when formal registers and written language are systematically established (Rothman 2007). This interplay of multilingualism and language development challenges traditional L1/L2 distinctions and conceptualisations of native speakers based on con-

## *1 Introduction: Investigating the dynamics of language contact situations*

structions of monolingual speech communities (Rothman & Treffers-Daller 2014, Wiese et al. 2022).

Heritage speakers and their communities thus offer a fascinating locus for research. The fact that both heritage and majority languages are used in daily life makes language contact ubiquitous. The effects of this constant contact are evident not only at levels of individual linguistic repertoires, but also at the larger level of the speech community, bringing about and spreading new patterns of language structure and use in a setting that is particularly open to linguistic variation and change.

In the public discussion of such countries as Germany or the US, the prevailing monolingual bias leads to a view of heritage speakers as a deviant case requiring special language support and intervention if speakers do not adhere to a perceived monolingual norm. This bias has also been visible in linguistic approaches where heritage speakers' competence has often been judged against monolingual standards, rather than assessed in its own right (cf. criticism in Rothman et al. 2023). For instance, "attainment" in the heritage language is typically measured against a monolingual standard language, with deviations classified as "errors", and linguistic areas characterised as "vulnerable" if they are likely to support such deviations, and "resilient" if they are not. Changes in a language are often measured with different yardsticks, depending on whether the language is being used as a heritage language or as a majority language. If we observe changes in heritage speakers, but not in majority language speakers, this is typically characterised as a "vulnerability" of the heritage language. On the other hand, if we observe changes in majority language speakers, but not in heritage speakers, heritage speakers tend to be characterised as "conservative" (e.g., Montrul 2015). In recent years, however, research has moved beyond supposed incomplete acquisition, attrition, and loss compared to monolingual norms, towards studies homing in on internal grammatical systematicity (e.g., Grosjean 2008, Pascual y Cabo & Rothman 2012, Rothman & Treffers-Daller 2014, Guijarro-Fuentes & Schmitz 2015, Scontras et al. 2015, Schroeder 2016, Tsehayé et al. 2021, Zuban et al. 2021, Wiese et al. 2022, Alexiadou & Rizou 2023, Alexiadou et al. 2023, Özsoy & Blum 2023, Bunk et al. forthcoming, 2024, Keskin et al. 2024; see Kupisch & Rothman 2018 for a detailed critique of accounts of what has been labelled "incomplete acquisition"). The following section illustrates the contribution of the Research Unit *Emerging Grammars in Language Contact Situations* (RUEG; see <https://hu.berlin/rueg>) to reframing heritage speakers and their languages as systematic developments within the spectrum of native grammars.

## 1.2 The Research Unit *Emerging Grammars in Language Contact Situations* (RUEG)

RUEG is a cluster of projects that have collaboratively investigated the dynamics of language contact in heritage speakers. Our research has been driven by a multilingual perspective: we approached heritage speakers' languages in terms of dynamics, rather than vulnerability; of development, rather than incomplete acquisition; and of innovation, rather than attrition and loss. As is also emphasised in other recent approaches taking the same perspective, we regard heritage grammars as reflecting a type of variation within native grammars (Scontras et al. 2018, Embick et al. 2020, Flores & Rinke 2020, in reply to Polinsky & Scontras 2020) as well as principles of language variation and change (Johannessen 2018, Aalberse et al. 2019, Lohndal et al. 2019, Muysken 2020).

A central target has been to investigate noncanonical phenomena in heritage speakers' repertoires as patterns in their own right in order to understand their status in the linguistic system and in language use. In a unified approach, projects examined the hypothesis that what is often regarded as speakers' limited proficiency in both the heritage and the majority languages can be re-interpreted as the emergence of new grammatical options and structures and, possibly, new linguistic varieties. While contact-induced variation may at first reflect an individual phenomenon, variants may also stabilise, spread and become part of shared linguistic repertoires and practices that support the emergence of new varieties.

Our collaborative research was guided by two overarching questions over two research periods (RUEG1: 2018–2021; RUEG2: 2021–2024):

1. What is the status of noncanonical phenomena in heritage speakers' two languages from the perspective of emerging grammars?
2. What are the linguistic dynamics in heritage speakers' repertoires?

In pursuit of the first research question, we investigated linguistic systematicity at different grammatical and pragmatic levels. Casting our net wide, we targeted noncanonical phenomena in general, that is, all those that differ from codified standard language norms. This allowed us to broadly take into account such phenomena in heritage speakers as well as in monolinguals, and to then assess which ones are characteristic for language contact situations and heritage speakers' repertoires. To capture a wider portion of speakers' repertoires, we included language use in different communicative situations: formal and informal, spoken and written. Building on this, the second research question targeted

## *1 Introduction: Investigating the dynamics of language contact situations*

dynamics in repertoires. We looked at crosslinguistic interactions between heritage speakers' two languages as well as developments that derive from existing language-internal tendencies, and at the dynamics of register in language contact.

The group encompassed altogether 16 research projects and a central project that was responsible for the overall coordination:

- P1 *Nominal morphosyntax and word order in Heritage Greek across majority languages*  
PI: Artemis Alexiadou  
Research Associate: Vasiliki Rizou
- P2 *Morphosyntax and word order in majority English across heritage speakers*  
PI: Shanley E. M. Allen  
Research Associate: Tatiana Pashkova
- P3 *Nominal morphosyntax and word order in heritage Russian across majority languages*  
PIs: Natalia Gagarina, Luka Szucsich  
Research Associate: Maria Martynova
- P4 *Clause combining and word order in heritage Turkish across majority languages*  
PI: Christoph Schroeder  
Research Associate: Kateryna Iefremenko
- P4a *Head directionality change in Turkic in contact situations: A diachronic comparison between heritage Turkish and Balkan Turkic*  
PI: Christoph Schroeder  
Research Associate: Cem Keskin
- P5 *Clause structure in heritage German*  
PI: Rosemarie Tracy  
Research Associate: Wintai Tsehayé
- P6 *Noncanonical constituent linearisation in German across heritage speakers*  
PI: Heike Wiese  
Research Associate: Oliver Bunk
- P7 *Intonation and word order in majority English and heritage Russian across speaker populations*

PI: Sabine Zerbian  
Research Associate: Yulia Zuban

- P8 *Dynamics of information structure in language contact*  
PIs: Shanley E. M. Allen, Oliver Bunk, Sabine Zerbian  
Research Associates: Kristina Barabashova, Tatiana Pashkova, Yulia Zuban
- P9 *Dynamics of discourse organisation in language contact*  
PIs: Shanley E. M. Allen, Christoph Schroeder, Heike Wiese  
Research Associates: Kateryna Iefremenko, Kalliopi Katsika, Annika Labrenz
- P9a *Clause combining in Balkan Turkic: Pathways and stages of contact-induced grammaticalization*  
PI: Christoph Schroeder  
Research Associate: Cem Keskin
- P10 *Dynamics of verbal aspect and pronominal reference in language contact*  
PIs: Artemis Alexiadou, Natalia Gagarina, Luka Szucsich  
Research Associates: Maria Martynova, Onur Özsoy, Vasiliki Rizou
- P11 *The heritage speaker lexicon: Dynamics and interfaces*  
PIs: Mareike Keller, Anke Lüdeling, Rosemarie Tracy  
Research Associate: Nadine Zürn
- Pc *Corpus linguistic methods*  
PIs: Anke Lüdeling, Anna Shadrova  
Research Associate: Martin Klotz
- Pd *“Emerging grammars”: A cross-linguistic corpus of comparative data in heritage and majority language use*  
PIs: Anke Lüdeling, Heike Wiese  
Research Associates: Martin Klotz, Annika Labrenz, Maria Pohle
- Pt *Family language dynamics: Empowering speakers of majority and heritage languages*  
PIs: Judith Purkarthofer, Rosemarie Tracy  
Research Associates: Sofia Grigoriadou, Johanna Tausch
- Pz *Coordination*  
PI: Heike Wiese  
Associates: Irem Duman Çakir, Esther Jahns, Pia Linscheid, Katrin Neuhaus

In addition to the researchers in individual projects, four Mercator Fellows cooperated with all projects: Maria Polinsky and Shana Poplack (Mercators in RUEG1); Cristina Flores and Jeanine Treffers-Daller (Mercators in RUEG2).

The remainder of the chapter is structured as follows. In Section 1.3, we provide some definitions for central concepts used in our group. Based on this, we describe our shared methodology and collaborative data collection (Section 2) and summarise central findings of our research (Section 3). We then offer an overview of the contributions to this volume (Section 4), and conclude with a reflection on RUEG's contributions and directions for future research (Section 5).

### **1.3 Central concepts and terminology**

We call **NONCANONICAL** all phenomena that differ from codified standard language norms. Accordingly, phenomena falling within codified standard norms are **CANONICAL**.

We understand a **HERITAGE LANGUAGE** as a language spoken in a family, typically as a result of migration, in cases where that language is not dominant in the larger society. The counterpart to a heritage language is a **MAJORITY LANGUAGE**: the dominant language of the larger society. Note that the larger society need not necessarily be the society at a national level, but can also be the speech community in a specific region. For instance, in a Kurdish-dominant region of Turkey, the majority language might be Kurdish.

A **HERITAGE SPEAKER** is a speaker who combines heritage and majority languages in their repertoire – a speaker who grew up in a bi- or multilingual home with a minority/heritage language that is not the majority language dominant in the larger society.

The term **MONOLINGUAL SPEAKER** should be understood as a short form of “monolingually-raised speaker”: this is a speaker who grew up without an additional minority/heritage language in the family.

Note that both heritage speakers and monolingual speakers will have the majority language as part of their repertoire. Accordingly, a **MAJORITY-LANGUAGE SPEAKER** is to be understood as a speaker who speaks a majority language, and this includes both monolingual and bi-/multilingual speakers.

Hence, if we look at two languages, say, English and German, and two countries, say, the US and Germany, and compare majority and heritage language use for monolingual and bilingual speakers, then this is not a one-to-one correlation. For instance, in the US, we could compare majority-language use in bilingual vs. monolingual speakers: this would be English spoken as a majority language (a) by monolingual speakers who grew up with only English regularly spoken in

the family, and (b) by bilinguals who speak English as a majority language, and who are, at the same time, heritage speakers of another language, in this case German. On the other hand, we could compare German in the US and Germany. In the US, German is spoken by heritage speakers, that is, bilingual speakers who use German as a heritage language in addition to English (the majority language), whereas in Germany, German is spoken as a majority language, that is, by majority-language speakers, who might be monolingual (only German in the family) or bilingual (heritage speakers of other languages, for instance, Turkish).

If we look at speakers' repertoires, we find a differentiation of language use across different COMMUNICATIVE SITUATIONS. This term refers to the setting of a communicative event. In this volume, it will be used as a superordinate term for the four conditions in which we elicited our data: informal-spoken, informal-written, formal-spoken, and formal-written (see Section 1.2; see Wiese 2020, 2023, for a detailed discussion of communicative situations). Communicative situations can be distinguished through two more specific terms for the contrasts we employed: MODE refers to the contrast of spoken vs. written communicative situations, FORMALITY to that of formal vs. informal ones.

Against this background, we understand REGISTER as the linguistic counterpart of communicative situation, that is, the language use associated with different communicative situations. Specifically, register distinctions are socially motivated and recurring intra-individual linguistic variations in different communicative situations (cf. Lüdeling et al. 2022).

For instance, if speakers use different linguistic patterns in informal vs. formal communicative situations, then they have different registers for this kind of contrast. If they do not differentiate between their linguistic options in these kinds of situations, they do not make a register distinction for informal vs. formal communicative situations for this kind of contrast. Such distinctions can also be specific to linguistic domains: if speakers, say, use different terms of address in informal and formal communicative situations, but they use noncanonical gender marking in both situations, then they make a register distinction with respect to address terms for informal vs. formal communicative situations, but not with respect to nominal gender.

The next section describes the kind of comparisons we made and the contrasts we targeted in more detail.

## **2 RUEG's comparative approach**

Our collective research supported a large-scale comparative investigation based on a unified methodology and a shared empirical basis. We included five lan-

## *1 Introduction: Investigating the dynamics of language contact situations*

guages: English, German, Greek, Russian, and Turkish. We investigated bilinguals in the US and Germany (heritage speakers of Greek, Russian, and Turkish in both countries, plus heritage speakers of German in the US), and monolinguals of all five languages in Germany, Greece, Russia, Turkey, and the US.

From a perspective of emerging grammars, analyses took into account the possibility that heritage speakers develop new grammatical options in both their languages in different communicative situations. Crucially, this meant that it was not sufficient to focus on standard language alone, but rather that we needed to investigate the larger repertoires of speakers, and their different choices across communicative situations. Accordingly, our approach included

- (a) heritage and majority language use;
- (b) language use in formal as well as informal situations;
- (c) written and spoken language;
- (d) language use in two different age groups: adolescents and adults.

Taking into account both languages of heritage speakers enabled us to identify domain-specific skills in different languages that might complement each other (cf. Grosjean 1997 on the “complementarity principle” for bilinguals, and Blommaert et al. 2005 for a discussion of “truncated multilingualism” for linguistic resources in different spatial contexts). It also allowed us to investigate mutual influence between the majority language and the heritage language, including the mutual reinforcement of linguistic options from different registers for majority and heritage languages.

The inclusion of informal as well as formal communicative situations, and of spoken as well as written communicative situations, further supported a genuinely bilingual perspective and a holistic view of competence. For instance, a frequent finding reported from heritage language research has been the loss of abilities in a heritage language after majority language immersion at school. However, such findings might be coloured by a sole focus on formal registers (e.g., Merino 1983; cf. also Kupisch & Rothman 2018 for a critique). By including language use in informal (spoken and written) communicative situations, we were able to capture possible register specialisations and interactions. This allowed us to assess speakers’ proficiency by register, rather than solely by comparison to formal standard language. Furthermore, it enabled us to identify the transfer of patterns across communicative situations as well as restrictions to certain registers.

For an appropriate assessment of noncanonical patterns, one needs to take into account the breadth of language use not only for heritage speakers, but also for monolinguals. Accordingly, we compared the repertoires of heritage speakers with language use of monolingual speakers.

If heritage language use is compared only to the respective standard language or to monolingual language from formal settings, noncanonical patterns might be attributed to the heritage context even if they represent linguistic variation that generally occurs in informal-spoken situations (see Wiese et al. 2025 [this volume]). Factoring this in, we incorporated the same breadth of language use in monolinguals as in heritage speakers. Our comparisons thus targeted

- (a) monolingual majority language speakers in countries where heritage speakers live, and
- (b) monolingual majority language speakers in the heritage countries (countries of origin for speakers' ancestors),

across the same written and spoken, formal and informal communicative situations.

Such register-sensitive, matched comparisons with monolingual language use allowed us to distinguish genuine language-contact phenomena from general patterns of variation. At the same time, the observed linguistic structures need not be an outcome of contact-induced change in the sense of (lexical or structural) transfer, but could also be due to language-internal change triggered or at least accelerated in contexts of language contact (see Silva-Corvalán 1994, Wiese 2013, Kupisch & Polinsky 2022). This possibility made the inclusion of different language pairs essential in order to identify specific language-contact effects. Consequently, our collaborative research allowed us to integrate results on

- (a) speakers of the same heritage language in countries with different majority languages,
- (b) speakers of different heritage languages for the same majority language,
- (c) speakers of the same language as heritage vs. majority language.

To support systematic, large-scale comparisons, we collected data in five countries in a closely synchronised endeavour, using the same methodology for eliciting naturalistic productions (see Wiese et al. 2025 [this volume] for methodological details). All data was processed and integrated into an annotated corpus

that served as a joint empirical basis for all projects and has, from its first version, been completely open access (<https://hu.berlin/rueg-corpus>; see Shadrova et al. 2025 [this volume] and Klotz et al. 2024 for corpus linguistic details). The cross-linguistic RUEG corpus is, to our knowledge, the first of its kind: it provides parallel data across countries, languages, contact-linguistic settings, and age groups, covering different communicative situations for heritage speakers and monolingual counterparts alike.

### **3 Major results and contributions from RUEG1 and RUEG2**

#### **3.1 Corpus**

A first major outcome of RUEG is the just-mentioned corpus – an annotated, multilayer and multimodal open-access corpus of comparable cross-linguistic production data from adolescent and adult heritage, majority and monolingually-raised speakers in informal and formal settings and written and spoken modes (Lüdeling et al. 2024; see Shadrova et al. 2025 [this volume] for more details). The data in the final version of this corpus come from 736 speakers: 381 adolescents and 355 adults; 411 bilingual and 325 monolingual. These speakers together contributed 528,709 normalised tokens to the corpus (typically graphemic words – 150,999 for English, 165,396 for German, 69,450 for Greek, 76,930 for Russian, and 65,934 for Turkish). An additional subcorpus of Kurdish speakers in Turkey ( $n = 29$ , 15169 tokens) and Germany ( $n = 9$ , 17710 tokens) provides Kurdish and Turkish language productions. In addition to the basic transcription, the entire corpus is annotated for language, lemma, and part of speech. Syntactic spans according to Universal Dependencies were added automatically. To pursue our specific research questions, manual annotations for topological Fields for German and Surface Syntactic Dependencies for German and Russian were added. Further, subsets of the data were annotated for specific phenomena such as intonation, nominal morphology, verb morphology, aspect marking, functions of discourse markers, referent introduction, and more.<sup>1</sup>

The corpus supports numerous types of systematic comparisons – across languages, countries, age groups, contact-linguistic settings, and communicative sit-

---

<sup>1</sup>The corpus is freely available at <https://zenodo.org/records/3236068> and is searchable with ANNIS (Krause & Zeldes 2014). At the time of publication of this volume, an ANNIS instance at Humboldt University of Berlin containing the RUEG corpora is hosted at <https://korpling.german.hu-berlin.de/annis/>. Full documentation for the corpus can be consulted at <https://korpling.german.hu-berlin.de/rueg-docs/latest/>.

uations (see Wiese et al. 2025 [this volume]), as well as detailed analyses within a heritage language and of individual speaker profiles (e.g. Tsehayé et al. 2025 [this volume], Zürn et al. 2025 [this volume]). Many of our studies compare the use of a heritage language in a majority setting and two different minority settings (Germany and the US – e.g., null subjects in Özsoy et al. 2025 [this volume], aspect in Rizou et al. 2025 [this volume], clause combining in Schroeder et al. 2025 [this volume], word order in Bunk et al. 2025 [this volume], discourse markers in Labrenz et al. 2025 [this volume]), while others compare the use of a majority language by three or four groups of bilingual speakers (clause combining in Schroeder et al. 2025 [this volume] for English and German, several structures in Bunk et al. 2025 [this volume] for German and Bunk et al. 2025 [this volume], Pashkova et al. 2025 [this volume] for English). Taken together, these comparisons allow for a more nuanced understanding of the factors underlying dynamic uses of language than is typically the case in other studies. Another strength of the corpus is that all of the studies on different areas of language (prosody, morphology, syntax, lexicon, pragmatics) come from the same speakers and the same data set, thus providing a more comprehensive picture of heritage language use than in previous research.

Many of the difficulties and rewards of the lived experience of our collaborative data collection, annotation and publication in an open-access corpus are detailed in Shadrova et al. (2025 [this volume]). From this experience, and in addition to the corpus itself, the RUEG project has several outcomes that are relevant for the field of corpus linguistics – especially for (distributed) interdisciplinary projects involving a variety of researchers and linguistic phenomena that are annotated manually in a multi-layer corpus. In particular, RUEG provides a model for how annotation, modeling, analysis, and the epistemological embedding of categories and analytical processes can proceed hand in hand. Our experience brings to light the importance of understanding that data acquisition, transcription and annotation are not trivial aspects of corpus work. Rather, all of the decisions involved in these steps are based on – sometimes implicit – assumptions and models which should be laid out explicitly to allow for scholarly sound use of the corpus data.

### **3.2 Factors driving noncanonical patterns**

A second major outcome of RUEG is the identification of numerous noncanonical patterns across the data, and insight into the various factors driving these patterns. We organised our corpus-linguistic studies within several joint ventures that involved all projects. These joint ventures targeted systematic developments,

their association with internal vs. external interfaces, their association with communicative situations (spoken vs. written, formal vs. informal), and the distinction of contact-induced change vs. language-internal developments and variation. Our corpus-linguistic and experimental studies revealed systematic non-canonical patterns in the domains of word order (Bunk et al. 2025 [this volume], Pashkova et al. 2025 [this volume]), sentence structure (Tsehayé et al. 2025 [this volume], Schroeder et al. 2025 [this volume], Keskin 2025 [this volume], Bunk et al. 2025 [this volume], Pashkova et al. 2025 [this volume]), morphosyntax (Özsoy et al. 2025 [this volume], Rizou et al. 2025 [this volume]), intonation (Zerbian et al. 2025 [this volume]), lexicon (Zürn et al. 2025 [this volume]), discourse organisation (Labrenz et al. 2025 [this volume], Katsika et al. 2025 [this volume]), and information structure (Bunk et al. 2025 [this volume], Zerbian et al. 2025 [this volume], Pashkova et al. 2025 [this volume]).

One key question was whether the noncanonical patterns we identified arose specifically due to language contact or if they emerged as part of the process of language-internal development and variation. We found clear indications for both in different situations. Numerous instances of cross-linguistic influence were evident – predominantly from the majority language to the heritage language (Tsehayé et al. 2025 [this volume], Rizou et al. 2025 [this volume], Zürn et al. 2025 [this volume], Labrenz et al. 2025 [this volume]) but also from the heritage language to the majority language (Pashkova et al. 2025 [this volume]). Other findings indicate change that bears no obvious relation to influence from another language but rather represents language-internal development since we saw similar noncanonical patterns in both bilingual and monolingual speakers of the language in question, or in bilingual speakers in contact with typologically different languages (Özsoy et al. 2025 [this volume], Keskin 2025 [this volume], Labrenz et al. 2025 [this volume]).

A second key question was whether noncanonical patterns would emerge mostly in structures at external interfaces incorporating knowledge from core grammar (e.g., phonology, morphology, syntax) and extragrammatical domains (e.g., discourse, pragmatics) (see Interface Hypothesis; Sorace 2011), or whether they would also emerge in structures that only required knowledge from core areas of grammar. We found a range of noncanonical patterns at external interfaces, for example at the interfaces of information structure with intonation (Zerbian et al. 2025 [this volume]), with word order in syntax (Tsehayé et al. 2025 [this volume], Bunk et al. 2025 [this volume], Pashkova et al. 2025 [this volume]), or with referent form in morphology (Özsoy et al. 2025 [this volume], Pashkova et al. 2025 [this volume]). However, we also found noncanonical patterns in areas

where arguably only core areas of the grammar were in play, such as with grammatical aspect or particle verbs (interface between morphology and semantics; Rizou et al. 2025 [this volume], Zürn et al. 2025 [this volume]).

A third key question addressed the role of communicative situations for the development of noncanonical patterns. Here we found more evidence of dynamic patterns in informal than formal situations, and in spoken than written situations. We also found evidence of register levelling in heritage speakers, such that patterns restricted to informal registers in monolingual speakers were also found in formal registers in heritage speakers, likely due to heritage speakers' lack of experience with formal registers (Tsehayé et al. 2025 [this volume], Schroeder et al. 2025 [this volume], Zürn et al. 2025 [this volume], Zerbian et al. 2025 [this volume], Katsika et al. 2025 [this volume]). These findings underline the importance of considering register in future studies of heritage speakers and other bilinguals, since most research elicits data in contexts that are more formal than informal. Our findings suggest that this methodological practice may be hampering our insights into one of the most fertile grounds of language change, namely informal language.

Finally, we evaluated the role of the social context in which a heritage language is used. In particular, we saw an effect of the size and cohesion of the speech community in several studies. For instance, the Turkish community in Berlin, where we collected our German data, is large and cohesive, whereas the Turkish community in the New York area, where we collected the US data, is smaller and more spread out.

Interestingly, we found more diverse patterns in clause combining (Iefremenko et al. 2021) and more register levelling (Schroeder et al. 2025 [this volume], Katsika et al. 2025 [this volume]) in the heritage speakers in the US. As another example, the German heritage speakers we studied in the US were all from recently immigrated families rather than the long-standing language islands that have existed in the US for generations. Our speakers were not part of a cohesive community, and we found considerably more inter-individual variation in clausal structure (Tsehayé et al. 2025 [this volume]) and the lexicon (Zürn et al. 2025 [this volume]) than has been found for heritage speaker groups living in more cohesive communities.

### **3.3 Resource-oriented perspective**

Arguably the most important result of RUEG is its support and furtherance of a resource-oriented rather than deficit-oriented view of heritage speakers. Wiese et al. (2022) offers a summary of our main results in this regard. First, we found

noncanonical patterns not only in bilingual speakers but also in monolingual speakers, including patterns that have so far been considered absent from native grammars, across linguistic domains (morphology, syntax, intonation, and pragmatics). Second, we found a degree of lexical and morphosyntactic inter-speaker variability in monolinguals that was sometimes higher than that of bilinguals. This finding challenges the model of the streamlined native speaker. Third, we observed that noncanonical patterns were dominant in spoken and/or informal communicative situations, and this was true for monolinguals and bilinguals. In some cases, bilingual speakers were leading quantitatively, but there was no qualitative difference in the noncanonical productions. In heritage settings where the language was not part of formal schooling, we found tendencies of register leveling such that patterns associated with informal situations in monolinguals were also used in formal communicative situations by heritage speakers.

Our findings thus indicate possible quantitative differences and different register distributions rather than distinct grammatical patterns in bilingual and monolingual speakers. Indeed, the linguistic systematicity that became clear here showed that, contrary to previous work, the language use of heritage speakers is not primarily characterised by errors, erosion and incomplete acquisition, but provides insights into options of developments and variation within the spectrum of a language. This further supports the integration of heritage speakers into the native-speaker continuum (e.g., Tsehayé et al. 2021, Wiese et al. 2022).

### **3.4 Dissemination**

Through conference contributions, workshops, invited talks and publications (see references throughout the following chapters), we were able to disseminate our research findings across sub-disciplines and theoretical schools. We have provided new findings on heritage speakers and on grammatical and pragmatic patterns in a range of languages, and new impulses for the study and conceptualisation of multilinguals and their linguistic resources.

We also were engaged in extensive outreach activities to transfer the knowledge gained through RUEG (e.g., through exhibitions, lectures, publications, workshops, web portals, and social media posts; see Purkarthofer et al. 2025 [this volume] for details). Our primary audience for these efforts has been educators of heritage language children, mainly through training workshops and collaborations with schools, as well as parents of heritage language children through collaboration with adult education programs. Last but not least, a number of outreach activities resulted in successful engagement and dialog with the general public.

Our experience in these various activities points to the importance of an approach that takes speakers' agency seriously and shows awareness of their multilingual environments (e.g., Bunk *in press*, 2024, Tausch & Tsehay 2023). As a result, we highlight the strategy of providing interactive activities that can be engaged with on multiple levels – by multilingual individuals themselves, by parents, by educators and others – including hands-on tasks, discussion prompts, quizzes, and information charts all based on the same materials developed from research results. We further emphasize the advantages of having the same resources and activities available in multiple languages to most effectively reach multilingual audiences (e.g. Purkarthofer et al. 2023). We also underline the value of communicating easily accessible yet research-grounded knowledge about bi- and multilingual language development, including the use of citizen science research activities.

## 4 Overview of individual projects and chapters

Part I of this volume (Chapters 2 to 4) introduces the methodology and application concerning corpus-based studies of heritage languages in the RUEG project. Part II (Chapters 5 to 9) focuses on the dynamics of morphosyntactic structure under a multilingual perspective, with studies of individual heritage languages and majority languages. It is complemented by Chapter 10 on the dynamics of the heritage speaker lexicon. Part III (Chapters 11 to 15) revolves around pragmatic aspects of heritage and majority varieties, including information structure, intonation, discourse markers and discourse openings and closings.

In Chapter 2, Wiese, Labrenz and Roy introduce the volume with their contribution *Tapping into speaker's repertoires*. They provide the conceptual-methodological background to RUEG's collaborative enterprise, arguing for comparative studies that take into account speaker repertoires across groups, including language from formal and informal communicative situations in heritage speakers and majority speakers, bilinguals and monolinguals alike. They present the Language Situations method used to elicit the data in all RUEG projects, as well as two studies that were conducted to evaluate the method. The authors argue that the set-up was successful in eliciting language productions which reflect register distinctions in different communicative situations, and that this data is naturalistic in the sense that it aligns with spontaneous data.

In Chapter 3, Shadrova, Klotz, Hartz and Lüdeling introduce the RUEG corpus and its infrastructure, providing insights into the methodological considerations concerning modeling, compilation and analysis of a structured and deeply annotated corpus. As is aptly foreshadowed in the title, *Mapping the mappings and*

*then containing them all*, instead of providing the reader with a manual of how to construct and maintain a complex linguistic corpus, the authors shine a light on a range of theoretical and practical concerns such as how to keep the corpus intact and sound within reasonable production cycles requiring the collaboration of dozens of researchers, and discerning which types of models we face in a large scale corpus project on heritage languages. Questions like these are seldom addressed in relation to large-scale corpus projects, even though they are fundamental to the quality, success and sustainability of open resources.

In Chapter 4, Purkarthofer, Tracy, Grigoriadou and Tausch address *Family language dynamics* by discussing the societal status of heritage languages, and by offering best practices for providing research-based knowledge about heritage language dynamics and repertoires to educators and bilinguals. As heritage languages are still often regarded as an obstacle rather than an asset, the authors target myths and misconceptions about heritage speakers. They also present different kinds of transfer activities revolving around three key aspects, namely communication, consultation and cooperative action, to emphasise the value of multilingualism from an early age.

In Chapter 5, Tsehayé, Tracy and Tausch target *Inter- and intra-individual variation* in German clause structure to show how a heritage language develops outside of historically established language islands in what they term "tiny language islands". The authors present results on clause-type optionality, finite subordinate clause distribution and clausal peripheries, concluding that in heritage speakers clausal architecture presents as a stable feature, and that variation is found mostly in the morphological domain. In their discussion of inter-individual variation they emphasise that generalisations over groups can be misleading given that, in many cases, non-canonical but nevertheless systematic patterns can clearly be traced to individual speakers and do not characterise the whole group.

In Chapter 6, Özsoy, Rizou, Martynova, Gagarina, Szucsich and Alexiadou discuss *Null subjects in heritage Greek, Russian and Turkish*, focusing on modality and animacy as influencing factors. They review results from corpus-based and experimental studies exploring whether heritage varieties, compared to monolingual data, pattern alike regarding the frequency of overtly realised pronominal subjects, especially in contact with Germanic majority languages. Their results indicate that animacy is a common factor affecting subject realization across groups. Further, quantitative differences between speaker groups and also between different communicative situations emerge depending on the various language combinations in the study.

In Chapter 7, Rizou, Martynova, Özsoy, Szucsich, Alexiadou and Gagarina discuss *Dynamics of verbal aspect in heritage Greek, Russian and Turkish*. The authors show that intra-linguistic factors like markedness and extra-linguistic factors like formality and mode variation affect speakers' preference for aspectual forms in heritage languages, although not to the same degree. The results from off-line tasks suggest that heritage speakers' repertoires vary depending on the languages in contact, motivating further discussions on aspect based on multi-lingual comparisons.

In Chapter 8, Schroeder, Iefremenko, Katsika, Labrenz and Allen take a comparative approach to *Clause combining in narrative discourse*, exploring subordination strategies in relation to formality and mode in the adolescents across all the RUEG speaker groups. The data show that typological characteristics of clause combining are largely preserved in the contact settings, despite speakers' bilingualism. Nevertheless, register levelling is observed in contact settings, less so in Germany and more so in the US where the heritage language has less community support. The authors advocate that future investigations should pursue individual variation in more detail to gain further insights into sociolinguistic variables explaining the observed effects.

In Chapter 9, Keskin takes *Balkan Turkic as a model for understanding contact-induced change in Turkish*, focusing on changes in head directionality and subordination. Considering the dialectal variation in Balkan Turkic, the author lays out trajectories of syntactic change in Turkic that can be used as a model to predict the changes in heritage Turkish. As regards head directionality change, he proposes that the observed heterogeneity across and within varieties can be traced back to intensity of contact, rates of change in different subsystems and universal constraints on grammar. For the domain of subordination Keskin proposes that in situations of prolonged contact between the the Turkic varieties under study, two subordination templates are blended into one.

In Chapter 10, Zürn, Keller, Tracy and Lüdeling discuss *Dynamic properties of the heritage speaker lexicon*, focusing on verbs in heritage German in an English majority context. The contribution begins with an analysis of the lexical inventory as a whole and then moves on to details concerning particle verbs with respect to their structural and semantic properties. They also include production phenomena like self-correction, repetition and and hesitation in the immediate environment of particle verbs to elucidate the process of lexical access and the selection of words from a pool of available options. The results challenge claims concerning the limited vocabulary of heritage speakers by showing the creative ways in which heritage speakers employ morphological building blocks to convey subtle nuances in meaning.

In Chapter 11, Bunk, Allen, Zerbian, Pashkova, Zuban and Conti illustrate the relationship between *Information packaging and word order dynamics in language contact*. The chapter reviews several studies discussing referent introduction in English as well as verb-third word order and modal particles in German, and presents a cross-linguistic study on left-dislocation constructions in English, German, and Russian. The authors propose that the greater flexibility exhibited by heritage speakers concerning their word order choices might be due to several factors, depending on their specific linguistic biographies. These factors include the constant exposure of heritage speakers to the various linguistic structures of their majority and heritage language, the different quantity and quality of input in the heritage languages, and an increased sensitivity to communicative situations.

In Chapter 12, Zerbian, Zuban, Böttcher and Bunk present their research on *Intonation in heritage languages*, a topic that has received less attention in heritage language research than, for example, morphosyntactic issues. The authors investigate different intonational features (phrasing and frequency of pitch accents) as they are used across oral narrations as well as at specific interfaces such as questions, focus and discourse linking. In accordance with other findings reported in this volume, they conclude that, overall, heritage language utterances show quantitative rather than qualitative differences, which could be attributed to register levelling and/or to maintenance of language features.

While all other chapters focus on heritage languages, Chapter 13 by Pashkova, Böttcher, Katsika, Zuban, Zerbian and Allen takes a closer look at *Majority English of heritage speakers*. The authors review studies relevant to the interface between information structure and four linguistic domains (prosody, article use, reference, and clausal syntax), as well as discourse structure. Their findings provide evidence that heritage speakers and monolingual speakers do not differ significantly in their use of the majority language in most of these areas. They suggest that the few differences found between these speaker groups in majority language use could be attributed either to crosslinguistic influence, or to a desire for explicitness and strict differentiation between their two languages.

In Chapter 14, Labrenz, Iefremenko, Katsika, Allen, Schroeder and Wiese present studies on *Dynamics of discourse markers in language contact* from English, German, Turkish, and Greek. While previous literature has mostly focused on the borrowability of discourse markers, this contribution highlights functional convergence in bilingual settings. Furthermore, the target domain is expanded to graphic discourse markers by including the three-dot sign. The qualitative analyses address language-specific developments, majority-language influence, and potential register levelling, expressed by functional extensions or restrictions

of discourse markers in heritage languages. Overall, the findings indicate differences between speaker groups concerning the frequency with which specific functions of discourse markers are used.

In Chapter 15, Katsika, Labrenz, Iefremenko and Allen discuss *Discourse openings and closings across language in contact*. A detailed analysis of their textual, subjective and intersubjective functions in the light of intercultural communication indicates that openings and closings are crucial in discourse organisation across communicative situations. In particular, the high frequency of openings across speaker groups suggests that monolingual and heritage speakers alike use situation-specific openings. The results also show evidence of language-specific patterns of discourse organisation that are similar across monolingual and heritage speakers, with pockets of evidence for cross-linguistic influence, register leveling, and individual variation.

## 5 Conclusion

Out of the many contributions to the study of heritage speakers and their language repertoires made by RUEG, perhaps the most tangible one is our open-access corpus, covering comparable informal vs. formal and spoken vs. written productions by adolescent and adult bilingual speakers of heritage Greek, Russian, and Turkish in Germany and the US, and of heritage German in the US, in both of their languages, as well as matching data from monolinguals in Germany, the US, Greece, Russia, and Turkey. This empirical resource can be used by researchers world-wide to investigate a wide variety of linguistic phenomena from a comparative perspective.

Our large-scale cross-linguistic study has allowed us to approach bilingual and monolingual speakers as native speakers of both their languages and to tease apart the various sources of noncanonical patterns in bilingual productions: language-internal developments and variation, cross-linguistic influence, general effects of bilingualism, and register levelling. Moreover, our research has provided new insights on register variation in different communicative situations among monolingual speakers and unveiled previously unreported patterns that are worthy of further investigation inside and outside a heritage language context.

Taken together, our findings suggest many profitable future directions for research. One possible direction is to expand the populations under investigation. The Language Situations method used in RUEG would adapt well to a range of ages, including both children and older adults which we did not examine, potentially uncovering developmental trends in noncanonical patterns. In addition,

it would be useful to extend this method to cross-linguistic studies involving typologically more varied languages in contact than the ones included in our sample. This could help to discern a sharper distinction between the sources of noncanonical patterns in heritage productions. We also hope that research on the monolingual grammar will continue to examine the sources of variation we have identified in our corpus work. Moreover, some of our studies have shown that experimental research can be useful to complement and build on our corpus linguistic findings, by targeting details of the noncanonical patterns our elicitations uncovered in different groups and communicative situations. Further experimental research could tackle the subtle differences observed among our various groups with quantitative paradigms designed to explore these in more depth.

Another interesting direction would be to expand the contexts of this research to different societal settings. One factor noted in several of our studies is the impact of the cohesiveness of the speech community on the uniformity of non-canonical patterns. It would be interesting to explore this in a more targeted way, such as in a study that explicitly compares one heritage-majority language pair in cohesive vs. more spread-out communities. Another direction would be to investigate heritage speakers in the Global South where many countries do not have such strong majority languages. This would allow us to tease apart what in our findings pertains to heritage speakers in general and what is particular to bilingual speakers under specific conditions in countries with a monolingual habitus in the style of European nation states.

We are pleased to present this volume offering a detailed overview of the contributions of RUEG to the field of heritage language studies. We hope it inspires further development in the field in addition to contributing knowledge in its own right.

## Acknowledgements

We gratefully acknowledge funding of the Research Unit *Emerging Grammars in Language Contact Situations* (FOR 2537) from the Deutsche Forschungsgemeinschaft. This includes the following grant numbers: 394836232, 394837597, 394839191, 394841858, 394995401, 394838878, 394844953, 394844736, 394821310, 313607803. Without this funding, such a large and complex project would never have been possible.

We also gratefully acknowledge the 736 participants who graciously gave their time to be involved in our study, the many individuals and organisations that helped us to contact these participants, the many research assistants who col-

lected and transcribed and annotated and processed the data, the doctoral students and postdocs and principal investigators who led the individual studies reported here, the cooperation partners and Mercator Fellows involved in RUEG, and the many colleagues who have provided insights and encouragement over the years of the RUEG project. We also gratefully acknowledge the 30 external reviewers and 23 internal reviewers whose careful commentary on earlier versions have substantially improved the quality of the chapters in this volume. Finally, we thank Lea Coy for help with pre-publication formatting.

## References

- Aalberse, Suzanne P., Ad Backus & Pieter Muysken (eds.). 2019. *Heritage languages: A language contact approach* (Studies in Bilingualism 58). Amsterdam: John Benjamins. DOI: 10.1075/sibil.58.
- Alexiadou, Artemis & Vasiliki Rizou. 2023. The use of periphrasis for the expression of aspect by Greek heritage speakers: A case study of register variation narrowing. *Register Studies* 5(1). 82–110. DOI: 10.1075/rs.20022.ale.
- Alexiadou, Artemis, Vasiliki Rizou & Foteini Karkaletsou. 2023. Agreement asymmetries with adjectives in Heritage Greek. *Languages* 8(2). 139. DOI: 10.3390/languages8020139.
- Blommaert, Jan, James Collins & Stef Slembrouck. 2005. Spaces of multilingualism. *Language and Communication* 25(3). 197–216. DOI: 10.1016/j.langcom.2005.05.002.
- Bonfiglio, Thomas Paul. 2010. *Mother tongues and nations: The invention of the native speaker*. Berlin, New York: De Gruyter Mouton.
- Brutt-Griffler, Janina & Keiko K. Samimy. 2001. Transcending the nativeness paradigm. *World Englishes* 20(1). 99–106. DOI: 10.1111/1467-971X.00199.
- Bunk, Oliver. 2024. What does linguistic structure tell us about language ideologies? The case of majority language anxiety in Germany. *European Journal of Applied Linguistics: Language ideologies: Old questions, new perspectives*. Naomi Truan & Esther Jahns (eds.). 91–116. DOI: 10.1515/eujal-2023-0049.
- Bunk, Oliver. In press. The anxious heritage speaker? Language anxiety and insecurity in multilingual contexts. In Artemis Alexiadou, Claudio Scarvaglieri, Christoph Schroeder & Heike Wiese (eds.), *The construction of multilinguals as others: Do we practice what we preach?* Berlin: Language Science Press.

## 1 Introduction: Investigating the dynamics of language contact situations

- Bunk, Oliver, Shanley E. M. Allen, Sabine Zerbian, Tatiana Pashkova, Yulia Zuban & Erica Conti. 2025. Information packaging and word order dynamics in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 379–412. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775181.
- Bunk, Oliver, Antje Sauermann & Fynn Raphael Dobler. 2024. Sociolinguistic variation in Kiezdeutsch and Namdeutsch: The case of the modal particles *halt* and *been*. In James M. Stratton & Karen V. Beaman (eds.), *Expanding variationist sociolinguistic research in varieties of German*. New York: Routledge.
- Bunk, Oliver, Britta Schulte & Heike Wiese. Forthcoming. Bare NPs in German: Results from a comparative corpus study. In Siegwald Lindenfelser, Angélica Prédiger & Renata Szczepaniak (eds.), *Deutsche Sprachminderheiten in der Welt: Empirische Studien zur Sprachvariation und Sprachideologie* (Germanistische Linguistik). Berlin: de Gruyter.
- Canagarajah, Suresh. 2013. *Translingual practice: Global Englishes and cosmopolitan relations*. London: Routledge. DOI: 10.4324/9780203073889.
- Cook, Vivian. 2016. Premises of multi-competence. In Li Wei & Vivian Cook (eds.), *The Cambridge handbook of linguistic multi-competence* (Cambridge Handbooks in Language and Linguistics), 1–25. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781107425965.001.
- Cook, Vivian & Li Wei (eds.). 2016. *The Cambridge handbook of linguistic multi-competence* (Cambridge Handbooks in Language and Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781107425965.
- Embick, David, Yosiane White & Meredith Tamminga. 2020. Heritage languages and variation: Identifying shared factors. *Bilingualism: Language and Cognition* 23(1). 21–22. DOI: 10.1017/S1366728919000476.
- Flores, Cristina & Esther Rinke. 2020. The relevance of language-internal variation in predicting heritage language grammars. *Bilingualism: Language and Cognition* 23(1). 25–26. DOI: 10.1017/S1366728919000464.
- Gogolin, Ingrid. 2002. Interkulturelle Bildungsforschung. In Rudolf Tippelt (ed.), *Handbuch Bildungsforschung*, 263–279. Wiesbaden: VS Verlag für Sozialwissenschaften. DOI: 10.1007/978-3-322-99634-3\_14.
- Grosjean, François. 1982. *Life with two languages: An introduction to bilingualism*. Cambridge, MA: Harvard University Press.
- Grosjean, François. 1997. The bilingual individual. *Interpreting* 2(1-2). 163–187. DOI: 10.1075/intp.2.1-2.07gro.
- Grosjean, François. 2008. *Studying bilinguals*. Oxford: Oxford University Press.

- Grosjean, François. 2010. *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press.
- Guijarro-Fuentes, Pedro & Katrin Schmitz. 2015. The nature and nurture of heritage language acquisition. *Lingua* 164. 239–250. DOI: 10.1016/j.lingua.2015.05.008.
- Iefremenko, Kateryna, Christoph Schroeder & Jaklin Kornfilt. 2021. Converbs in heritage Turkish: A contrastive approach. *Nordic Journal of Linguistics* 44(2). 130–154. DOI: 10.1017/S0332586521000160.
- Johannessen, Janne Bondi. 2018. Factors of variation, maintenance and change in Scandinavian heritage languages. *International Journal of Bilingualism* 22(4). 447–465. DOI: 10.1177/1367006918762161.
- Katsika, Kalliopi, Annika Labrenz, Kateryna Iefremenko & Shanley E.M. Allen. 2025. Discourse openings and closings across languages in contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 527–576. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775189.
- Keskin, Cem. 2025. Balkan Turkic as a model for understanding contact-induced change in Turkish. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 287–323. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775175.
- Keskin, Cem, Kateryna Iefremenko, Jaklin Kornfilt & Christoph Schroeder. 2024. Hybrid clause combining strategies in Turkish language contacts. *Studia Linguistica* (Early View). DOI: 10.1111/stul.12256.
- Klotz, Martin, Rahel Gajaneh Hartz, Annika Labrenz, Anke Lüdeling & Anna Shadrova. 2024. Die RUEG-Korpora: Ein Blick auf Design, Aufbau, Infrastruktur und Nachnutzung multilingualer Forschungsdaten. *Zeitschrift für germanistische Linguistik* 52(3). 578–592. DOI: 10.1515/zgl-2024-2026.
- Krause, Thomas & Amir Zeldes. 2014. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1). 118–139. DOI: 10.1093/llc/fqu057.
- Kupisch, Tanja & Maria Polinsky. 2022. Language history on fast forward: Innovations in heritage languages and diachronic change. *Bilingualism: Language and Cognition* 25(1). 1–12. DOI: 10.1017/S1366728921000997.
- Kupisch, Tanja & Jason Rothman. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5). 564–582. DOI: 10.1177/1367006916654355. (27 April, 2023).

## *1 Introduction: Investigating the dynamics of language contact situations*

- Labrenz, Annika, Kateryna Iefremenko, Kalliopi Katsika, Shanley E.M. Allen, Christoph Schroeder & Heike Wiese. 2025. Dynamics of discourse markers in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 493–526. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775187.
- Lohndal, Terje, Jason Rothman, Tanja Kupisch & Marit Westergaard. 2019. Heritage language acquisition: What it reveals and why it is important for formal linguistic theories. *Language and Linguistics Compass* 13(12). e12357. DOI: 10.1111/lnc3.12357.
- Lüdeling, Anke, Artemis Alexiadou, Aria Adli, Karin Donhauser, Malte Dreyer, Markus Egg, Anna Helene Feulner, Natalia Gagarina, Wolfgang Hock, Stefanie Jannedy, Frank Kammerzell, Pia Knoeferle, Thomas Krause, Manfred Krifka, Silvia Kutscher, Beate Lütke, Thomas McFadden, Roland Meyer, Christine Mooshammer, Stefan Müller, Katja Maquate, Muriel Norde, Uli Sauerland, Stephanie Solt, Luka Szucsich, Elisabeth Verhoeven, Richard Waltereit, Anne Wolfsgruber & Lars Erik Zeige. 2022. Register: Language users' knowledge of situational-functional variation. *Register Aspects of Language in Situation* 1(1). 1–58. DOI: 10.18452/24901.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehayé, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Merino, Barbara J. 1983. Language loss in bilingual Chicano children. *Journal of Applied Developmental Psychology* 4(3). 277–294. DOI: 10.1016/0193-3973(83)90023-0.
- Montrul, Silvina. 2015. *The acquisition of heritage languages*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO978139030502.
- Muysken, Pieter. 2020. The case for contact induced-change in heritage languages. *Bilingualism: Language and Cognition* 23(1). 37–38. DOI: 10.1017/S1366728919000373.
- Myers-Scotton, Carol. 2006. *Multiple voices: An introduction to bilingualism*. Malden: Blackwell.

- Ortega, Lourdes, Andrea E. Tyler, Hae In Park & Mariko Uno. 2016. *The usage-based study of language learning and multilingualism*. Georgetown University Press.
- Özsoy, Onur & Frederic Blum. 2023. Exploring individual variation in Turkish heritage speakers' complex linguistic productions: Evidence from discourse markers. *Applied Psycholinguistics* 44(4). 534–564. DOI: 10.1017/S0142716423000267.
- Özsoy, Onur, Vasiliki Rizou, Maria Martynova, Natalia Gagarina, Luka Szucsich & Artemis Alexiadou. 2025. Null subjects in heritage Greek, Russian and Turkish. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 179–217. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775167.
- Pascual y Cabo, Diego & Jason Rothman. 2012. The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics* 33(4). 450–455. DOI: 10.1093/applin/ams037.
- Pashkova, Tatiana, Marlene Böttcher, Kalliopi Katsika, Sabine Zerbian & Shanley E. M. Allen. 2025. Majority English of heritage speakers. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 449–491. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775185.
- Polinsky, Maria & Gregory Scontras. 2020. Understanding heritage languages. *Bilingualism: Language and Cognition* 23(1). 4–20. DOI: 10.1017/S1366728919000245.
- Purkarthofer, Judith, Sofia Grigoriadou & Rosemarie Tracy. 2023. Von der Forschung in die Familie oder Schule. In Judith Purkarthofer (ed.), *Sprach-verhandeln* (Schulheft), 123–128. Innsbruck: Studienverlag.
- Purkarthofer, Judith, Rosemarie Tracy, Sofia Grigoriadou & Johanna Tausch. 2025. Family language dynamics: Strengthening heritage speakers' linguistic resources. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 111–138. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775163.
- Rizou, Vasiliki, Onur Özsoy, Maria Martynova, Luka Szucsich, Artemis Alexiadou & Natalia Gagarina. 2025. Dynamics of verbal aspect in heritage Greek, Russian and Turkish. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 219–249. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775170.
- Romaine, Suzanne. 1989. *Bilingualism* (Language in Society 13). Oxford: Blackwell.

## 1 Introduction: Investigating the dynamics of language contact situations

- Rothman, Jason. 2007. Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism* 11(4). 359–390.
- Rothman, Jason, Fatih Bayram, Vincent DeLuca, Grazia Di Pisa, Jon Andoni Duñabeitia, Khadij Gharibi, Jiuzhou Hao, Nadine Kolb, Maki Kubota, Tanja Kupisch, Tim Laméris, Alicia Luque, Brechje van Osch, Sergio Miguel Pereira Soares, Yanina Prystauka, Deniz Tat, Aleksandra Tomić, Toms Voits & Stefanie Wulff. 2023. Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics* 44(3). 316–329. DOI: 10.1017/S0142716422000315.
- Rothman, Jason & Jeanine Treffers-Daller. 2014. A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics* 35(1). 93–98. DOI: 10.1093/applin/amt049.
- Schroeder, Christoph. 2016. Clause combining in Turkish as a minority language in Germany. In Mine Güven, Didar Akar, Balkız Öztürk & Meltem Kelepir (eds.), *Exploring the Turkish linguistic landscape: Essays in honor of Eser Erguvanlı-Taylan* (Studies in Language Companion Series), 81–102. Amsterdam: John Benjamins. DOI: 10.1075/slcs.175.06sch.
- Schroeder, Christoph, Kateryna Iefremenko, Kalliopi Katsika, Annika Labrenz & Shanley E. M. Allen. 2025. Clause combining in narrative discourse: A contrastive analysis across heritage and majority languages. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 251–286. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775173.
- Scontras, Gregory, Zuzanna Fuchs & Maria Polinsky. 2015. Heritage language and linguistic theory. *Frontiers in Psychology* 6. 1545. DOI: 10.3389/fpsyg.2015.01545.
- Scontras, Gregory, Maria Polinsky & Zuzanna Fuchs. 2018. In support of representational economy: Agreement in heritage Spanish. *Glossa: A journal of general linguistics* 3(1).
- Shadrova, Anna, Martin Klotz, Rahel Gajaneh Hartz & Anke Lüdeling. 2025. Mapping the mappings and then containing them all: Quality assurance, interface modeling, and epistemology in complex corpus projects. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 69–110. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775161.
- Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles* (Oxford Studies in Language Contact). Oxford: Clarendon Press.

- Sorace, Antonella. 2011. Cognitive advantages in bilingualism: Is there a “bilingual paradox”? In *Multilingualism: Language, Power, and Knowledge*, 335–358. Edistudio.
- Tausch, Johanna & Wintai Tsehay. 2023. Mehrsprachigkeit als Ressource: Ein Einblick in mehrsprachiges Sprachverhalten im Alltag. In Judith Purkarthofer (ed.), *Sprach-ver-handeln* (Schulheft), 109–122. Innsbruck: Studienverlag.
- Tsehay, Wintai, Tatiana Pashkova, Rosemarie Tracy & Shanley E. M. Allen. 2021. Deconstructing the native speaker: Further evidence from heritage speakers for why this horse should be dead! *Frontiers in Political Science* 12. 1–14. DOI: [10.3389/fpsyg.2021.717352](https://doi.org/10.3389/fpsyg.2021.717352).
- Tsehay, Wintai, Rosemarie Tracy & Johanna Tausch. 2025. Inter- and intra-individual variation: How it materializes in Heritage German and why it matters. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 141–177. Berlin: Language Science Press. DOI: [10.5281/zenodo.15775165](https://doi.org/10.5281/zenodo.15775165).
- Wiese, Heike. 2013. What can new urban dialects tell us about internal language dynamics? The power of language diversity. In Werner Abraham & Elisabeth Leiss (eds.), *Dialektologie in neuem Gewand: Zu Mikro-/Varietätenlinguistik, Sprachenvergleich und Universalgrammatik* (Linguistische Berichte Sonderheft 19), 207–245. Hamburg: Buske.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers’ repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike. 2023. *Grammatical systems without language borders: Lessons from free-range language* (Conceptual Foundations of Language Science 9). Berlin: Language Science Press. DOI: [10.5281/zenodo.10276182](https://doi.org/10.5281/zenodo.10276182).
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: [10.3389/fpsyg.2021.717973](https://doi.org/10.3389/fpsyg.2021.717973).
- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers’ repertoires: Elicitation of register-differentiated productions across groups. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 33–67. Berlin: Language Science Press. DOI: [10.5281/zenodo.15775159](https://doi.org/10.5281/zenodo.15775159).

## *1 Introduction: Investigating the dynamics of language contact situations*

- Zerbian, Sabine, Yulia Zuban, Marlene Böttcher & Oliver Bunk. 2025. Intonation in heritage languages. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 413–448. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775183.
- Zuban, Yulia, Maria Martynova, Sabine Zerbian, Luka Szucsich & Natalia Gagrina. 2021. Word order in heritage Russian: Clause type and majority language matter. *Russian Linguistics* 45(3). 253–281. DOI: 10.1007/s11185-021-09246-1.
- Zürn, Nadine, Mareike Keller, Rosemarie Tracy & Anke Lüdeling. 2025. Dynamic properties of the heritage speaker lexicon. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 325–375. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775179.



## **Part I**

# **Methodology and application**



# Chapter 2

## Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups

Heike Wiese<sup>a</sup>, Annika Labrenz<sup>a</sup> & Albrun Roy<sup>a</sup>

<sup>a</sup>Humboldt-Universität zu Berlin

This chapter addresses the challenge of tapping into speakers' repertoires to elicit ecologically valid data. It describes a set-up that draws on the LangSit method (Wiese 2020) and has been implemented by the RUEG group to elicit data as a basis for the RUEG corpus (cf. Shadrova et al. 2025 [this volume]). Elicitations yielded comparable productions from informal and formal, spoken and written situations across bilingual and monolingual speakers, adults and adolescents, and heritage and majority languages. We discuss strengths and limitations of the set-up and present analyses that illustrate two key strengths of this method: (1) it is successful in eliciting register-differentiated data across different communicative situations, and (2) this data is naturalistic in the sense that it follows the patterns of spontaneous, non-elicited language productions.

### 1 Introduction

In this chapter, we discuss a method of elicitation that addresses the challenge of tapping into speakers' repertoires to elicit ecologically valid data: we present the set-up that has been implemented by the RUEG group<sup>1</sup> to elicit the data for

<sup>1</sup>Research Unit “Emerging Grammars in Language Contact Situations: A Comparative Approach”; see Wiese et al. (2025 [this volume]) and <https://hu.berlin/rueg> (last accessed October 23rd, 2023). The corpus created by the group, the open-access RUEG corpus, is described in Shadrova et al. (2025 [this volume]) and can be accessed via <https://hu.berlin/rueg-corpus> (last accessed October 23rd, 2023).



the RUEG corpus. We show that this kind of set-up yields production data that is elicited and thus comparable across different settings and groups, but at the same time representative of speakers' natural behaviour in different communicative situations.

When investigating natural languages, it is important to get a realistic picture of language users' practices and competences. Language use always involves different situation-dependent choices from a broader linguistic repertoire, and accordingly we should take into account such variation. To get a realistic picture of linguistic practices and language competence, we cannot confine ourselves to, say, formal situations associated with standard language. Rather, we need to tap into broader speaker repertoires and cover different registers. At a minimum, this should include language use in informal as well as formal communicative situations.<sup>2</sup>

While this might seem obvious, it is something that often gets overlooked in research. One area where we see this as a desideratum is research on heritage-language speakers. These are bi- or multilingual speakers who grew up in a bi- or multilingual home with a minority language in addition to the majority language of the larger society (see Introduction).

Heritage languages are often used primarily in informal domains: they are acquired in the family and often remain associated with informal communicative situations. In contrast, languages spoken as a societal majority language will also be used in more formal situations, and the acquisition of formal registers of such languages will be further supported as part of schooling (see also Kupisch & Rothman 2018).

Yet, heritage language research often looks at data elicited under lab conditions that are favourable to more formal language, and compares heritage speakers' data to that of monolinguals who speak the language as a majority language. In such lab settings, monolinguals will choose formal registers, while heritage speakers might use more informal versions if formal registers of the heritage language are not part of their repertoire. This, then, brings in an additional factor: informal patterns in bilinguals are measured against formal ones in monolinguals. Hence, if we do not want to compare apples and oranges, we need to take into account informal as well as formal communicative situations across groups, in monolinguals and multilinguals alike.

---

<sup>2</sup>Following the terminology introduced in the Introductory chapter (Wiese et al. 2025 [this volume]), we understand communicative situations as the settings of a communication, and registers as the linguistic counterparts of communicative situations, that is, the language use associated with different communicative situations.

Furthermore, it can be valuable to include both spoken and written language. Heritage speakers typically do not acquire literacy in their heritage language as part of formal schooling, so it might depend on their media consumption (books, newspapers) in how far they acquire formal-written registers. Informal-written registers of the heritage language might be used in digital social media, for instance transnationally in interactions within extended families, but this might be not as common as, say, text messaging in the majority language. Including spoken vs. written as well as formal vs. informal communicative situations can therefore allow us to tap into and tease apart different relevant domains of speakers' linguistic resources.

Taking into account such register distinctions provides us with ecologically more valid data on both monolingual and bi-/multilingual speakers' competences.<sup>3</sup> Furthermore, comparing both groups on equal grounds gives us a better picture of heritage language characteristics, since it allows for a better assessment: What phenomena are really specific to heritage languages, language contact or multilingualism? What phenomena, in contrast, might actually be found across speaker groups if we do not restrict ourselves to, say, formal situations?

This is not only relevant for the heritage language, but also for another part of heritage speakers' resources, namely the majority language. Heritage speakers who grow up in a country with a different majority language are often not acknowledged as part of the core speaker group of that language, and their linguistic practices are then approached from a deficit perspective. Such effects of a widespread monolingual habitus in the Global North are evident not only in public discourse, but also in linguistic papers.<sup>4</sup> As a result, patterns of linguistic variation in informal or spoken registers might not be recognised as such, but attributed to bi-/multilingualism. If we take into account register-differentiated use in bilinguals and monolinguals alike, such misattributions can be avoided. This can provide new insights into native grammars, their dynamics in multilinguals, and their variability across registers (see also Wiese 2013, Kupisch & Polinsky 2022 on the special dynamics of multilingual settings).

Wiese et al. (2022) discuss cross-linguistic examples for the kind of findings that such an approach can yield. Tapping into speakers' broader repertoires re-

---

<sup>3</sup>In corpus-linguistic research, the problem of the authenticity (or naturalness) of corpus data was discussed as early as the 1990s: one of the questions was what kind of texts (e.g. informal/-formal; oral/written; literary/ordinary) should be included in a sample so that a corpus reflects "authentic" language (e.g. Sinclair 1991) or is representative of a particular language use (e.g. Biber 1993).

<sup>4</sup>See Wiese et al. (2022) for a critique; Adli & Guy (2022) for a call to broaden our perspective and reduce the bias towards Western societies.

vealed some unexpected patterns not just in heritage speakers, but also in monolinguals. We found, for instance, noncanonical word order patterns, such as new referents in post-verbal position in Turkish<sup>5</sup> or the placement of two constituents before the finite verb in main declaratives in German.<sup>6</sup> Both of these patterns are ruled out by established grammatical descriptions of Turkish and German, respectively. Yet, we found them to be systematically available in native grammars. In majority language settings, such patterns were associated with informal and/or spoken language, indicating their register-related status. Hence, when heritage speakers use such noncanonical patterns in more formal settings, this points to register levelling rather than novel patterns.

Such findings underline the need to cast our net wider and target ecologically valid data in a broader range of communicative situations: had we focused on formal situations alone, such patterns might only have shown up in heritage speakers and might then have been misattributed to bilingualism. In contrast, if we tap into actual language use and look at broader repertoire data from multilinguals and monolinguals alike, then this can reveal interesting similarities of noncanonical patterns and inform us on the dynamics of native grammars (see also Bayram et al. 2019 on the problem of using an idealised monolingual standard as a yardstick).

However, it is notoriously difficult to gain ecologically valid data that is representative of speakers' natural behaviour and covers not only formal, but also informal settings. The empirical basis for much of heritage language research so far has often been limited to data elicited under lab conditions (see also Montrul 2015: Chapter 6, Polinsky 2018: Chapter 3 for examples). Common set-ups are, for instance, ones that use acceptability, sentence completion, or picture matching tasks. This kind of elicitation has two major advantages. For one, it yields controlled data that is suitable for targeting specific linguistic phenomena that have been of interest for the analysis, for instance, certain morphological or syntactic patterns in a heritage language. Second, it yields comparable data across speaker groups, and it has typically been employed to compare heritage speakers to monolinguals (but see Rothman et al. 2023 for a call to further reflect on the choice of comparison groups in heritage language research). However, such elicited data also has a major disadvantage: it is not representative of speakers' natural behaviour across different communicative situations (see also Aalberse et al. 2019: Chapter 5 for a discussion of this problem). Such data reflects only those responses in a somewhat artificial setting that will favour formal language

---

<sup>5</sup>Wiese et al. (2022); for detailed discussion of this pattern see also Schroeder et al. (2024).

<sup>6</sup>Wiese et al. (2022); see also Wiese & Müller (2018) for an analysis of verb-third in German.

use, and as we have argued above, this can mean missing out on important patterns in multilinguals and monolinguals alike.

A type of method that avoids artificial settings and has also been employed in minority language research is the collection of spontaneous data, for instance through speakers' self-recordings (see Aalberse et al. 2019: Chapter 5 for examples). This kind of data can be regarded as the gold standard from the point of view of ecological validity, but it is much more difficult and time-consuming to collect. Furthermore, since such data is spontaneous and thus by its very nature not controlled, it does not offer the same advantages as elicited data: specific linguistic patterns might just not occur (at all) in the database, and it does not yield comparable productions across speakers and speaker groups. Additionally, in an uncontrolled setting, the data might only represent a specific part of a speaker's repertoire, e.g., informal conversations with friends, but not, say, formal registers used with strangers.

A kind of methodology that addresses these challenges and combines the advantages of controlled and spontaneous data is the elicitation of naturalistic data. This is data that is representative of speakers' natural behaviour, but at the same time is elicited and thus can be more controlled, making it suitable for targeted and comparative investigations. In what follows, we present a set-up that achieves this and provides data that covers speakers' behaviour across different communicative situations. This is the "Language Situation" (LangSit) method. In Section 2, we describe this methodology and its implementation by the RUEG group to yield comparable data across a broad range of different settings and speakers, which supported such empirical findings on heritage speakers and monolinguals as mentioned above. Section 3 will present analyses that illustrate two key strengths of this method: (1) it is successful in eliciting *register-differentiated* data across formal and informal, written and spoken communicative situations, and (2) this data is *naturalistic* in the sense that it follows the patterns of spontaneous, non-elicited language productions. In the final section (Section 4) we conclude this chapter and discuss the implications of our results.

## 2 Naturalistic elicitations across groups based on the LangSit method

### 2.1 LangSit: General features

The LangSit method has been used to elicit naturalistic data across different communicative situations in a range of different studies, in particular on multilingual

settings, but not restricted to these, and it provides the set-up for RUEG's empirical basis, the RUEG corpus (see Shadrova et al. 2025 [this volume] and Klotz et al. 2024). In this subsection, we describe its core features; for a more detailed discussion (including an overview of previous applications) see Wiese (2020) and the LangSit website (<https://hu.berlin/LangSit>).

The main idea underlying the LangSit method is to encourage speakers to play-act *natural communication in different situations*. The first part aims at eliciting naturalistic language, the second part targets register-differentiated productions, allowing us to cover different parts of speakers' repertoires. Finally, in order to make productions *comparable* across the different communicative situations, speakers are asked to talk about the same event in each situation.

This event is presented to participants through a nonverbal stimulus (to prevent linguistic priming), for instance a photo story or a short video. In order to make the setting ecologically valid, the event should be interesting enough to motivate people to talk about it. For instance, for a corpus of Namibian German (the DNAm Corpus<sup>7</sup>), we showed participants a photo story where a car bumped into a trolley on a shopping mall car park, causing a woman to fall down and spill the contents of her bag (cf. Wiese et al. 2017, Zimmer et al. 2020).

Participants are asked to imagine being a witness to the event and then to play-act telling different people about this. Through the choice of different interlocutors, researchers can specify different communicative situations. For instance, asking speakers to tell a friend or a relative about what they saw will simulate an informal situation; asking them to tell a police officer or a teacher simulates a formal situation. Further domains of speakers' repertoire can be targeted by manipulating additional aspects of the language production, for instance the mode by asking participants to write or to speak, or the monologic vs. dialogic communicative character by asking them to leave a message for someone or to speak to each other.

To elicit this kind of language production, participants are familiarised with the event and then asked to act out telling someone about this. For instance, in the informal situation, they are told something along the lines of "Imagine you just witnessed this event. What would you do? You might, for instance, call your friend and tell her about this, or you might be asked to describe this to a police officer who needs a witness. Let's act this out. Talk to a friend about it. Which friend would you call? Now take your phone and do as if send them a voice message."

---

<sup>7</sup>See <https://www.linguistik.hu-berlin.de/de/institut/professuren/multilinguale-kontexte/korpora/dnam> (last accessed 2025-02-13).

The fact that participants are asked to talk to different types of interlocutors that are associated with informal vs. formal communicative situations is advantageous not only because it supports register-differentiated data. It also reduces speakers' reluctance to produce informal language in the setting of an elicitation, since it is clear that this will not reflect badly on their language competence, given that they are asked to produce formal language as well. This focus on different situational choices thus helps speakers to overcome the restraints of standard language ideologies, something that is a notorious obstacle to eliciting informal language use (Wiese 2020). On the other hand, for speakers who might not have formal registers of a heritage language in their repertoire, it can be of help for them to know that they are also asked to produce informal language: the latter allows them to act out in a setting that they associate with their heritage language and where they feel secure to use it. This can reduce the kind of language anxiety that might hold them back and distort their responses in a formal lab setting (see Sevinç & Dewaele 2018 on language anxiety in bilingual speakers).

Finally, the choice of event can guide speakers towards specific topics (e.g., make them talk about a trolley or about someone falling down) and thus control to some degree the kind of linguistic material they use, which can help target specific linguistic domains (e.g. nouns of a certain gender, or motion verbs).

Taken together, the LangSit method is open enough to yield ecologically valid, naturalistic data; it is powerful enough to cover different communicative situations and thus elicit register-differentiated productions; yet it is restrictive enough to control the type of situations and topics and to yield comparable data across speakers, speaker groups, and (contact-linguistic) contexts.

## **2.2 RUEG's LangSit implementation**

The RUEG group employed the LangSit method to collect data for our shared empirical basis. As described in the introductory chapter (Wiese et al. 2025 [this volume]), our overarching goal was to investigate heritage speakers' linguistic competences and the dynamics of language contact. To this aim, we explored the distinctive grammatical and pragmatic characteristics of heritage language use, the role of bilingualism and the impact of the contact situation. To achieve this, we conducted large-scale comparisons across languages, speaker groups, and settings (see Wiese et al. 2025 [this volume]). The close comparability that LangSit affords was hence a major advantage for us. A second advantage was its power to yield naturalistic data, for bilinguals and monolinguals alike.

When implementing the LangSit method, we made a number of specifications to the general set-up, which we describe in the following paragraphs. To ensure uniform data collection, we conducted an on-site training course for all elicitors that also included a video tutorial. All materials we used are fully open-access, available through RUEG's site on the Open Science Foundation (<https://osf.io/cm96g/>). This includes an overview of the set-up, all stimuli, detailed instructions for elicitors, and the training video for data collection.

We used a *video stimulus* to present the event, in order to give participants a lively and natural impression that would make it easier for them to imagine themselves as a witness. Following previous LangSit implementations, we chose a minor traffic accident, as an event that is notable enough to motivate telling about it, but not so severe as to disturb viewers.

Since we conducted our elicitations in *five countries* (Germany, Greece, Russia, Turkey, and the US), we tried to avoid any elements in this accident that were too specific to work equally well in each country. When piloting the set-up, we had used an accident between a car and a bike at a traffic light. The results of piloting showed that this was too country-specific: among other things, traffic lights look different across countries, and cyclists are less common in some countries than in others. Accordingly, for the main study we set the accident on a car park, with no bike involved. This said, one cannot altogether eliminate the possibility that country-specific differences have an impact, for instance we might find different attitudes and experiences with speaking to the police in different countries or for different speaker groups (for instance, depending on the groups, there might be higher levels of language anxiety in heritage speakers in such communicative situations).

The story is about a woman with a dog on a leash who is loading groceries into her car when a young man and a woman with a baby pram walk towards her from across the car lane. The man plays with a ball that suddenly rolls into the lane towards the woman with the dog. The dog breaks free and runs into the lane towards the ball, making the woman drop her groceries. Meanwhile, two cars have approached; the driver of the first car has to break abruptly for the dog, and as a result the second car bumps into the first car. The two drivers get out and call the police. The young man checks on them and then helps the dog owner retrieve her groceries.

We used the same video across all countries, except for one element: in order to show that the car drivers called the police, there was a close-up of one driver's mobile phone showing the police telephone number, which is different in different countries, so we used country-specific pictures for that.

To cover a broad spectrum within speakers' repertoires, we defined *four communicative situations* that differed along the parameters of formality and mode: informal-spoken, informal-written, formal-spoken, and formal-written. (In-)formality was induced by different *interlocutors*: for the informal productions, participants were asked to play-act telling a friend about the incident; for the formal ones, describing the accident as a witness for the police. Hence, the parameter of formality was operationalised through different interlocutors, with a friend representing a familiar interlocutor in a more symmetric and closer relationship to the speaker, and a police officer representing an unfamiliar interlocutor in a more asymmetric and professional relationship to the speaker. This does not cover all possible aspects that can contribute to the formality vs. informality of a communicative situation, of course, but it captures some key distinctions that are likely to constitute situations of different formality for speakers.

To enhance comparability across spoken and written productions, communication was set up as monologic in both modes. Participants were asked to type texts in the written mode and to leave voice messages in the spoken mode; hence all productions were *non-simultaneous*. In the case of Russian and Greek, we left the choice of script (Latin, Cyrillic, Greek) open for heritage speakers.

In the informal productions, participants sent messages via WhatsApp from a mobile phone provided by us, with auto correction, swiping, and suggestions switched off. For formal-spoken productions, we used the same mobile phone, where we had created a contact "Police Department – eyewitness line"; participants were told to imagine that this was a line for witnesses to leave a voice mail with their reports, and that they had been asked to do this for the accident they had seen. For formal-written productions, participants were asked to type a witness report for the police on a laptop we provided.

Across participants, we *balanced* the *order* of communicative situations, that is, we had a balanced order of informal and formal parts, and of written and spoken elicitations within informal and formal parts.

After elicitations, participants filled in a digital *questionnaire* with questions on their personal details (e.g., age, date and place of birth, educational background, family members), language biographies, language and media use. This information was recorded as a basis for analyses targeting the differential impact such factors might have on language productions.<sup>8</sup>

It is generally challenging to elicit informal productions. As discussed above, the LangSit set-up somewhat mitigates this problem by its emphasis on speakers'

---

<sup>8</sup>The importance of documenting metadata that might have an impact on language variation (e.g., available metadata on texts and producers) has been emphasised in corpus linguistics for a long time (cf. Sinclair 1991, Granger 2008).

repertoires, which makes it clear for participants that informal language use will not be judged as incorrect or bad. To make it easier for participants to differentiate informal and formal elicitations, there was a break between them. In addition, we included a conversational session (*chit chat*) before the informal elicitation, in order to help participants to get into the spirit of an informal conversation.

We further supported the production of naturalistic informal as well as formal language by creating different settings for informal and formal elicitations, with different surroundings and elicitors. We used two elicitors for each meeting. “*Formal*” elicitors wore a formal outfit (for instance, a suit) and interacted with participants in a distanced and formal manner, including through their way of speaking, for instance, they used the formal *Sie*-form in German (the *vos* variant in the T-V distinction, see e.g. Bilá et al. 2020), addressed participants with their surname, and used language close to the standard. In addition to the elicitation, they also organised all the formalities, e.g., they handed out speaker information forms, consent-forms, and receipts for the reimbursement. “*Informal*” elicitors wore informal clothes (for instance, jeans, a sweatshirt and a basecap), acted in a relaxed and accessible manner, offered cookies and drinks, and used more vernacular language, using participants’ first name, and e.g., addressing them with the informal *du*-form (*tu* variant) in German. They also interacted with participants during the chit-chat sessions.

Formal and informal elicitations took place in different rooms designed to further induce (in)formality: the *formal room* was an office room with a desk and no decoration; the *informal room* had comfortable seating and was made to look cosy and informal, for instance, by draping colourful cloths over chairs, putting a candle on the table, using mismatched mugs etc.

Since we were interested in speakers’ competences and language use across their repertoire, we included *both languages* for bilingual speakers. This meant that bilingual speakers did two sessions, one in their heritage language and one in the country’s majority language. The sessions were at least three days apart. Again, the order was balanced. Each session was conducted in a monolingual mode, with elicitors who used only the respective language (either the heritage or the majority language). This was owed to our research interest in the dynamics within heritage and majority languages, both for bilinguals and monolinguals.

Taken together, this allowed us to collect comparable data targeting speakers’ repertoires across speaker groups, countries, and heritage and majority languages. The data was processed and integrated into a shared empirical basis, the *RUEG corpus* (Lüdeling et al. 2024), an annotated, multi-layer and multimodal open-access corpus which is freely available in an open repository (see also Shadrova et al. 2025 [this volume] and Klotz et al. 2024).

As mentioned above, one of the strengths of this database was that it revealed unexpected patterns in monolingual as well as bilingual speakers, allowing for new insights into native grammars. The large-scale comparisons that it made possible enabled us to tease apart different factors underlying linguistic variation and the development of noncanonical patterns, and to pinpoint the role of language contact (see contributions in this volume). In the next section, we evaluate the power of our methodology to yield naturalistic register-differentiated productions: How successful was our method in tapping into speakers' repertoires, that is, did we manage to elicit data that is representative for different registers? And how close is this elicited data to spontaneous language productions under natural conditions, that is, how high is the price we pay for using controlled elicitations that yield comparable data?

In what follows, we arrive at a positive evaluation when addressing these questions: we show that our set-up was successful in eliciting register-differentiated data across formal and informal, written and spoken communicative situations (Section 3.1), and we show that the elicited data is naturalistic in the sense that it follows the patterns of spontaneous, non-elicited language productions (Section 3.2). For the first part, we present comparisons within the RUEG corpus, for the second, we present comparisons of the informal-spoken RUEG corpus data with a corpus of spontaneous WhatsApp voice messages.

For comparisons within the RUEG corpus, we target phenomena at different linguistic levels, namely syntax (Section 3.1.1), discourse organisation (Section 3.1.2), and the lexicon (Section 3.1.3), and show that they vary according to our four communicative situations. Comparisons of our corpus data with spontaneous voice messages indicate similar patterns, and again we show this for a range of phenomena at the levels of syntax (Section 3.2.2), discourse organisation and the lexicon (Section 3.2.3). Analyses of the RUEG data are based on the RUEG-1.0-SNAPSHOT corpus version (2023).

### **3 Register-differentiated naturalistic productions**

#### **3.1 Register distinctions within the RUEG corpus**

##### **3.1.1 Syntax**

A central domain that has been found to be sensitive to register distinctions is that of syntactic categories, and a prominent example is the distribution of nouns vs. pronouns. As Biber & Conrad (2019: Chapter 3) showed, pronouns tend to be more frequent in situations with a higher degree of interaction between

interlocutors where they refer to referents that the addressee is familiar with or that are present in the communicative situation. In contrast, full noun phrases can convey information that is unknown to the addressee and are therefore more frequent in situations where interlocutors are less familiar with each other.

Accordingly, if our elicitation schema worked out, we should expect a more frequent use of pronouns in the corpus data from spoken and informal elicitations, and a more frequent use of nouns in the corpus data from written and formal elicitations.

To test this, we used the universal part-of-speech tags that the RUEG corpus annotations provide for all five languages (English, German, Greek, Russian, and Turkish).<sup>9</sup> We conducted a corpus query based on the tags “NOUN” and “PROPN” for nouns (common nouns and proper nouns) and “PRON” for pronouns. For comparison of frequencies, we provide normalised frequencies, calculated as occurrences per 100 tokens, where the total number of tokens included all tokens except punctuation marks and filled pauses. We conducted Kruskal-Wallis tests and posthoc pairwise Wilcoxon tests.<sup>10</sup> Figure 1 gives the results for nouns, and Table 1 shows normalised rates of occurrences per 100 tokens for nouns across languages and registers.

The pattern confirms the expected register differences for all languages (“>” indicates significant differences of  $p < 0.01$ ):

formal-written > informal-written, formal-spoken > informal-spoken

We hence see an impact of both the parameters we manipulated for register distinctions: formality as well as written mode supported the use of nouns. The two parameters reinforced each other, with the effect that the setting with two positives (formal and written) elicited most nouns, that with two negatives (informal and spoken) the least, and the two conditions that combine a positive and a negative instance (informal and written or formal and spoken) occupy a shared intermediate position. Table 2 gives the results for pronouns.

---

<sup>9</sup><https://universaldependencies.org/u/pos/> (last accessed: 2025-02-14; see Shadrova et al. 2025 [this volume] for more details)

<sup>10</sup>Note that the standard deviations are not high in this domain, which is advantageous for the comparisons and might not hold for some other linguistic domains. As is frequently mentioned in heritage language research, we often observe a large variability in heritage speaker data (see, e.g., Montrul 2015, Polinsky 2018). Note, though, that this is not restricted to this speaker group: as shown in Wiese et al. (2022), we can find the same or even a higher degree of variability in monolinguals (see also Shadrova et al. 2025 [this volume] on the massive variability in naturalistic language in general).

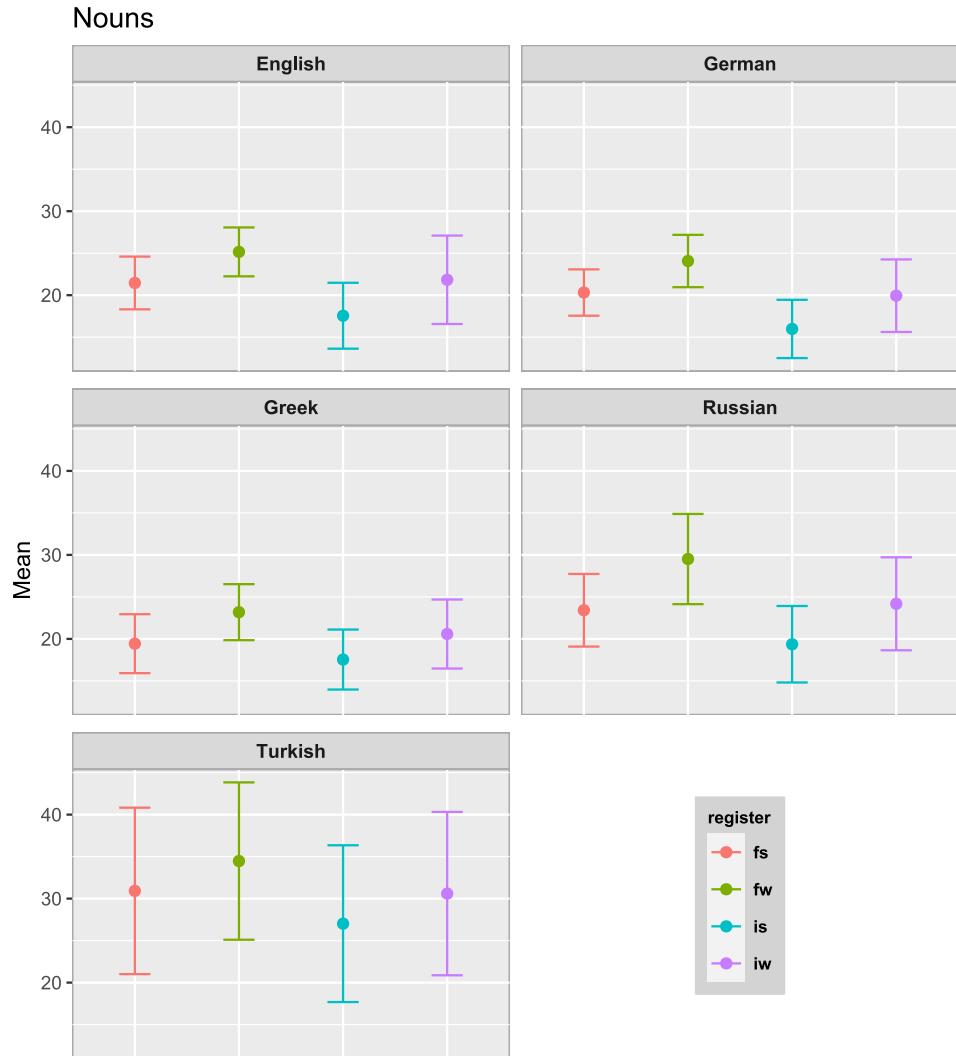


Figure 1: Means (dots) and standard deviations of normalised frequencies of nouns across languages

Table 1: Means of normalised occurrences per 100 tokens (“No.”) and standard deviations (“SD”) for nouns across languages and registers

| Language | Formal  |      |        |      | Informal |      |        |      |
|----------|---------|------|--------|------|----------|------|--------|------|
|          | written |      | spoken |      | written  |      | spoken |      |
|          | No.     | SD   | No.    | SD   | No.      | SD   | No.    | SD   |
| English  | 25.16   | 2.91 | 21.46  | 3.14 | 21.83    | 5.27 | 17.56  | 3.92 |
| German   | 24.07   | 3.12 | 20.32  | 2.76 | 19.94    | 4.32 | 16.00  | 3.47 |
| Greek    | 23.18   | 3.33 | 19.43  | 3.51 | 20.58    | 4.12 | 17.54  | 3.57 |
| Russian  | 29.51   | 5.37 | 23.41  | 4.32 | 24.18    | 5.54 | 19.36  | 4.55 |
| Turkish  | 34.47   | 9.37 | 30.91  | 9.91 | 30.60    | 9.33 | 27.02  | 9.73 |

Table 2: Means of normalised occurrences per 100 tokens (“No.”) and standard deviations (“SD”) for pronouns across languages and registers

| Language | Formal  |      |        |      | Informal |      |        |      |
|----------|---------|------|--------|------|----------|------|--------|------|
|          | written |      | spoken |      | written  |      | spoken |      |
|          | No.     | SD   | No.    | SD   | No.      | SD   | No.    | SD   |
| English  | 4.49    | 2.23 | 7.38   | 3.05 | 8.94     | 3.43 | 10.05  | 3.50 |
| German   | 6.22    | 2.34 | 8.02   | 2.70 | 8.35     | 3.35 | 10.10  | 3.39 |
| Greek    | 7.29    | 3.07 | 9.23   | 3.20 | 8.90     | 3.62 | 9.50   | 3.17 |
| Russian  | 6.76    | 3.25 | 10.05  | 3.98 | 9.03     | 3.92 | 10.74  | 3.82 |
| Turkish  | 4.07    | 2.66 | 4.96   | 3.08 | 5.66     | 3.33 | 6.10   | 3.85 |

This pattern confirms our expectation to find least pronouns in the formal-written condition, but unlike the pattern for nouns, we find some cross-linguistic differences for the other conditions. For English, and German, the pattern mirrors that for nouns (“>” indicates significant differences of  $p < 0.01$ ):

informal-spoken > formal-spoken, informal-written > formal-written

For Turkish, Greek, and Russian the pattern is similar, but with slight differences: For Turkish and Greek, there is no significant difference between frequencies in the informal-spoken condition and the two intermediate ones (formal-spoken and informal-written):

informal-spoken, formal-spoken, informal-written > formal-written

For Russian, there is no significant difference between frequencies in the formal and informal-spoken modes, but there is one between those and the written conditions as well as between the informal-written and formal-written conditions. This brings mode to the fore as a relevant parameter for the frequent use of pronouns:

informal-spoken, formal-spoken > informal-written > formal-written

The difference between German and English, as opposed to Turkish, Greek, and Russian is probably due to typological differences. Russian, Turkish and Greek are languages characterised by frequent pro-drop (see Özsoy et al. 2025 [this volume]<sup>11</sup>). As shown in Table 3, we found notably fewer personal pronouns in Turkish and Greek compared to the other languages. For Russian this holds especially for the written mode. These differences might account for the lack of significant differences between registers regarding the category of pronouns. It seems that in pro-drop languages, pronouns are a less decisive linguistic indicator for register differences.

Table 3: Means of normalised rates of occurrences per 100 tokens of personal pronouns across languages and registers

|         | Formal  |        | Informal |        |
|---------|---------|--------|----------|--------|
|         | written | spoken | written  | spoken |
| English | 5.39    | 3.74   | 6.22     | 7.95   |
| German  | 4.34    | 3.21   | 4.17     | 4.91   |
| Greek   | 1.07    | 0.41   | 0.24     | 1.19   |
| Russian | 3.42    | 1.67   | 2.55     | 3.77   |
| Turkish | 2.36    | 1.73   | 1.43     | 2.03   |

Across communicative situations, nouns are more frequent overall than pronouns. This is presumably due to the specific communicative task that participants had, namely to retell a car accident where a lot of new referents had to be introduced (several persons, a dog, a ball, cars, a pram etc.). This points to a general challenge of the LangSit method: the topic under discussion can have an

<sup>11</sup>For specific effects of mode and formality on subject drop in Russian and Turkish and on demonstratives in Russian, Turkish, and Greek cf. Özsoy et al. (2025 [this volume]).

impact on the linguistic means that speakers use, in this case the frequency of nouns, and this is something one has to take into account when comparing such corpus data with spontaneous data or data elicited in other set-ups. This is not an issue when we compare LangSit data elicited in the same set-up within our corpus, since we used the same stimulus video throughout, and these comparisons point to systematic patterns of differences between communicative situations.

Taken together, our data hence indicates that our set-up was successful in eliciting register-differentiated data: it constituted different communicative situations that were perceived by participants as suitable for differentiated language use. In this dynamic, (in)formality and mode augmented each other in a way that the formal-written condition was located at one pole of the continuum and the informal-spoken one at the other pole.

As the findings presented in the other chapters show, such register distinctions are also evident at the level of complex syntactic patterns, for instance for clause types in English (Pashkova et al. 2025 [this volume]), left dislocation in German, Russian, and English , verb-second violations in German and OVS vs. SVO word order in Russian (for the last three examples see Bunk et al. (2025 [this volume])).

### **3.1.2 Discourse organisation**

In the domain of discourse organisation, openings are particularly interesting for distinctions between informal and formal registers, since the choice of openings is closely related to the relationship between the speaker/writer and the interlocutor.

Our corpus data is well suited to investigate this: it contains a lot of openings since the task was to retell an incident, and openings are typical for narratives in general. They are usually less frequent in WhatsApp messages between friends (which is what we chose for the informal conditions) because these are often embedded in a longer, continuous communication process. However, this was not the case in our elicitations since speakers had to act out sending a message out of the blue rather than relating to a previous conversation. Accordingly, we find enough openings even in the informal conditions.

Qualitative analyses show clear differences between the data elicited in informal versus formal conditions, and very similar patterns across spoken and written modes, in particular within productions in informal conditions.

(1) through (4) give examples from majority-German, from informal productions in spoken ((1) and (2)) and written mode ((3) and (4)).<sup>12</sup>

---

<sup>12</sup>The speaker codes are constructed as follows: The first two letters refer to the country of elicita-

- (1) Ey [name] du weißt nicht was passiert ist  
ey [name] you know not what happened has  
'Ey [name] you don't know what happened.' [DEbi71FG\_isD]
- (2) Hallo meine liebe du weißt nich was schon wieder passiert is  
hello my dear you know not what yet again happened has  
'Hello my dear, you don't know what happened yet again'  
[DEbi16FT\_isD]
- (3) Jo dicker, du weiss nicht, was ich grad erlebt hab  
yo fat.one, you know not what I just experienced have  
'Yo dude, you don't know what I just saw'  
[DEbi04FT\_iwD]
- (4) [Name]! du weißt nicht was passiert ist  
[name]! you know not what happened has  
'[Name]! you don't know what happened'  
[DEbi85FR\_iwD]

(5) through (8) illustrate the contrast to formal productions in majority-German, in spoken (5 and 6) and written mode (7 and 8):

- (5) ja guten tag DEbi10MR mein name  
yes good day DEbi10MR my name  
'Yeah, good afternoon, my name is DEbi10MR'  
[DEbi10MR\_fsD]
- (6) Hallo ja aso mein name is DEbi68FR  
hello yes so my name is DEbi68FR  
'Hello, yes, so, my name is DEbi68FR'  
[DEbi68FR\_fsD]
- (7) Sehr geehrte Damen und Herren, nachfolgend finden Sie meinen  
very honoured ladies and gentlemen following find you my  
Zeugenbericht zum Unfall vom xxx  
witness.report on.the accident of xxx.  
'Dear Madam/Sir, please find below my witness report on the accident of  
xxx.'  
[DEbi04MT\_fwD]

---

tion (DE=Germany, TU=Turkey, GR=Greece, US=USA RU=Russia, TU=Turkey); the next two to the speaker's language background (bi=bilingual, mo=monolingual) followed by speaker number, which also indicates one of two age groups (speakers numbered 1–49 are adults, those with numbers >50 are adolescents); then gender self-identification (F=female, M=male), heritage language (G=Greek, D=German, T=Turkish, R=Russian) or majority language in case of German (D) and English (E) monolinguals; communicative situation (iw = informal-written, is=informal-spoken, fw = formal-written, fs = formal-spoken); language of elicitation (D=German, E=English, G=Greek, R=Russian, T=Turkish).

- (8) Sehr geehrte Damen und Herren, meine Fallnummer ist F16  
very honoured ladies and gentlemen, my case.number is F16  
'Dear Madam/Sir, my case number is F16' [DEmo56FD\_fwD]

As the examples illustrate, we find characteristic patterns that differ for data elicited in informal vs. formal conditions.<sup>13</sup> In the informal conditions, the imagined interlocutor is a friend, and participants used an informal greeting, usually followed by a form of address and an attention getter along the lines of "You do not know/believe what just happened" that motivates the narrative. In formal conditions, participants were asked to imagine a police officer as an interlocutor, and in this condition, they used a formal greeting, often gave their name or the case number and then introduced their report.

The marked differences between these patterns indicate register distinctions at the level of discourse organisation: they confirm that the set-up was successful in eliciting distinct discourse openings in settings presented as informal vs. those presented as formal. The findings suggest that participants perceived the communicative situations as different along the parameter of (in)formality and acted accordingly.

### 3.1.3 Lexicon

Interacting with discourse organisation, we also find differences between openings in informal vs. formal conditions at the lexical level. As the examples above illustrate, participants chose colloquial terms (*jo, ey*) in informal openings and addressed the interlocutor with her/his first name, a term of endearment (e.g. *my dear*), terms of address associated with peer-group situations among young people (e.g. *Dicker*), or a nickname. In contrast, in the formal conditions they used formal greetings (*Guten Tag, Sehr geehrte Damen und Herren*) that indicate a social distance between interlocutors, and then presented themselves or the case number. An exception is *hallo*, which serves as a neutral term of greeting that is used across communicative situations.

As is discussed in Labrenz et al. (2025 [this volume]), we also find register distinctions within lexical elements at the level of discourse functions (see also Labrenz 2023): the German word *also* shows functional variation across registers that reflects the linguistic needs of the respective communicative situations. For instance, *also* occurs more often as an adverbial consecutive connector in written productions, but as a repair marker in spoken ones; the function of an elaboration

---

<sup>13</sup>For a detailed analysis of openings and closings, based on findings from all five languages, see Katsika et al. (2025 [this volume]).

or specification marker is more frequent in the formal conditions, whereas it is used for evaluation more often in the informal ones.

A similar register-differentiated distribution can be found for a graphic element, namely the three-dot sign <...>. As a graphic element, the three-dot sign is restricted to written productions. Within this mode, it shows clear functional distinctions along the parameter of (in)formality: in productions from the formal-written condition, it is used exclusively as a placeholder for lexical material, in those from the informal-written condition, it is used in discourse-pragmatic functions, for instance as a discourse-organisational device (see Labrenz et al. (2025 [this volume]) and Labrenz et al. (2022) for more details).

Another kind of graphic element we found in our data are emoji. Emoji have been shown to function as markers of informality (Siebenhaar 2020), and in line with this, in our corpus they occur exclusively in the WhatsApp messages, that is, in the informal-written condition, where they can be analysed as graphic discourse markers (Wiese & Labrenz 2021 for details).

While the three-dot sign and emoji are translinguistic elements by virtue of their graphic nature, we also found evidence for cross-linguistic lexical integration, and again, this reflected register-differentiations. Since our elicitation procedure favoured a monolingual mode (see Section 2.2 above), we did not expect to find a lot of language mixing. However, while not very frequent, we did find some evidence for this. ((9)) through ((16)) illustrate this with data from the German (9 through 11), English (12), Russian (13), Greek (14), and Turkish ((15), (16)) subcorpora.

The phenomena include code-switching between majority and heritage language (9), the use of discourse markers and interjections from international English outside the US ((9), (10)), from the heritage language in the majority language ((11), (12)), and vice versa ((13) through (16)).

- (9) und guess what        in dem Moment kamen zwei Autos  
                                   and guess what<sub>English</sub> in that moment came two cars  
                                   hintereinander   und ja kopek sokagin ortasina      gidince  
                                   behind.each.other and yes dog street in.the.middle when.goes  
                                   araba ani        fren yapti  
                                   car suddenly brake make<sub>Turkish</sub>  
                                   'And guess what, at that moment two cars came up behind each other and  
                                   yes, the car breakes suddenly as the dog runs into the middle of the road.'

[DEbi01FT\_iwD]

- (10) Alter, ich bin gerade so nach Hause gelaufen und da ist ja old.one, I have just like to home walked and there is yes dieser Parkplatz, **you know?**  
this car.park *you know*<sub>English</sub>?  
'Dude, I was just walking home and there's this car park, *you know?*'  
[DEmo53FD\_iwD]
- (11) Die riefen danach die Polizei. **Hadi** bis spaeter dicker they called afterwards the police. *OK then*<sub>Turkish</sub> until later fat.one  
'They called the police afterwards. *OK then*, see you later dude'  
[DEbi04MT\_iwD]
- (12) *Okay privat* [name], so I just saw this car accident and it was kind of weird  
'Okay *hi*<sub>Russian</sub> [name], so I just saw this car accident and it was kind of weird'  
[USbi67MR\_isE]
- (13) первая машина так затормозила что вторая **цак** и first car so braked that second *wham*<sub>German</sub> and въехала!  
drove.in  
'the first car braked so hard that the second *wham* and drove into it!'  
[DEbi18MR\_iwR]
- (14) αυτά οκέι τα λέμε **τσούς** that.was.it okay these talk *bye*<sub>German</sub>  
'that was it okay see you then *bye*'  
[DEbi07FG\_isG]
- (15) **omg** kız bugün araba kazasını gördüm !  
*omg*<sub>English</sub> girl today car accident saw  
'omg girl I saw a car accident today!'  
[USbi02FT\_iwT]
- (16) nerdeyse araba da hemen e **vollbremsen** gibi almost car also immediately uh *fully.break*<sub>German</sub> like bir şey yaptı **einfach** hemen durdu arkadaki something did *simply*<sub>German</sub> immediately stopped in.the.back araba da arabada da [laughing] önde kine çarptı car also in.the.car also [laughing] the.one.in.front hit

pek bir şey olmadı **eigentlich**  
not.much something happened *after.all*<sub>German</sub>  
'almost immediately the car did something like a *full stop* and *simply*  
stopped immediately and the car behind hit the car in front, nothing  
much happened *after all*' [DEbi52FT\_isT]

Such patterns were limited to productions from the informal conditions.<sup>14</sup> This is what would be expected for naturalistic register differentiations: in the countries we investigated, monoglossic language ideologies generally stigmatise language mixing in formal settings, and translinguistic fluidity and creativity is mostly restricted to informal contexts with friends. Hence, cross-linguistic integration can serve as an informality marker. What is interesting is that we can observe some evidence for this even though our elicitations were conducted in a generally monolingual mode. This, then, further underlines the power of our set-up to elicit register distinctions that are close to those expected in spontaneous data. The following section targets this aspect through comparison with spontaneous data.

### 3.2 Parallels with spontaneous data

### 3.2.1 Informal spontaneous productions as a point of comparison

For our comparison, we focussed on informal-spoken data since this is the kind of data that is notoriously difficult to elicit. A previous study revealed similarities of LangSit data and spontaneous data for the example of bare local NPs. In this study, Wiese & Pohle (2016) compared informal-spoken productions of adolescents from a smaller corpus of LangSit data with spontaneous spoken productions from the KiDKo corpus.<sup>15</sup> this corpus is based on self-recordings conducted by adolescents in peer-group situations. In both kinds of data, they found bare local NPs that would be ungrammatical in standard German, and these NPs followed the same syntactic and semantic patterns in the elicited LangSit data and the spontaneous Kiezdeutsch data. This is a first indication that the LangSit method is suitable to elicit naturalistic data, that is, language productions that are similar to spontaneous ones.

<sup>14</sup>In majority language use, cross-linguistic integration was generally restricted to informal productions. In heritage language use, there were also some integrations in formal productions, but these were lexical elements like nouns and verbs rather than discourse markers (and there was no code switching).

<sup>15</sup>Wiese et al. 2012; <https://hu.berlin/kidko> (last accessed: 2024-10-23).

In the present study, we followed up on these results with a broader comparison based on the informal-spoken LangSit data from the RUEG corpus and naturally occurring spontaneous data of a kind that was maximally close to the RUEG data.

The corpus of spontaneous data (the “Roy corpus”; see Roy 2022) consists of 15 not-elicited voice messages sent to friends via WhatsApp that had been donated to one of us (A. Roy), who knew all donors personally. This data set is part of a larger set of donations; the criterion for inclusion in the corpus was that the messages needed to contain some narration, given that the RUEG data was based on narrations of the accident. The speakers were young adults in Germany who had all grown up in families with German as the only family language.

To generate a comparative data set from the RUEG corpus, we extracted 15 informal-spoken productions from the German subcorpus, of the speaker group of monolingual adults in Germany, choosing speakers who were similar to the donors with respect to age/generation and education: speakers’ average year of birth was 1991 in the Roy corpus and 1992 in the RUEG subcorpus, their average age was 28 and 27 years, respectively, and the majority of speakers had a university degree (11 and 8 of the speakers, respectively). The spontaneous data was transcribed in the same way as the RUEG data. For our analyses, we mostly concentrated on the narrative parts (except for the investigation of discourse openings), that is, we ignored organisational passages that often occurred at the beginning or end in the spontaneous spoken data. Counting only the narratives from the spontaneous data, this yielded a data set with a size of 3,450 words and 327 communicative units (CUs, defined as higher-level segmentation units consisting of an “independent clause with its modifiers”, Loban 1976: 9). By comparison, the size of the RUEG extract was 2,931 words and 298 CUs.

For our investigation, we targeted all noncanonical patterns, which yielded 30 types of phenomena. Analyses revealed far-reaching similarities of the patterns in both data sets. In what follows, we describe some central findings at levels of syntax, discourse organisation, and lexicon; a detailed discussion of all results can be found in Roy (2022). Given the small size of the comparative (sub)corpora, we focus on qualitative patterns.

### **3.2.2 Syntax**

Syntactic parallels were evident in a number of word order patterns that are characteristic for informal-spoken German and would not be expected in formal and/or written registers that are usually close to the standard. Examples are

verb-first (V1) declaratives, right dislocations of modal expressions, and left dislocations in a topic-comment structure (see, for instance, Önnerfors 1997, Imo 2014 for such patterns in German). (17) through (19) give examples from both data sets. Note that the V1 examples also include ones that do not contain an ellipsis of obligatory constituents ((17b) through (17d)). This is a type considered “genuine V1” in the literature (e.g. Önnerfors 1997), which are generally less acceptable (Auer 1993) and which we found in both data sets.

(17) V1

- a. hatte einen zahlendreher drin hab mich zum dritten mal  
had a reversal.of.numbers in have me for.the third time  
mit meinem text vorgestellt  
with my text introduced  
'had a reversal of numbers, introduced myself with my text for the  
third time' [Roy, Bmo9fD\_1]

b. und is ihr halber Einkauf auf die Straße gefallen (ne)  
and is her half purchase on the road fallen QUESTION.TAG  
'and half her purchase has fallen on the road, right'  
[RUEG, DEmo43FD\_isD]

c. und sagt die Person so ja sie sind hier gar nicht bei  
and says the person like yes you[V-FORM] are here at.all not at  
dem Amt  
the department  
'and the person says, like, well, you aren't at the department here at  
all' [Roy, Bmo9fD\_1]

d. ä:h genau muss ich auch noch warten  
uh exactly must I also still wait  
'uh right, on top of everything I need to wait'  
[RUEG, DEmo39FD\_isD]

(18) Right dislocation

- a. andererseits als ich dann vorgelesen habe habe ich mir auch  
on.the.other.hand when I then read.out have have I me also  
so ein bisschen bewusst gemacht was der Inhalt eigentlich ist  
such a little.bit aware made what the content actually is  
**irgendwie**  
**somehow**  
'on the other hand when I read it out loud, I somehow became a little  
bit aware of what the content was, actually' [Roy, Bmo2fD 1]

- b. und du glaubs es nich wat [laughing] was ich hier was äh  
and you believe it not what [laughing] what I here what uh  
gesehen hab schon wieder  
seen have yet again  
'and you won't believe what uh I saw yet again'

[RUEG, DEMo43MD\_isD]

(19) Left dislocation

- a. ja aber dieses system so was so den Stoff aufzubauen das  
yes but this system like that like the stuff to.build.up that  
kommt irgendwie schwer in meinen Kopf  
comes somehow hard in my head  
'yes but this system how to build up the stuff, I somehow don't get it'  
[Roy, Bmo18mD\_1]
- b. der Mann der hat die ganze Zeit mit =n Ball mit m dem  
the man who/he has the whole time with =a ball with a  
Fußball gedribbelt  
football dribbled  
'the man, he dribbled with a ball – with a football all the time'  
[RUEG, DEMo17MD\_isD]

Of the V1-sentences that can be analysed as pronoun ellipses, we found differences with respect to the category of person: ellipsis was mostly of 1<sup>st</sup> person singular pronouns in the spontaneous data, and mostly of 3<sup>rd</sup> person singular pronouns in the elicited data. This is presumably due to the difference in the topic under discussion, which in the RUEG data was always an event where the speaker was not involved. Note that despite these quantitative differences, we found both patterns in both data sets.

### 3.2.3 Discourse organisation and lexicon

When we look at openings, we find similar patterns in both data sets, namely those prototypical for narratives, where speakers use an attention-getter and make the narrative relevant (e.g. Quasthoff 1979, Ochs & Capps 2002: Chapter 1, Gülich 2020 for German; see examples in Section 3.1.2 from the RUEG corpus). Not all messages include such an opening, though, and the frequency is higher in the RUEG data: all voice messages in the RUEG subcorpus start with an opening, compared to only 2 of the spontaneous messages. This can be attributed to the context in which the respective messages occur. As mentioned in Section 3.1.2

above, the RUEG narratives were not embedded in a larger conversation. In contrast, the spontaneous messages were often integrated. Accordingly, in the Roy corpus, speakers would presumably have felt less of a need to make their narrative relevant and could assume tellability without any comment (c.f. König 2019). (20) gives an example from the spontaneous data set where this is particular clear. In this message, the speaker talks about the behaviour of a mutual friend and starts her narrative without any introduction, beginning the message with this:

- (20) äh na ja also sie war irgendwie so voll überschwänglich  
 uh well yes so she was somehow like really effusive  
 ‘uh well, so somehow she was like really effusive’ [Roy, Bmo5fD\_4]

Both data sets alike contained a range of discourse phenomena typical of narratives, for instance markers of moral stance where speakers evaluate the narrative or a part of it (21 and 22), references to common ground ((23) and (24)), and hesitations at points of complication in the narrative ((25) and (26)).

- (21) es war nett aber auch irgendwie komisch  
 it was nice but also somehow strange  
 ‘it was nice but also strange somehow’ [evaluating a university seminar]  
 [Roy, Bmo3fD\_1]

- (22) es war ein bisschen unsinnig  
 it was a little.bit nonsensical  
 ‘it was a little bit nonsensical’ [evaluating the accident]  
 [RUEG, DEmo18FD\_isD]

- (23) genau jetzt bin ich grad mit [name] und [name] auf  
 exactly now am I just with [name] and [name] on  
 [name of an island] ist total geil  
 [name of an island] is totally cool  
 ‘right, I am with [name] and [name] on [name of an island] now and it’s  
 totally cool’ [Roy, Bmo13mD\_1]

- (24) den Parkplatz an sich den kennste ja da is ja die Hauptstraße  
 the car.park at itself that knows.you yes there is yes the main.road  
 vorne  
 in.front  
 ‘you know the car park after all, there is this main road in front of it’  
 [RUEG, DEmo17MD\_isD]

- (25) allerdings beim Aufwachen äh hatte er Lungenprobleme und although during.the waking.up uh had he lung.problems and somit auch äh wohl schwerwiegende Atemprobleme äh accordingly also uh probably serious breathing.problems uh sodass er dann noch ne Nacht äh auf der Intensivstation bleiben so.that he then still a night uh on the intensive.care.unit stay musste had.to  
'but waking up he had problems with the lungs and accordingly probably also serious breathing problems and so he needed to stay in the intensive care unit for another night' [narrative about an operation that went well but then some complications occurred afterwards] [Roy, Bmo1fD\_1]

- (26) und äh dann war da so n Hund der äh der halt auf den Ball and uh then was there such a dog that uh that just on the ball abgegangen ist on.went has  
'and uh then there was this dog that uh just went for the ball'  
[RUEG, DEmo18FD\_isD]

Again, we found the same patterns in both kinds of data. At the lexical level, we found similar patterns for instance for discourse markers in the left periphery. Qualitative analysis indicated that such discourse markers occurred at the same narrative loci in both data sets. Typical loci were, e.g., at the beginning of a narrative (cf. (27) and (28)) or of a longer explanation ((29) and (30)):

- (27) äh also die OP von [name] ist gut verlaufen  
uh so the OP(eration) of [name] has well gone  
'uh so the operation of [name] went well' [Roy, Bmo1fD\_1]
- (28) und zwar ähm wir ham uns ja grad verabschiedet  
and in.fact uh we have us yes just said.goodbye  
'and uh we just said goodbye to each other' [RUEG, DEmo17MD\_isD]
- (29) also ich wollte es kaufen damit er seine Zeit nicht vergeudet  
so I wanted it buy so.that he his time not wastes  
'So I wanted to buy it so he doesn't waste his time' [Roy, Bmo8mD\_1]

- (30) also der konnte gar nich groß auf die Straße renn  
 so he could at.all not big on the road run  
 'so he couldn't really run onto the road properly'

[RUEG, DEmo20FD\_isD]

Taken together, we hence find parallels of the RUEG corpus data with spontaneous data at levels of syntax, discourse organisation, and lexicon, and we find them for informal-spoken productions, which represent a register that is particularly difficult to elicit, given its comparably large distance from standard language.

## 4 Conclusions and outlook

In this paper, we addressed the methodological challenges one faces when aiming for data that is ecologically valid and representative of speakers' repertoires. We argued that such data is important if we want to account for actual linguistic patterns and linguistic variation. In particular we argued that we cannot confine ourselves to data elicited under conditions that favour formal language use in, e.g., a lab setting. To get a realistic picture of linguistic practices and language competence, we need to cast our web wider and include informal as well as formal communicative situations, ideally across both written and spoken modes. Including a broader range of registers also allows us to tease apart phenomena associated with informal registers and nonstandard patterns that might be due to other factors, for instance language contact, heritage language developments, or acquisitional stages in language learning. For an informed analysis of linguistic patterns, we hence need to use ecologically valid data from a range of communicative settings, and we need to do so for multilingual and monolingual speakers alike.

This is what the RUEG group set out to do, and in this paper, we described and discussed the methodology we used for creating a database for this endeavour. Since our shared enterprise included large-scale comparisons, our data needed to be controlled enough to allow for comparisons across different communicative situations, languages, speaker groups, and (contact-linguistic) contexts. Accordingly, we needed to conduct controlled elicitations. For this, we developed a method that was based on the "Language Situations" set-up where participants see a nonverbal (video or photo) stimulus showing an interesting event, for instance, a minor traffic accident, and are asked to imagine witnessing this incident and play-act telling different interlocutors about it (Wiese 2020).

We described the specifications we made to this general set-up in order to ensure homogeneity across our large-scale elicitations and to maximally support naturalistic language productions. As we have shown in this paper, our elicitation method was successful in yielding data that was (a) representative of different communicative situations, and (b) close to naturally occurring data.

First, we showed that the data we elicited in the formal vs. informal, and written vs. spoken conditions showed systematic differences that indicate register-differentiated productions. We discussed examples for such differences at levels of syntax, discourse organisation, and the lexicon. Our analyses revealed qualitative and quantitative differences along the lines expected for register distinctions.

Second, we showed that the corpus data displays similar patterns as spontaneously occurring data. We targeted informal-spoken productions because this is a type of data that is notoriously difficult to elicit. We compared RUEG corpus data of German monolingual adults with spontaneous messages by a comparable speaker group. Again, we discussed examples from syntax, discourse, and the lexicon, and showed that we find the same patterns in both data sets.

Taken together, this underlines the value of our methodology for collecting ecologically valid, register-differentiated data. It makes the RUEG corpus a suitable basis for investigations that do not just evaluate noncanonical patterns against the yardstick of a putative monolingual standard language: using the RUEG corpus, we can systematically tap into speakers' repertoires, taking into account both formal and informal registers, and doing so for multilinguals and monolinguals alike.

Our discussion pointed to a number of aspects one should keep in mind when implementing this methodology. We see three major points. First, a frequent challenge for data gathering are the resources available, in terms of time, money, and personnel. The RUEG group elicited data in 5 countries, for 5 languages, in a large range of different contact settings, from 2 age groups, and from bi- and monolingually raised speakers. This makes our corpus a rich database with a huge potential for further research (see also Shadrova et al. (2025 [this volume]) on the RUEG corpus): it provides a basis for a wealth of different comparisons that we hope will continue to be a useful resource for other researchers as well. Creating such a database was possible because of the large number of projects and researchers our group has brought together. However, applying the LangSit method is also possible with much less resources. As we argued in this paper, one major advantage is that it is less time- and cost-intensive than gathering spontaneous, naturally occurring data, but at the same time yields data that is naturalistic in the sense of showing parallels to such spontaneous data. Applying the LangSit method to a smaller number of participants, e.g., concentrating on a

specific speaker group, does not require a large amount of resources, and it can be done even in the context of a Master thesis (see Wiese 2020 for examples of this). Such smaller data sets can also be used for larger comparisons with existing data from the RUEG corpus, in particular if they were elicited with the same stimuli (which is straightforward since all RUEG resources including stimuli are open access, as described in our discussion).

A second point are the requirements on elicitation. In order to support participants' natural productions, it is important to create a relaxed atmosphere. We have argued that this is particularly true for the informal productions, since language in informal communicative situations often deviates from standard language, is thus considered "incorrect" and might be modified in a situation perceived as testing. As described, setting up an informal atmosphere to avoid this can be achieved through room decoration, elicitor's habitus (clothes, behaviour) and introductory chit-chat sessions. Nevertheless, one should expect at least some participants to have problems (play-)acting naturally, and take this into account in analysis.

A third aspect to keep in mind – kind of the other side of the coin – is the degree to which the data is controlled. On the one hand, as we discussed in this paper, the choice of the event presented in the stimulus guides speakers towards specific topics, and this is something that analyses should take into account (for instance, the topic under discussion can have an impact on the frequency of nouns vs. pronouns). On the other hand, LangSit data is, by virtue of its naturalistic nature, not as controlled as experimental data, and this means that for our analyses we face a challenge that is well known from corpus linguistics in general: some phenomena might not occur, and their absence does not prove that they are not part of a speaker's repertoire or that they are ungrammatical in a variety or a language. Hence, LangSit data is not suitable for providing negative evidence. As we have shown in this paper, its strengths lie in revealing novel and unexpected grammatical and pragmatic patterns in naturalistic language productions, in capturing a broader part of speakers' linguistic repertoires, and in allowing meaningful comparisons of bilingual and monolingual data. As such, LangSit data can reveal interesting linguistic patterns in language use and provide evidence for parallels and differences between speaker groups and (contact-linguistic) settings. If negative evidence is needed for further testing the details of such patterns, LangSit findings can be complemented by experimental data, e.g., acceptability judgments or sentence completion tasks. As several of the RUEG projects have shown, this kind of multi-method approach can be particularly fruitful for investigations into morphological and syntactic structure (see Tsehayé et al. (2025 [this volume]) and Özsoy et al. (2025 [this volume]) for examples).

We see set-ups of such controlled experiments as another domain that might benefit from the results of the methodological discussion we presented in this paper. In particular, we should aim for such experiments to also allow for register distinctions, and to do this in a way that encourages naturalistic language use. For instance, an experiment might not only be conducted in a formal lab room, but also in more informal surroundings, and stimuli should account for informal as well as formal communicative situations, e.g., through target sentences in acceptability tests that are introduced as examples from either a WhatsApp message to a friend or from a formal report. This way, our findings might inspire further set-ups for controlled experiments, with a potential for revealing novel findings on monolinguals and bilinguals alike.

Finally, as we described in this paper, our elicitations favoured a monolingual mode in both of heritage speakers' languages, given that in a first step, we wanted to target developments within their heritage and their majority language varieties. In future research, one should also include bilingual modes that encourage language mixing, especially in informal settings. Such mixing is something that is common even in societies dominated by a monolingual habitus (as is the case in the countries we investigated), and it showed up in our data even though the set-up did not support this. This points to the normality of such mixing in natural language use and should encourage us to further move away from bilingual-/monolingual dichotomies and take a more translinguistic perspective on grammatical patterns and linguistic resources. The methods we discussed in this paper lend themselves to such research, and we hope they will be a useful resource for future investigations into native grammars and linguistic diversity, taking into account the breadth of speakers' linguistic repertoires and their actual language use in different communicative situations.

## Acknowledgements

We would like to thank Maria Pohle and our fantastic and dedicated student assistants Luisa Koch, Sabine Hainsfurth, Claudia Czarniak, and Tjona Sommer. Maria Pohle was involved in the first two years of project Pd-2 and was very much missed in the subsequent years. Her contribution to the project is gratefully acknowledged. We also thank the anonymous reviewers for their helpful and very constructive suggestions, and Lea Coy for help with pre-publication formatting. The research was supported through funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit *Emerging Grammars in Language Contact Situations*, project Pd (grant number: 394844736).

## References

- Aalberse, Suzanne P., Ad Backus & Pieter Muysken (eds.). 2019. *Heritage languages: A language contact approach* (Studies in Bilingualism 58). Amsterdam: John Benjamins. DOI: 10.1075/sibil.58.
- Adli, Aria & Gregory R. Guy. 2022. Globalising the study of language variation and change: A manifesto on cross-cultural sociolinguistics. *Language and Linguistics Compass* 16(5-6). DOI: 10.1111/lnc3.12452.
- Auer, Peter. 1993. Zur Verbspitzenstellung im Gesprochenen Deutsch. *Deutsche Sprache* 23. 193–222.
- Bayram, Fatih, Tanja Kupisch, Diego Pascual Y Cabo & Jason Rothman. 2019. Terminology matters on theoretical grounds, too! Coherent grammars cannot be incomplete. *Studies in Second Language Acquisition* 41(2). 257–264. DOI: 10.1017/S0272263119000287.
- Biber, Douglas. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8(4). 243–257.
- Biber, Douglas & Susan Conrad. 2019. Analyzing linguistic features and their functions. In *Register, genre, and style*, 2nd edn., 51–84. Cambridge: Cambridge University Press. DOI: 10.1017/9781108686136.
- Bilá, Magdaléna, Alena Kačmárová & Ingrida Vaňková. 2020. The encounter of two cultural identities: The case of social deixis. *Russian Journal of Linguistics* 24(2). 344–365. DOI: 10.22363/2687-0088-2020-24-2-344-365.
- Bunk, Oliver, Shanley E. M. Allen, Sabine Zerbian, Tatiana Pashkova, Yulia Zuban & Erica Conti. 2025. Information packaging and word order dynamics in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 379–412. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775181.
- Granger, Sylviane. 2008. Learner corpora. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 259–275. Boston: De Gruyter. DOI: 10.1515/9783110213881.2.
- Gülich, Elisabeth. 2020. Alltägliches erzählen und alltägliches Erzählen. In Elisabeth Gülich, Stefan Pfänder, Carl Eduard Scheidt & Elke Schumann (eds.), *Mündliches Erzählen: Verfahren narrativer Rekonstruktion im Gespräch* (Narratology), 3–26. Boston: De Gruyter.
- Imo, Wolfgang. 2014. Elliptical structures as dialogical resources for the management of understanding. In Susanne Günthner, Wolfgang Imo & Jörg Bücker (eds.), *Grammar and dialogism: Sequential, syntactic, and prosodic patterns be-*

- tween emergence and sedimentation* (Linguistik: Impulse & Tendenzen 61), 139–176. Boston: De Gruyter.
- Katsika, Kalliopi, Annika Labrenz, Kateryna Iefremenko & Shanley E.M. Allen. 2025. Discourse openings and closings across languages in contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 527–576. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775189.
- Klotz, Martin, Rahel Gajaneh Hartz, Annika Labrenz, Anke Lüdeling & Anna Shadrova. 2024. Die RUEG-Korpora: Ein Blick auf Design, Aufbau, Infrastruktur und Nachnutzung multilingualer Forschungsdaten. *Zeitschrift für germanistische Linguistik* 52(3). 578–592. DOI: 10.1515/zgl-2024-2026.
- König, Katharina. 2019. Narratives 2.0: A multi-dimensional approach to semi-public storytelling in WhatsApp voice messages. *Journal für Medienlinguistik* 2(2). 30–59. DOI: 10.21248/jfml.2019.10.
- Kupisch, Tanja & Maria Polinsky. 2022. Language history on fast forward: Innovations in heritage languages and diachronic change. *Bilingualism: Language and Cognition* 25(1). 1–12. DOI: 10.1017/S1366728921000997.
- Kupisch, Tanja & Jason Rothman. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5). 564–582. DOI: 10.1177/1367006916654355. (27 April, 2023).
- Labrenz, Annika. 2023. Functional variation of German *also* across registers and speaker groups. *Contrastive Pragmatics* 4(2). 289–320. DOI: 10.1163/26660393-bja10077.
- Labrenz, Annika, Kateryna Iefremenko, Kalliopi Katsika, Shanley E.M. Allen, Christoph Schroeder & Heike Wiese. 2025. Dynamics of discourse markers in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 493–526. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775187.
- Labrenz, Annika, Heike Wiese, Tatiana Pashkova & Shanley E. M. Allen. 2022. The three-dot sign in language contact. *Pragmatics & Cognition* 29(2). 246–271. DOI: 10.1075/pc.21021.lab.
- Loban, Walter. 1976. *Language development: Kindergarten through grade twelve*. Tech. rep. 18. National Council of Teachers of English, 1111 Kenyon Road, Urbana, Illinois 61801 (Stock No. 26545). <https://eric.ed.gov/?id=ED128818> (27 April, 2023).

- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Anni Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Montrul, Silvina. 2015. *The acquisition of heritage languages*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139030502.
- Ochs, Elinor & Lisa Capps. 2002. *Living narrative: Creating lives in everyday storytelling*. Cambridge: Harvard University Press.
- Önnerfors, Olaf. 1997. *Verb-erst-Deklarativsätze: Grammatik und Pragmatik* (Lund germanistische Forschungen 60). Stockholm: Almqvist & Wiksell International.
- Özsoy, Onur, Vasiliki Rizou, Maria Martynova, Natalia Gagarina, Luka Szucsich & Artemis Alexiadou. 2025. Null subjects in heritage Greek, Russian and Turkish. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 179–217. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775167.
- Pashkova, Tatiana, Marlene Böttcher, Kalliopi Katsika, Sabine Zerbian & Shanley E. M. Allen. 2025. Majority English of heritage speakers. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 449–491. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775185.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Quasthoff, Uta. 1979. Die Partikeln der deutschen Sprache. In Harald Weydt (ed.), *Verzögerungsphänomene, Verknüpfungs- und Gliederungssignale in Alltagsargumentationen und Alltagserzählungen*, 39–57. Boston: De Gruyter. DOI: 10.1515/9783110863574.39.
- Rothman, Jason, Fatih Bayram, Vincent DeLuca, Grazia Di Pisa, Jon Andoni Duñabeitia, Khadij Gharibi, Jiuzhou Hao, Nadine Kolb, Maki Kubota, Tanja Kupisch, Tim Laméris, Alicia Luque, Brechje van Osch, Sergio Miguel Pereira Soares, Yanina Prystauka, Deniz Tat, Aleksandra Tomić, Toms Voits & Stefanie Wulff. 2023. Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics* 44(3), 316–329. DOI: 10.1017/S0142716422000315.

- Roy, Albrun. 2022. *Elizitierte Informalität: Ein Vergleich elizitierter und spontansprachlicher Sprachnachrichten*. Supervisor: Heike Wiese. Universität Potsdam. (MA thesis).
- Schroeder, Christoph, Kateryna Iefremenko & Mehmet Öncü. 2024. The postverbal position in heritage Turkish: A comparative perspective with a focus on non-clausal elements. In Zeynep Kalkavan-Aydin & Yazgül Şimşek (eds.), *Deutsch-Türkische Zweisprachigkeit mit besonderem Fokus auf Jugendliche [Turkish-German bilingualism with a special focus on adolescents]* (Mehrsprachigkeit [Multilingualism] 58), 109–132. Münster: Waxmann.
- Sevinç, Yeşim & Jean-Marc Dewaele. 2018. Heritage language anxiety and majority language anxiety among Turkish immigrants in the Netherlands. *International Journal of Bilingualism* 22.2. 159–179. DOI: 10.1177/1367006916661635.
- Shadrova, Anna, Martin Klotz, Rahel Gajaneh Hartz & Anke Lüdeling. 2025. Mapping the mappings and then containing them all: Quality assurance, interface modeling, and epistemology in complex corpus projects. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 69–110. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775161.
- Siebenhaar, Beat. 2020. Informalitätsmarkierung in der WhatsApp-Kommunikation. In Jannis Androutsopoulos & Florian Busch (eds.), *Register des Graphischen: Variation, Interaktion und Reflexion in der digitalen Schriftlichkeit*, 67–92. Boston: De Gruyter. DOI: 10.1515/9783110673241-004.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Tsehayé, Wintai, Rosemarie Tracy & Johanna Tausch. 2025. Inter- and intra-individual variation: How it materializes in Heritage German and why it matters. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 141–177. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775165.
- Wiese, Heike. 2013. What can new urban dialects tell us about internal language dynamics? The power of language diversity. In Werner Abraham & Elisabeth Leiss (eds.), *Dialektologie in neuem Gewand: Zu Mikro-/Varietätenlinguistik, Sprachenvergleich und Universalgrammatik* (Linguistische Berichte Sonderheft 19), 207–245. Hamburg: Buske.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.

- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagaria, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Shanley E. M. Allen, Mareike Keller & Artemis Alexiadou. 2025. Introduction: Investigating the dynamics of language contact situations. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 1–29. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775157.
- Wiese, Heike, Ulrike Freywald, Sören Schalowski & Katharina Mayr. 2012. Das KiezDeutsch-Korpus: Spontansprachliche Daten Jugendlicher aus urbanen Wohngebieten. *Deutsche Sprache* 2. 97–123. DOI: 10.37307/j.1868-775X.2012.02.02.
- Wiese, Heike & Annika Labrenz. 2021. Emoji as graphic discourse markers: Functional and positional associations in German WhatsApp messages. In Daniël Van Olmen & Jolanta Šinkūnienė (eds.), *Pragmatic markers and peripheries* (Pragmatics & Beyond New Series), 277–302. Amsterdam: John Benjamins. DOI: 10.1075/pbns.325.10wie.
- Wiese, Heike & Hans Georg Müller. 2018. The hidden life of V3: An overlooked word order variant on verb-second. In Mailin Antomo & Sonja Müller (eds.), *Non-canonical verb positioning in main clauses* (Linguistische Berichte Sonderheft 25), 202–223. Hamburg: Buske.
- Wiese, Heike & Maria Pohle. 2016. „Ich geh Kino“ oder „... ins Kino“? Gebrauchsrestriktionen nichtkanonischer Lokalangaben. *Zeitschrift für Sprachwissenschaft* 35(2). 171–216. DOI: 10.1515/zfs-2016-0012.
- Wiese, Heike, Horst Simon, Christian Zimmer & Kathleen Schumann. 2017. German in Namibia: A vital speech community and its multilingual dynamics. Péter Maitz & Craig A. Volker (eds.). 221–245.
- Zimmer, Christian, Heike Wiese, Horst J. Simon, Marianne Zappen-Thomson, Yannic Bracke, Britta Stuhl & Thomas Schmidt. 2020. Das Korpus Deutsch in Namibia (DNam): Eine Ressource für die Kontakt-, Variations- und Soziolinguistik. *Deutsche Sprache* 48(3). 210–232. <https://nbn-resolving.org/urn:nbn:de:bsz:mh39-100570> (27 April, 2023).



# Chapter 3

## Mapping the mappings and then containing them all: Quality assurance, interface modeling, and epistemology in complex corpus projects

Anna Shadrova<sup>a</sup>, Martin Klotz<sup>a</sup>, Rahel Gajaneh Hartz<sup>a</sup> & Anke Lüdeling<sup>a</sup>

<sup>a</sup>Humboldt-Universität zu Berlin

Building on our experience with the structured and deeply annotated RUEG corpus, this contribution summarizes methodological aspects of corpus modeling, compilation and analysis. We argue that corpus data – as opposed to experimental data – is characterized primarily by its open-endedness, and thus corpus-linguistic work is characterized primarily by the task of structuring and ontologizing data on a range of linguistic levels based on a range of linguistic models and with the aid of a range of technical models. As the unification of these diverse models requires interface definitions, it follows that annotation, modeling, analysis, and the epistemological embedding of categories and analytical processes must go hand in hand. For any large-scale corpus project, this complexity requires a division of labor that creates distributed and decentralized knowledge, which we understand as more interfaces requiring definitions. We conclude that the epistemic gains that can be procured from corpus projects go beyond the actual corpus resource and beyond discussions of technical representations, and that it would benefit the field to synthesize this type of modeling knowledge more productively for a positive impact on future projects as well as a better understanding of the emergent properties of situated language production.



## 1 Introduction

This paper summarizes aspects of the corpus-methodological research that took place within the Research Unit *Emerging Grammars* (RUEG) in 2018–2024 and the compilation of the corpus it created from elicited data, which we will refer to as the RUEG corpus (Lüdeling et al. 2024, Klotz et al. 2024). RUEG collected and analyzed data from heritage speakers of Greek, Russian, and Turkish in Germany and the US, German in the US, and Kurmanji (Kurdish) in Turkey, with all participants elicited in both of their languages; and from monolingually-raised participants in the countries where those heritage languages represent the majority language (e.g. Turkish in Turkey, Greek in Greece etc.). This was done with the purpose of comparing emergent patterns on various linguistic levels fostered by the speaker-internal language contact situation in the minds of multilingual individuals, and in correlation with situational aspects such as formality and modality. While challenges and constraints occur as concrete and specific in any research project, many challenges we have encountered arise from a generalizable requirement of mapping models and interfaces. We believe this requirement should receive more attention for the benefit of future corpus-related work. This paper thus is not primarily concerned with a description of the corpus, neither is it about technical standards of corpus compilation or data architectures, or about specific linguistic phenomena.<sup>1</sup> Rather, it presents a conceptual reflection of the complexity and interdependence of modeling decisions that this type of research requires and that have not found much attention in the debate until now.<sup>2</sup>

RUEG and its corpus are complex units created in a strongly collaborative effort by contributors from a variety of linguistic subfields and levels of experience. Combining data in six languages elicited in five countries from over twenty speaker cohorts,<sup>3</sup> and in recording spoken and written data based on carefully

---

<sup>1</sup>For a more in-depth discussion of either of those aspects, please view Wiese (2020), the continuously updated corpus documentation under <https://korpling.german.hu-berlin.de/rueg-docs/latest/>, and the other chapters in this volume.

<sup>2</sup>As a reflection of this complexity and interdependence as well as the current lack of methodological and meta-theoretical integration, a certain degree of oscillation between the concrete examples from the RUEG process and corpus and the abstraction of principles of corpus modeling could not be avoided. This may appeal more or less in different subsections to readers from different specializations, e.g. annotators, technicians, or linguistic experts. Readers are encouraged to skip ahead from subsections that provide too many low-level details for their interest.

<sup>3</sup>Adolescent/adult monolingual/bilingual speakers in each country plus some trilinguals, see more in Section 2.

prepared prompt material, the RUEG corpus is one of the most complexly designed corpus projects to exist to date. This is a reflection of the RUEG project's research goals, as it brings together a speaker-centric with a variety- or linguacentric perspective on language data (describing speaker behavior vs. language patterns), while also attempting a comparison of phenomena on various linguistic levels which are not perfectly identical between the chosen languages. It further compares linguistic structures in different modalities and degrees of formality and aims to find and map emergent patterns in multilingual contexts, i.e., patterns that are influenced by language contact and individual variability, and that may have not yet been fully linguistically described for the respective languages. Moreover, RUEG intends both quantitative and a qualitative descriptions and thus also interfaces with diverse analytical frameworks.

As visualized in Figure 1, the corpus compilation process, the project structure and the data architecture reflect the complexity of RUEG's research agenda. This is partially due to the phenomena of interest, which require deep linguistic annotation that cannot be performed automatically at adequate quality. Data therefore has to be annotated manually, collaboratively, and additively. In this, some layers of annotation and pre-processing depend on others, therefore requiring an iterative workflow and many decisions around the adequacy and usefulness of the chosen categorization, i.e., the linguistic model, at each level.

In this contribution, we argue that the central challenge of the compilation, infrastructure provision, maintenance, and analysis of task-based corpora like the RUEG corpus lies in the mapping of different models and their interfaces. "Modeling" here refers to the process of explicitly defining the object or process of interest in terms of a selected theoretical, conceptual, methodological, or other tradition or framework (which are also models themselves).<sup>4</sup> "Interface mapping" refers to the process of identifying and defining the connective properties and elements between two or more of those modeled objects. We will discuss these notions in greater depth in Section 3.

With this, we offer a perspective on corpora that does not primarily center the result of their compilation, namely the corpus itself as a reusable resource, but one that views the process of thought and construction with the interdependence of all steps in its own epistemic right. The corpus resource is then just a view of

---

<sup>4</sup>We are aware that different disciplines and areas understand modeling in different ways and that this is only one of the possible meanings of modeling (see for example Stachowiak 1973, White 1973) and more generally the debates in the JLM (<https://jlm.ipipan.waw.pl/index.php/JLM>). We will not be able to discuss other approaches and philosophies of modeling here, and will instead provide definitions of our understanding where necessary.

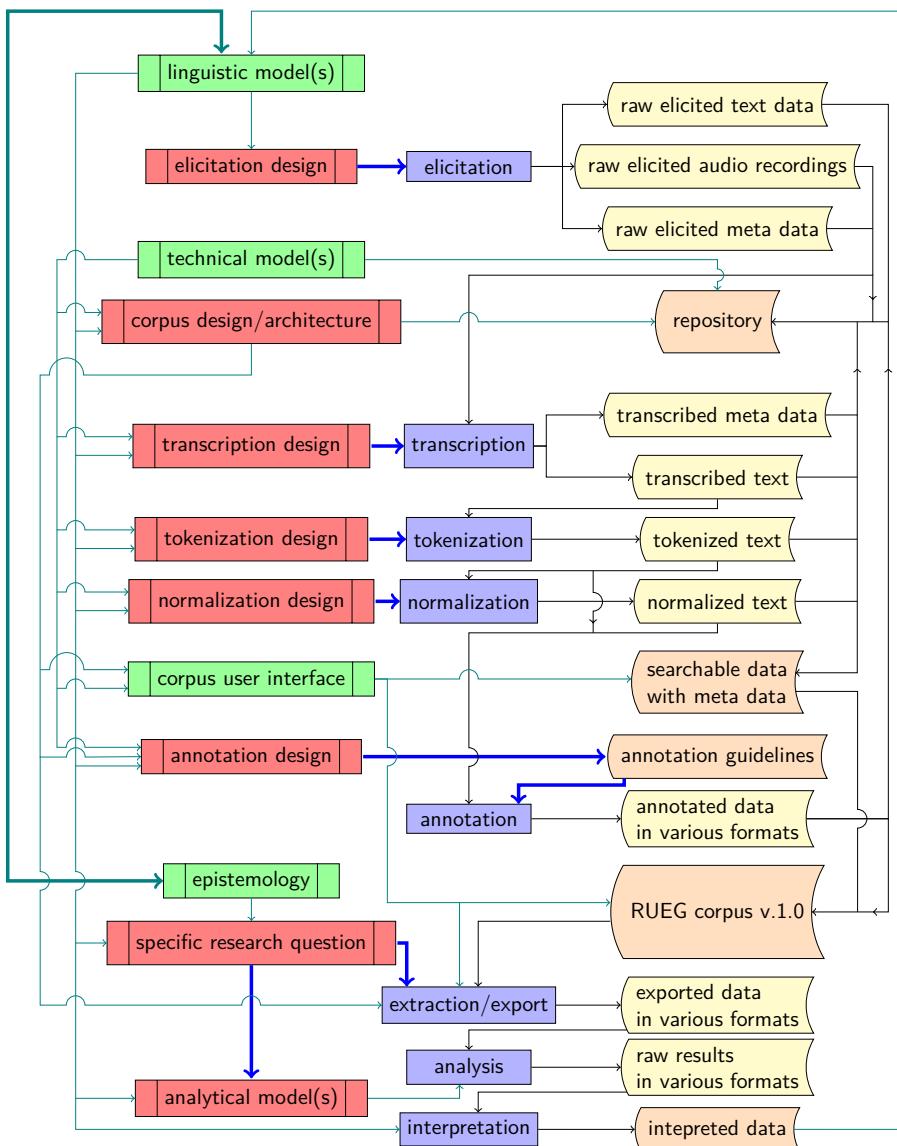


Figure 1: Processes involved in the compilation of the RUEG corpus. Arrow colors signify different types of relations: teal – influences, limits, or defines access to; blue – defines; black – results in or is passed over to. Colors and shapes signify type of process, model, or framework: boxes: green – underlying, project-independent models or infrastructures; red – design choices made within the project; violet – processes performed by project members; book-shaped: yellow – process products; orange – stable, reusable products.

### *3 Mapping the mappings and then containing them all*

one stage in its development. Since the RUEG corpus was developed under significant time constraints relative to its complexity, it also provides insight into the challenge of handling subprocesses in parallel in spite of their mutual interdependence. This simultaneity poses specific demands on the technical realization as well as high interactional and communicative pressure on the collaborators, especially since different types of models, processes, and objects (marked in different colors in Figure 1) are often handled by different project members. In RUEG, this is further complicated by the groups' decentralized organisation across several locations in Germany.

The text is structured as follows. We will first provide an overview of the corpus design, the elicited material, and the core research questions in Section 2. We will then introduce the notion of linguistic modeling and discuss how different models interlock in an interdisciplinary project and why that poses a specific problem in corpus data (Section 3), followed by a discussion of some of the modeling processes and common challenges in Section 4. Specifically, we will discuss differences between sequential (what happens in time) and structural aspects (which choices depend on previous ones) of the model; how to map diverse and partially incompatible models; and how to map models that are not yet fully defined. Wherever useful, we will illustrate our points with examples from the RUEG corpus and context, although we believe that the problems are more structural and relevant to all of corpus linguistics and especially projects building and analyzing task-based corpora.

We offer five central theses in this paper:

1. Since corpus data is not experimental in nature, most of the epistemological weight of any argument derived from corpus data lies in the post-elicitation modeling, selection, and analysis.
2. Modeling decisions are interface definitions connecting different models. Models exist on the technical side (such as data structures or database architectures) and on the linguistic side (such as different linguistic models projected onto the data through annotation). In a collaborative environment, each collaborator's understanding is also a (usually implicit) model. With a distribution of labor between e.g. project managers, annotators, technical experts, statistical analysts, and so on, corpus projects create distributed knowledge, which necessitates interface definitions between collaborators in order to remain consistent.
3. Interface definitions are mappings of the categories of different models onto one another and definitions of their potential interactions. They are

not usually obvious or available a priori for sufficiently diverse models, e.g. technical infrastructure and linguistic content, or linguistic theory at the lexical vs. the syntactic level. The number of interfaces is at least a quadratic function of the number of collaborators, which means that raising the number of collaborators shifts the balance of a corpus project towards more required mapping, i.e., communicative overhead. It is questionable whether these communicative demands can be satisfied within the constraints of common research infrastructure for very large corpus projects.

4. Linguistic depth is therefore most approachable through small to medium-sized, well-controlled, and manually annotated corpora and within infrastructures that allow for as many of the collaborators as possible to participate in communication and analytical decision-making at each step.
5. The corpus and scientific publications based on its data are not a comprehensive representation of the set of epistemes that arise from corpus projects. Rather, procedural aspects, including linguistic and technical modeling decisions, represent knowledge in their own epistemic right. This type of knowledge cannot be passed on in full through continued access and maintenance of a corpus-database or through corpus-specific documentation. Rather, the field needs more reflection on generalizable corpus-methodological epistemes beyond the discussion of technical standards.

While task-based corpus elicitations limit the active potential of participant responses, they do not effectively limit the variability or complexity of the data. Rather, even within this limitation, variability remains massive and the emergent properties of language are omnipresent. Corpus linguistics has the unique benefit of modeling the interaction of the full emergent quality of situated language production, but can only do so under epistemological clarity and demarcation of all relevant categories and interrelations. This requires a strong focus on the distillation and synthesis of models that are applicable to the specific case from a large number of available ones, which involves a process of idiosyncratic expertise development. It cannot (helpfully) be rushed and should not be viewed solely in terms of publications or resources generated.

## 2 The RUEG corpus

Within the RUEG project, language contact phenomena are investigated from different viewpoints. One perspective is a focus on possible influences of different heritage languages on the same majority language. Another one is a focus on possible influences of different majority languages on the same heritage language. A list of all the projects, their research topics and their perspectives can be found in chapter (Wiese, Allen, et al. 2025 [this volume]).

A heritage language is a language acquired naturally during first language acquisition that is different from the majority language spoken by the surrounding society (Pascual y Cabo & Rothman 2012, Montrul 2015). Literature points out that differences can be observed in heritage speakers compared to monolingual speakers regarding the heritage language as well as the majority language (Rothman 2009, Polinsky 2018). Differences of this kind, though, are not undisputed (Wiese et al. 2022, Özsoy & Blum 2023). Definitions for the terms *heritage language*, *heritage speaker* and *monolingual speaker* are also provided in the introductory chapter (Wiese, Allen, et al. 2025 [this volume]).

The design of the RUEG corpus is driven by several general research questions investigated by multiple research groups from different cities in Germany in several overlapping phases of data elicitation, annotation, correction, and statistical analysis.

To obtain data appropriate for the research questions, an elicitation procedure was defined (Wiese 2020), for which participants are presented with a short stimulus video showing a series of events that lead to a minor traffic accident, which they are then asked to report in four different situations defined across two modes (spoken vs. written) and two degrees of formality (formal vs. informal). The setup acknowledges the sensitivity of language use to the production situation (in order to assess register knowledge, see e. g. Lüdeling et al. 2022). For differentiating findings specific for heritage language from general trends within a majority language of a country, speakers with no heritage language background were included as well.<sup>5</sup> To also address questions of speaker generation, elicitations spanned across two age groups, adolescents and adults. Additional knowledge on the participants' socioeconomic background, language biography and habits, and personality was collected with a digital questionnaire. Data collection took place over several months in five countries and for six languages (see Table 1).

---

<sup>5</sup>For simplicity, we will refer to this speaker cohort as monolinguals, although we are aware that most speakers are not truly monolingual in an urban and digitized environment.

Table 1: Languages elicited and country of elicitation by subcorpus

| Language    | Country |        |        |        |     |
|-------------|---------|--------|--------|--------|-----|
|             | Germany | Greece | Russia | Turkey | USA |
| Core corpus |         |        |        |        |     |
| German      | ✓       |        |        |        | ✓   |
| Greek       | ✓       | ✓      |        |        | ✓   |
| Russian     | ✓       |        | ✓      |        | ✓   |
| Turkish     | ✓       |        |        | ✓      | ✓   |
| English     |         |        |        |        | ✓   |
| Extensions  |         |        |        |        |     |
| Kurmanji    |         |        |        | ✓      |     |

We chose a multi-layer architecture (Zeldes 2018) because of its extensibility and flexibility with regard to number and types of annotation layers. Additionally, such an architecture allows to enrich the corpus with extensive metadata obtained from the participant questionnaire. The research questions require linguistic analysis of phenomena from different linguistic domains at different degrees of granularity. Spoken data is carefully transcribed, written data is automatically extracted and reformatted (Schmidt & Wörner 2014, Boersma 2001). Then, both types of data are passed on to further processing. We follow the tradition from historical linguistics to call the layer that is as close as possible to the original the *diplomatic* layer. As a baseline for follow-up annotations, the dipl layer is then normalized to canonicalized tokens, to which lexical and functional annotations (lemma, part of speech (PoS), morphological features), syntactic annotations (dependencies and constituency trees), or annotations regarding information status and referents are applied. A subgroup directly annotates prosodic features for the diplomatic tokens.

A centralized data repository is used to store the annotated data and their versions throughout the entire project. The repository is interfaced with automatic preprocessing pipelines for enriching the data with new annotations, testing it against several criteria of consistency and integrity, and to deploy a combined resource for search and visualization under a common technical model (Zipser & Romary 2010, Krause & Zeldes 2014, Krause & Klotz 2023). Also, the repository interface served as knowledge base for generating the future corpus documentation. It furthermore served for the coordination of the complex annotation

process handled by multiple parties with different annotation tasks and scopes (for an illustration see Figure 1). The infrastructure around the RUEG corpus is designed to provide easy process for further iterations of annotation, documentation, investigation, and publication, i.e., linguistic research in general.

As previously mentioned, the RUEG corpus and its data are the outcome of a coordinated collaborative effort. Multiple research groups have annotated and continue to annotate the raw data in parallel, basing their work in specific expertise on different subfields of linguistics, which we consider the most interesting and challenging aspect of this endeavour. Annotations are designed and applied relative to each respective research question. For this, there are regular conceptual exchange meetings, after which annotation groups work mostly autonomously. Annotation processes are frequently executed by student assistants and PhD students in exchange, but not usually in direct collaboration, with postdoctoral researchers and professors. Annotation tasks can focus on a specific speaker group or a single language, but can also range across multiple languages. Depending on the type of annotation and its structure, different annotation tools are involved in an annotation workflow, which results in variable formats (representations).

### **3 Data, models, and interfaces**

On the most abstract level, a corpus can be described as a collection of data. But what is data?

On the right side of Figure 1 we have visualized in orange a range of stable, reusable resources that have resulted from the work of the RUEG project. This most obviously includes the final 1.0 version of the RUEG corpus, but also its predecessors without manual annotations, a repository storing all materials and data in their various stages, annotation guidelines that can be reused for future research, and of course a number of publications that will serve the development of heritage language research and related fields. Typically, the corpus itself would be referred to as “data”, or as processed data, and the elicited material, the audio recordings, typed text, questionnaires and so on, as “raw data”.

Beyond this, RUEG has produced many intermediate stage products, visualized in yellow in Figure 1, which we will refer to as process products. This includes each of the transcribed, tokenized, and normalized text, and a range of annotation styles or tiers (e.g. syntactic, phonetic, or discourse-level), as well as data that was exported from the corpus and entered into other analytical processors, such as statistics scripts, and the results thereof. It is also common for researchers

to regroup and filter data to their convenience outside of the provided corpus interface and add on-the-fly annotations during analysis. Data interpretation (visualization, etc.) may therefore be based on only a subset or an extension of the data actually provided in RUEG 1.0, and there may be many more versions of the RUEG corpus on individual researchers' computers that the authors are not aware of.

At each step of the process from the top left to the bottom right, "data" presents as a different type of object. In Figure 1, this can be read from the black arrows. The process of transcription, for instance, takes the raw elicited audio recordings and the raw elicited metadata, and yields transcribed metadata and transcribed text as a result. This *result* of one process is then passed on to the next process of tokenization and normalization *as data*. Similarly, the process of annotation takes *tokenized text data* that may or may not also be normalized, i.e., be the result of one or two processes, as data, and yields *annotated data* as a result. Finally, the researcher working on the final analysis will take the annotated data, or more generally the *corpus data* as an input to their analytical model. This results in new, e.g. quantitative *data* that is then interpreted.

At this point, the researcher is working with the result of a range of processes which each involve a number of linguistically relevant and often complexly interdependent choices. In Figure 1, these are summarized in the red and green boxes representing project-specific and external sets of views and choices respectively. These sets of views and choices are models.

A model in our sense is an abstract description of a phenomenon (an object or a process) that contains, i.e., defines, the elements and parameters relevant to the context of use and their interrelations. For example, in the RUEG context, the transcription design contains information on how to adequately transpose phonetic to graphemic information. But there is no information on how to adequately transcribe archaic handwriting to modern graphemes, because this is irrelevant in the RUEG context. The transcription design is a model, not a 1:1 representation of the process. It does not and cannot reflect or predict every single mapping from phonetic quality to graphemic representation. Rather, it is itself a product of a process of communication and mediation between several participants (frequently more experienced researchers instructing student assistants on how to transcribe, and transcribers bringing questions on unresolved cases). This process is informed by other models, for instance conventions of orthography, which in turn rely on etymology, language policy, etc.

In this way, all research is a reflection of a range of processes that typically result in a model (the result of a meeting; a paper; an annotation guideline; a cor-

### 3 Mapping the mappings and then containing them all

pus; a new theory; etc.) and are also informed by other models, which themselves are results of modeling processes, and so on.<sup>6</sup>

Any research process, and any type of data, in this way is explicitly or implicitly influenced by the models available to a field (paradigms, theories, hypotheses, explanations, etc.). This also means that data is never neutral, it is always a model that reflects underlying (presumed) ontologies.

All research further requires a definition of the relevant interfaces of all objects and processes involved. Any well-designed experiment, for instance, is defined in terms of how the expected or unexpected result would fit with the theoretical framework; and any function of a programming language is defined in terms of which data format it requires in order to be correctly executed. Interface definitions are hence simply descriptions of how the different elements (processes, objects, phenomena, etc.) fit with one another. Interface definitions are also models, i.e., functional abstractions of complex and fuzzy real-world phenomena into manageable sets of properties.

These processes are fundamental to all research, corpus-linguistic or otherwise. There are, however, aspects of corpus-linguistic research of the kind performed within the context of RUEG that pose specific challenges. These are, in no particular order:

- a. the number of interfacing elements and processes – as Figure 1 shows, dozens of relations occur even in this synthesized and simplified description which does not reflect the iterations of subprocesses required to arrive at a higher-level concept such as *normalization*;
- b. the path dependence or mutual interdependence of some of the processes involved (see Section 4.2);
- c. the number of people involved, who each carry their own understanding of the material, i.e., their own model, effectively creating an interface with every person, every general and every project-specific model involved;
- d. the open-ended character and fundamental abstractness of language;
- e. the fact that corpora contain *naturalistic* language, which as an emergent phenomenon results in more than the sum of its parts through manifold and complex interrelations. This is independent of whether the context of utterance is also considered realistic or natural.

We will discuss the latter two points in detail below.

---

<sup>6</sup>Leonelli (2019) shows this very convincingly at the example of plant phenotyping in an interdisciplinary and digitized research setting.

### 3.1 The open-endedness of corpus data

Modern day linguistics frequently attempts a degree of objective categorization as it is traditionally sought in the natural sciences and computational subjects: linguistic phenomena are described to be *observed* in their context similar to observations of, say, the leaves of a plant.<sup>7</sup> However, unlike the leaves of a plant, we cannot touch, store away, cut, rotate, etc. language without a priori knowledge.<sup>8</sup> The multiplicity and variability of linguistic models in use shows that this can take very different shapes. The pluralicity of ways to view the elements of a language is a central issue of corpus linguistics, which at many stages is concerned precisely with the demarcation of linguistic units at different levels of abstraction and granularity. On a very basic level, this is the case in tokenization, i.e., the separation of character or audio strings into meaningful units (like phones, letters, words, punctuation signs, etc.).<sup>9</sup>

But it is equally the case in all annotation decisions (*does a certain string represent a case of category X?, in what shape does phenomenon Y occur in the data?*). Even for the classification of a seemingly delimited phenomenon like PoS, we find numerous models from various grammatical traditions, and several more from a more pragmatic computational perspective (for a discussion see e. g. Robins 1986, Knobloch & Schaefer 2000, Atwell 2008, Petrov et al. 2012).

Since there is no single, theory-neutral, “obviously natural” categorization, i.e., no clear-cut, extra-linguistically definable way of determining what constitutes a unit at any layer of linguistic granularity, a demarcation of meaningful units relative to a research question is of central importance to any corpus-based project. A project interested in discourse markers, for example, first has to decide what

---

<sup>7</sup>This is not to imply that the STEM fields always succeed in reaching objective conclusions, as debates around the replication crisis show. Recent research in biology also shows that results from the same dataset majorly depend on subjective choices even in seemingly objective analytical frameworks like inferential statistics (Gould et al. 2025).

<sup>8</sup>A biologist arriving on an island with previously unrecorded plants can go to a plant and touch, rotate, etc. its leaves, unless the leaves have some extremely unusual properties. A linguist hearing upon arrival on a remote island a previously unrecorded language has to put some effort in and meet informants to begin identifying even the basic building blocks of the unknown language, even if the language is not particularly unusual.

<sup>9</sup>While it may seem obvious where one word begins and another ends in many modern European languages, the process of tokenization actually requires many design choices regarding for instance the status of compounds, neologisms, the role of inflections, corrections and repairs. For languages that are not separated into orthographic words by convention (such as non-alphabetic or non-existent writing systems; agglutinative and polysynthetic languages; signed languages; historical languages prior to orthographic normalization; etc.), even more decisions have to be made (Bauer 2000, Schmid 2008). Even in European languages, we find phenomena like clitics or amalgamations that add complexity.

### *3 Mapping the mappings and then containing them all*

counts as a discourse marker, and then identify all occurrences in the data. This may seem obvious, but in our experience is not usually *a priori* defined in actual corpus projects and may not be *a priori* definable due to the inherent emergentist quality of the data.

Once this identification is performed in the form of annotation, the decisions, i.e., models, guiding said annotation will transcend through any further research that uses those annotations. A compiled and annotated corpus then will contain many modeling decisions upon which all further analysis depends, i.e., path-dependences in the data. Some categories may appear as though they fell into place naturally, but this only means that the categorization was performed based on implicit understanding, i.e., a hidden model. For annotators, this impression usually quickly dissipates when they are confronted with the full variability of their data. At this step, the general practice of dividing labor between annotators and model designers can become challenging. This is because, unless both parties participate in frequent in-depth discussion of the data, their implicit models of the data may easily diverge. It is not uncommon for those divergences to go unnoticed, only causing problems at later stages. Between annotators, this can be alleviated via agreement measures and guidelines, but when different models need to be fitted with one another, other strategies are required.

Why does corpus data give rise to divergent models so easily? We believe that this may be due to a mistaken perspective of *elicitation as quasi-experiment*: when analyzing corpus data, or when designing a task-based corpus, researchers often implicitly view it as a kind of experiment in which speakers are asked to perform task *x*, and the data that is extracted from the corpus is their response. However, that is not an accurate model. An experiment is the recording of a process in an *a priori* artificially constrained world, in which a participant can only respond within a predetermined response space (a response time; an eye movement pattern; a Likert scale; one out of a limited number of specified behaviors). The response may be multimodal (response time and Likert scale), but each response *space*, i.e., the number of options, is limited, hence also limiting the number of combinations. If a participant does not provide responses of the envisioned kind (e.g., if they glance away, or fall asleep and produce extremely long reaction times), they are excluded from the analysis. This also means that, in an experiment, the plausible kinds of analyses are embedded in the design. Experimental data on a Likert scale cannot be re-interpreted as eyetracking data instead. Rather, once the experiment is complete, the data exists in the expected format and can be passed on to the preconceived analytical framework. This means that in an experiment, the interface between the hypothesis/the theoretical model,

the experiment design, and the outcome variable are all modeled during design, namely in constraining the response space in a finite and specific way.

This is not to say that all experimental data is unproblematic in its design or interpretation, or that experimental designs always succeed at isolating their factors as desired (Dewaele 2019, Gozli 2019). Rather, the point is that the data elicited from an experiment exists in a world closed and discrete by design, and hence the objects of interest for analysis and interpretation are known by design: as soon as the experimental data comes in, the results are determined. The data is delimited and speaks for itself (within the realms of the experiment – in experimental data, production roughly equals result with a deterministic transformation through a method of analysis).

In corpus data, we find more *qualitative* variability between responses, i.e., speakers responding to the task in cooperative, acceptable, and interpretable ways, but their responses lying outside our expectations in elicitation. Corpus data is continuous prior to analysis and open-ended (Moisl 2009). Continuous here refers to its quality of not naturally falling into categories, linguistic or otherwise. All demarcations of categories have to be performed post-hoc, including demarcation of words or sounds from the audio or string. This is even true of typed text, where the demarcation of words seems to be defined by the speaker, but will often be changed for normalization. It is open-ended in several ways:

- a. Corpus data is not constrained to a limited response space – in responding to a task like the one prompted in RUEG, participants have a universe of choices at all linguistic levels.
- b. Speakers make use of this freedom by responding in astoundingly variable ways, particularly in the lexical domain (Shadrova 2025). This shows even in the communicatively narrow space of the RUEG corpus, which means that a task-based limitation does not necessarily result in greater predictability of the lexical forms that will occur in the corpus.<sup>10</sup>
- c. Beyond choosing different expressions for the same communicative purpose, speakers also frequently diverge in their interpretations of the situation, the task, or the communicative purpose and engage in playacting, change the genre, invent content, and so on.
- d. Corpus data is not analytically deterministic and cannot *as such* be handed over to an analytical framework. It needs to always be filtered and rearranged before it becomes quantifiable.

---

<sup>10</sup>For instance, speakers in the German subcorpora find over 600 verbs to refer to the events and actions taking place in the 40 second video, see Shadrova et al. in preparation.

- e. This process of filtering and rearranging itself is open-ended, as it depends on a multiplicity of analytical decisions and framework specifications.
- f. Corpus data is also open-ended in a historical or dynamic sense, i.e., with respect to the emergence of structures that cannot be predicted because they are idiosyncratic, subject to ongoing language change, or as of yet undescribed. The latter is especially interesting in the RUEG context because the project aims at the identification of shifts in form and function, for which the models in use must be able to handle unknown forms or forms in new functions.

The complexity and open-endedness of the corpus linguistic analysis can hence be summarized as (1) massive variability and (2) interconnectivity with (3) an unpredictable number of productive interrelations between means of expressions and (4) concepts describing those means of expression (5) in a historically changing, dialectic, and non-deterministic space which (6) can be passed on to an open class of analytical models. In short, corpus data never speaks for itself, because it is *a priori* undelimited and analytically non-deterministic.

One may try to avoid the emergent space by filtering the data for a specific category. For instance, if interested in expressions of directionality, one might query for the occurrences of directional prepositional phrases or selected lexemes or morphemes, provided the corpus is equipped with the relevant annotations. But the full variability and complexity of how speakers express directionality (lexically/adverbially, constructionally, pragmatically, etc.) is *still present*, and raises the question of what should be considered in the conceptual and analytical model; it cannot be avoided in the linguistic discussion of the data. Filtering a corpus for a specific type of information does not undo the complexity and emergence that speakers bring to the data. At the very least, the analytical model has to specify the cases it excludes, adding more interrelations and descriptive complexity. In this way, corpus data is open-ended with respect to the factor combinations that emerge from it, because an unlimited number of demarcations and interpretations can be combined with one another, resulting in a perplexing combinatorial space even in small data.

The major challenge of corpus linguistics then is to delineate meaningful constructs in corpus data and to model their interrelations in an epistemologically convincing way relative to the linguistic and analytical frameworks employed. This makes corpus linguistics a *constructionist*, rather than an *analytical* field until rather late stages of the research process.<sup>11</sup> Only after all categories and in-

---

<sup>11</sup>This distinction might be reminiscent of the discussion about corpus-driven, theory-driven, or corpus-based approaches (Biber 2015). Framed in these terms, our approach may be labelled as both theory-driven and corpus-based.

terrelations have been constructed (e.g. annotated), can the analytical process of separating effects truly begin. It is of course this very work that provides insight into naturalistic language production in context, and frequently yields fascinating and surprising results. In order to do so, clear epistemologies and problem definitions are required. We will now present a range of ways in which this work was designed and performed in RUEG.

## **4 Models and interfaces in RUEG**

We have mentioned that a complex project like RUEG consistently requires the mapping of various models into a common space. This process is defined by a number of factors, including the number of models to fit; their mutual compatibility, i.e., freedom of contradiction in representation, theoretical assumptions, or epistemologies; their interdependence; and their certainty, well-definedness or degree of ambiguity. Better understanding of each model contributes to ease of mapping. However, in reality, models are not always perfectly defined in every detail, and some happen to be partially incompatible. In the following sections, we will discuss some of the issues arising from the combinations of different types of models. One particularly interesting case in the RUEG context is the mapping of unknown models, i.e., the description of emergent phenomena which can only be described as a research outcome, rather than mapped onto the data *a priori*. Of course this is the most complex case, since it maps open-ended data to an open-ended modeling space at high uncertainty. Unfortunately, we do not have the space to discuss this case in detail and will only mention few aspects in Section 4.4. We will provide a more in-depth discussion in Klotz & Shadrova (in preparation).

### **4.1 Mapping different models into one**

We will first discuss the mapping of a number of models into one annotation representation at the example of the realization of aspectual meaning in heritage vs. monolingual speakers of Greek, Russian, and Turkish. The linguistic details are discussed in contribution (Wiese, Allen, et al. 2025 [this volume]). For our purposes, it suffices to understand that aspectual meaning can be encoded in a range of linguistic realizations, which include morphosyntactic (aspectual verb markers), syntactic (periphrastic constructions; subject-drop can play a role), and lexical realizations (verb semantics/Aktionsart; adverbs). These features are diversely distributed between the languages represented in RUEG.

The RUEG perspective predicts systematic shifts in language contact situations, i.e., systematic differences between monolingual speakers and heritage speakers outside of the majority language environment. For example, heritage speakers of different languages might move away from a morphosyntactic realization of aspect and towards a periphrastic or adverbial realization. Of course neither of those categories are identically distributed in the three languages in question in monolinguals and heritage speakers. This means that all aspect realizations in each language need to be annotated. Those are typically well-described in linguistic theory, although the descriptions may not be fully exhaustive due to the open-ended character of language. If there is truly a systematic shift between speaker groups, the data might challenge the existing linguistic models and require adaptations for annotation (see also Section 4.4). Finally, the hypothesis predicts a systematic shift *in a specific direction* for all three languages. To detect this, a comparability between the language-specific descriptions is required. What does it mean, for instance, for Greek aspect to shift from one distribution to another, and (how) can this be quantified in comparison with Russian aspect realization distribution?

Language-specific models of morphosyntactic description will frequently not only differ by the contents of specific categories (e.g. different aspect markers), but also in relevant theoretical assumptions (e.g. UG-based/minimalist grammar vs. usage-based/construction grammar), which may result in drastically different annotations. Will annotation categories for both, or all three, languages hence mean the same, and will the different tag distributions provide conceptual comparability?<sup>12</sup>

The above is only an example of the many mappings required in the annotation of aspect in a corpus like RUEG, as is visualized in Figure 2 in a simplified manner. A comprehensive map of all interrelations would require a reflection of all entities contained by the larger boxes and how they interact with one another. We will not attempt this here. Even in this simplified manner, the modeling requirements for the annotation of a single phenomenon are demanding. This is partially due to the diverse range of language-specific realizations represented in

---

<sup>12</sup>The fact that linguistic categories are almost never exactly the same between languages has been discussed time and again in typology; see e.g. the controversies and discussion in the construction of the World Atlas of Linguistic Structures (among many others Cysouw et al. 2008) or Croft's (2001) Radical Construction Grammar. There were several attempts to build ontologies for linguistic categories in computational linguistics. The danger here is that the level of abstraction needs to be so high that the tags are no longer informative enough for the problems at hand. The surface-syntactic universal dependencies (SUD; Kahane et al. 2021) annotation we are using for syntax is one case in point – while it may show similarities between languages, it necessarily misses language-specific details we need.

RUEG. However, this multilingual approach only adds another layer of complexity to an already complex map. Even for a single language, a phenomenon that can be realized lexically, morphosyntactically, syntactically, or through other parameters, requires a reflection of all of those linguistic levels separately and in combination; and within the frameworks respectively used. Even if every project member works from the same framework and language, different frameworks and language-specifics may be at work in pre-existing annotations, for instance those created by annotation tools such as syntactic parsers.

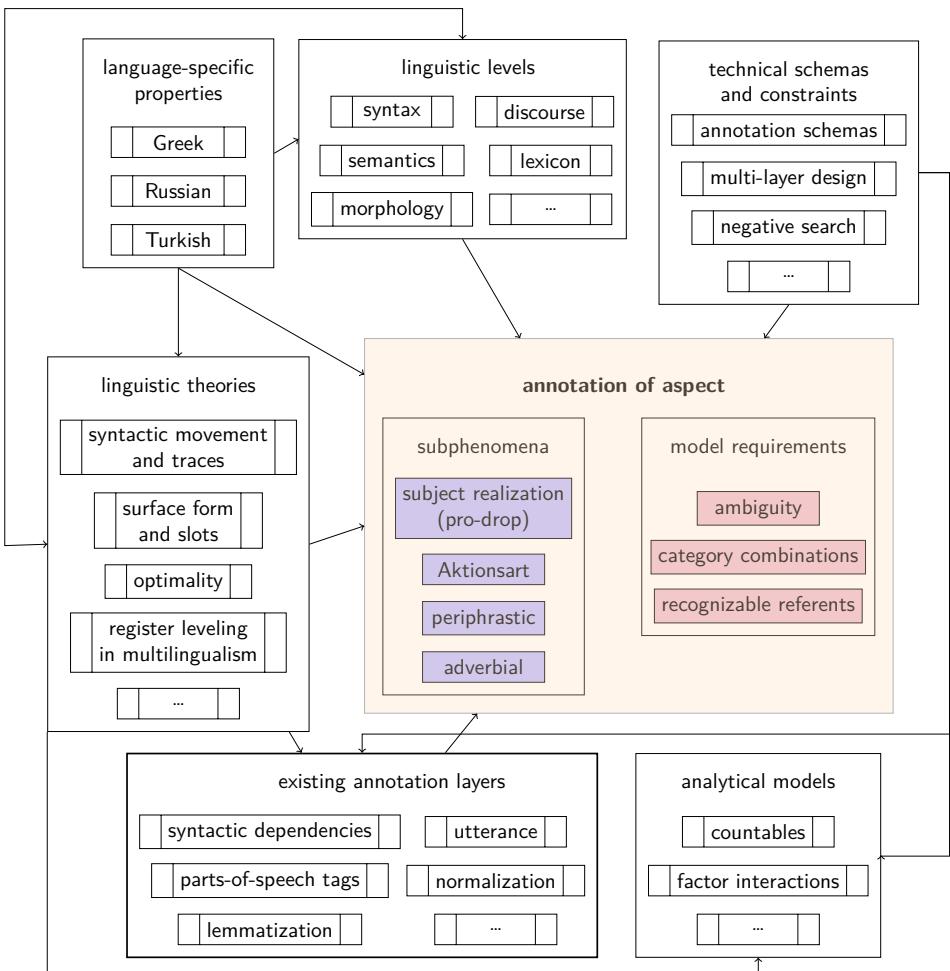


Figure 2: Model interaction in the design of annotations of aspectual expressions in Greek, Russian, and Turkish data

### 3 Mapping the mappings and then containing them all

In this way, while the actual annotation aimed for may be definable within the scope of a project (visualized as boxes with singular outlines in Figure 2), most of the models influencing those definitions are not (visualized as boxes with double outlines). As previously mentioned, this has repercussions on the workflow in a corpus project. Expertise is typically distributed by the different fields (technicians vs. linguistic experts) and the subfields they contain (syntax vs. semantics; or Greek vs. Russian). However, a common occurrence in a corpus project like RUEG is that the *specialized* expertise for the data represented in the corpus is acquired primarily by the active annotators. Active annotators may not always have as much experience with the models as designers of the research questions. Neither active annotators nor other linguistic experts often have a full understanding of the computational and technical perspective; and it is not uncommon for neither the active annotators, nor the linguistic experts, nor the technicians to have considered the interface between annotations and the final analytical model envisioned. In the simplest case, this refers to the countable elements and combinations of factors as they need to be made extractable via annotation.

In a best-case scenario, all project members involved would gain awareness of all the interfaces they are handling in the process of devising and performing the relevant annotations. In some sense, annotation can be viewed almost as a separate data collection from existing data.

The result of this work is threefold, as it creates the annotated data, the annotation guidelines, and the procedural and conceptual knowledge generated by all those who participate in annotating and annotation design. The process of annotation thus is a practice of mapping and specifying all of the linguistic models involved to the research question at hand. This requires an ongoing reflection of the perspective the annotated data lends on the existing linguistic models (hence the double-sided arrow in Figure 2). In this lies the unique advantage of corpus linguistics – in their application, we cannot only test linguistic categorizations with respect to their internal consistency, but we also learn on a case-by-case basis what *precisely* our theories have failed to predict. This sound and specific conceptual knowledge around linguistic theories and their combinatorics, i.e., their interfaces, is epistemically rich and beneficial. Conceptualizing annotation and annotation design as actual research, rather than a menial stage of preprocessing, offers many opportunities to clarify the *linguistic* understanding of a phenomenon (see also Shadrova et al. 2025). In spite of this, it is not usually in focus in corpus-linguistic study.

Rather, annotation is often rushed in order to quickly proceed to analysis and interpretation. This is unfortunate for two reasons: One, it puts tremendous pres-

sure on researchers in active annotator roles. In order to annotate well, they have to acquire expertise in all the external models and all the interfaces they are handling, while also acquiring very detailed specialized expertise on the data. They gain rich and useful knowledge that can pave the way to important new research trajectories in the understanding of language; and they acquire critical reasoning skills that are of high value in- and outside of academic employment. In spite of these objective benefits, the work of annotators is often not considered to be “real” research, and the academic recognition is limited, although it is actually of great worth to overall linguistic understanding. Two, when modeling is rushed and interfaces are left undefined, more challenges arise as the lack of clarity carries over to later research stages. This may result in frustrating workflows with many back-and-forth iterations, and can endanger the success of entire projects. This is not to be confused with a workflow that is iterative by design, in which guidelines are developed, tested, and then applied to larger sets. Rather, when interfaces are unintentionally left undefined, annotations are inconsistent, contradictory, and guidelines are changed back and forth, resulting in repeated annotations of the same data *without* increased clarity.

The work of mapping a range of models into a single data space (a corpus) requires an abstraction of the models involved, i.e., a model of each model as is relevant to the current research question. This may imply competing models, which in a multi-layer architecture can be represented in separate annotation layers. However, the outcome of research cannot be fully known in advance, which means that it is impossible to know which details might matter at the end. Thus the abstraction is often avoided by representing as many details of each linguistic theory as possible in the annotation. This, however, only creates more problems by a) burdening annotators with high demands on their working memory; and b) by creating *even more* interfaces between each detail of every model involved, which will then not only need to be reflected in the data (creating bigger data, which can turn out problematic in technical representation especially after reimport into analytical frameworks), but also in the analytical framework itself.

## 4.2 Mapping partially incompatible models

Mapping multiple models into one to obtain a common linguistic resource poses representational as well as conceptual challenges. Think of a situation (like RUEG’s) where an input data model passes through different research groups simultaneously in order to be extended and adapted to the needs of their investigations. The issue of compatibility arises when all output models require unification. A unification procedure generates a combined model from multiple models

and has its technical substantiation as a persistent resource, e. g. a linguistic corpus. There is usually more than one unification procedure available. A technical prerequisite for unification is that there is a way of unambiguously representing all models together. In contexts where this is not possible, we deal with representational incompatibility. If it is possible, we may move on to conceptual mappings limiting unification: The output of unification needs to be stable – in the sense that what is represented in a single model is represented again to the same extent and manner in the unified model – and sound, i. e. what emerges from combining each model’s aspects together still leads to valid conclusions about the modeled items. If unification fails to be stable and sound, we face conceptual incompatibility.

#### **4.2.1 Representational incompatibility**

Failure to represent a unified model is strictly bound to a context, at least for linguistic corpora. For illustration, RUEG’s German subcorpus contains two types of syntactic annotations: Syntactic dependencies following the surface-syntactic universal dependencies annotation scheme (SUD; Kahane et al. 2021), and hierarchical topological fields derived from KiDKo’s syntax scheme (Wiese et al. 2010), which for the sake of this explanation, can be understood as constituent syntax. While dependency syntax such as SUD appears technically simple, since the input ingredients are tokens with PoS annotations that are enriched with direct and labelled relations between them, the case of constituent syntax seems slightly more complicated: Apart from tokens and PoS annotations, there are non-terminal nodes that do not directly connect to the tokens.<sup>13</sup> The process of annotating dependency syntax is drawing an edge from one token to another and assigning it a label. The process of annotating constituent syntax, though, includes defining non-terminal nodes and the adjacent member tokens or non-terminal node members of it. Thus, the demands towards an annotation tool are quite distinct, and tools which can do both are quite rare. The representation of those two types of annotations and thus an underlying unified model of both constituent and dependency syntax on one screen fails, i. e. the models are incompatible in the context of the actual creation process. Similarly, it holds for many linguistic data formats, whose representational capabilities are usually a reflection of the features of an annotation tool, that they can either represent dependency syntax or constituent syntax, but not both together. Some (but not the only) exceptions are the TCF format of the online pipelining tool Weblicht

---

<sup>13</sup>Alternatively, the scheme introduces empty nodes.

(Hinrichs et al. 2010), serializations of UIMA CAS (Ogren & Bethard 2009), or the annotation tool Hexatomic (Druskat et al. 2023). In the context of search and visualization – or information retrieval – combining a model of dependency syntax and constituent syntax succeeds through apt representations (Krause & Zeldes 2014) and unification procedures (Zipser & Romary 2010, Fei et al. 2021, Krause & Klotz 2023).

#### **4.2.2 Conceptual incompatibility**

Some problems that appear to be technical issues on the surface reveal themselves to be more fundamental conceptual incompatibilities at closer inspection. It simply takes a rigid formal environment for underlying conceptual issues, that might otherwise have never been detected, to come to light. To understand under which conditions models are conceptually incompatible, we need to look at the circumstances under which their unification might be unstable or unsound. In our example of two syntactic schemes, we can think of a unification procedure that leads to different results about our annotated language samples in a unified model compared to the dependency model only. Unifying dependency and constituent trees in one model must answer the question on how to proceed with non-terminal nodes when defining the common set of tokens. Non-terminal nodes in constituent syntax can be understood as empty nodes, since they do not connect to an actual token. A unification procedure that does not map non-terminals directly to tokens, but models them as indirectly connected through their descendent nodes, is stable. It does not affect dependency trees and what can be revealed about them.<sup>14</sup> An alternative unification procedure, though, that maps non-terminal constituent nodes to newly created token objects with an empty text value between the existing input token objects, does. Even though we can still map the original dependency trees somehow onto all previously existing tokens, since none of them was deleted: By inserting new token objects, the unification alters what will be revealed about the linear order of tokens, e. g. in an n-gram analysis. Thus it is not stable.

Incompatibility arising from unsound unification is the more dangerous one, since it easily goes unnoticed. Representational compatibility is already given and we might not realize that one of our unification's mappings produces an inaccurate model of our data. Parts of the Russian and English spoken RUEG data have been annotated for prosodic features; pitch accent and boundary tones (Zerbian et al. 2022, 2024). The same subset of the data has simultaneously been an-

---

<sup>14</sup>It furthermore is sound, since everything we learn about dependence structures in combination with constituents is true.

### 3 Mapping the mappings and then containing them all

notated for basic functional and lexical features (e. g. PoS and lemma, henceforth subsumed as *morphology* or *morphological*) by another group. The shared input model to both remodeling processes, morphology and prosody, is a sequence of sub-segmentations and segmentation into tokens of an audio stream. While the morphology group bases their annotations on normalized tokens, which are a reinterpretation of the input tokens, the prosodic annotations require a subdivision of the input tokens into syllables. Normalization includes orthographic correction, split, and merge of tokens. Subdivision into syllables splits an input token at least zero times (one-syllable tokens). Both output models still contain the shared input model's entities and add new base units and annotations on top. They are illustrated on the left in Figure 3 along the example *everyone's*, a merged form of *everyone* with a clitic form of *is*, which in the diplomatic input is a single token. The prosodic group divides it into syllables ε, v.i, and wənz and assigns a pitch accent to the first. For morphology, the input token is normalized as two tokens *everyone* and *is*, of which the former is annotated to be a pronoun, the latter to be an auxiliary. Associations of annotations with their annotated unit is visualized as horizontal overlap. At first glance, the two models' representations seem comparable and thus compatible. An attempted unification is illustrated in Figure 3 on the right. We can easily diagnose stability, the content of each model completely transfers to the unified model. Unfortunately, the unified model is not sound. By the models logic of association by overlap, we actually do not know how to map syllables onto normalized tokens, but the model requires us to provide such a mapping for unification. How annotations for normalized tokens map onto syllables is undefined, so by guessing we end up with wrong prosodic information on lexical items and vice versa.

| <b>pos</b>       | PRON              | AUX |
|------------------|-------------------|-----|
| <b>norm</b>      | everyone          | is  |
| <b>dipl</b>      | <i>everyone's</i> |     |
| <b>dipl</b>      |                   |     |
| <b>syl</b>       | ε v.i wənz        |     |
| <b>pitch-acc</b> | +                 |     |

| <b>pos</b>       | PRON              | AUX      |
|------------------|-------------------|----------|
| <b>norm</b>      | everyone          | is       |
| <b>dipl</b>      | <i>everyone's</i> |          |
| <b>syl</b>       | ε                 | v.i wənz |
| <b>pitch-acc</b> | +                 |          |

Figure 3: Two input models (on the left) are unified in a stable, but unsound manner (on the right). The mapping between syllables (*syl*) and normed (*norm*) segments is undefined. The example is taken from speaker and situation USbi53MR\_isE.

Luckily, the conceptual incompatibility of the prosodic and morphological model can be resolved by updating the unification procedure to the point that it does not require a mapping of syllables to normalized tokens. A mapping is initially required, because the logic of association by overlap is transitive.<sup>15</sup> But there is no way to map the two annotation target groups *syl* and *norm* onto each other in a sound way, that keeps our unification stable, as long as we rely on a transitive relation between entities in our model and as long as we do not add a common minimal segmentation of both, *norm* and *syl*. The solution in Figure 4 provides a minimal segmentation layer (*min*)<sup>16</sup>, to which syllables and normalized tokens refer, thus a mapping between the two is given. In practice such a minimal layer is expensive in terms of annotation effort, and not necessarily defined for all potential cases. However, if we replace the transitive association relation by a semantically weaker alignment pointer, a more feasible solution can be achieved (cf. Figure 5). This way, both input models keep a copy of their diplomatic tokens and we align those input segments as if they were a parallel corpus. This way, we avoid transferring morphological features to syllables. Also, overlap is not given anymore and cannot transitively transfer morphological features to prosodic units. As a consequence, we do not require a mapping between them.

|                  |            |     |   |   |   |      |   |   |     |
|------------------|------------|-----|---|---|---|------|---|---|-----|
| <b>pos</b>       | PRON       |     |   |   |   |      |   |   | AUX |
| <b>norm</b>      | everyone   |     |   |   |   |      |   |   | is  |
| <b>dipl</b>      | everyone's |     |   |   |   |      |   |   |     |
| <b>min</b>       | e          | v   | e | r | y | o    | n | e | ' s |
| <b>syl</b>       | ɛ          | v‿i |   |   |   | wənz |   |   |     |
| <b>pitch-acc</b> | +          |     |   |   |   |      |   |   |     |

Figure 4: A stable and sound unification: A common minimal basic tokenization layer for morphology and prosody leads to a clean mapping between all annotations. Nevertheless, such a mapping is expensive, disputable, and not always available.

Conceptual incompatibility must not be mistaken for theoretical contradiction. A not very faithful syntactician might very well use dependency and constituent syntax together when annotating their data, because they see the two schemes' advantages adding up very well in their favour; so said researcher unifies two

<sup>15</sup>Here we refer to transitivity in the mathematical sense: A relation that applies for two entities *a* and *b* as well as *b* and *c* thus also applies to *a* and *c*.

<sup>16</sup>The layer is minimal w. r. t. to layers *syl* and *norm* since each unit in one of the two latter layers can be modelled as a sequence of one or more units of said layer.

### 3 Mapping the mappings and then containing them all

|                  |            |      |
|------------------|------------|------|
| <b>pos</b>       | PRON       | AUX  |
| <b>norm</b>      | everyone   | is   |
| <b>dip1</b>      | everyone's |      |
| $\uparrow$       |            |      |
| <b>dip1</b>      | everyone's |      |
| <b>syl</b>       | $\epsilon$ | v̞i  |
| <b>pitch-acc</b> | +          | wənz |

Figure 5: A sound and stable unification: A parallel corpus architecture maps only the common base units onto each other and keeps prosodic and morphological layers aligned, but separated.

data models of different syntactic annotations. Annotations are first and foremost a means of retrieving information and must not be dismissed for not being what they were never meant to be (a linguistic theory). Rather, they need to be understood as a (unified) model and in this an interpretation and reduction. At the same time, we cannot take this view on annotations so far as to consider them labels without any theoretical underpinning (Bubenhöfer 2018). Especially in the context of multi-layer architecture, the theoretical underpinnings gain importance from the process perspective of corpus creation (meaning all of the processes that lead to one released corpus) being a prerequisite of successful information retrieval. This is again exemplified in context of the RUEG project in Section 4.3 below.

#### 4.3 Interdependence between annotation layers

Designing annotation layers, especially lower-level layers containing basic information such as normalization, lemmatization, or PoS, always constrains which dependent annotation layers can be defined on their basis later on. Usually, in corpus linguistics, we are then interested in higher-level concepts that use these layers for their domain of annotation, i. e. they directly or indirectly depend on them. This dependence implies that designing and annotating lower-level layers requires transparent annotation decisions and a diligent annotation procedure. At the same time, the dependence relations between higher-level layers and lower-level layers always need to be explicitly and formally documented.

The terms *dependent (annotation) layer* and *independent (annotation) layer* will be used in the following. The *dependent layer* denotes the layer that is based on a layer added to the corpus earlier. The *independent layer* is that earlier annotated layer another layer builds on. This does not mean the independent layer cannot

itself be dependent on any of the other layers in the corpus. It only makes a statement about relations between two layers within one pairwise comparison. Since annotations necessarily rely on specific underlying models, interdependence between annotation layers is not only relevant from a technical perspective. Ignoring interdependence between annotation layers can lead to at least two different problems:

*Problem A:* Annotation layers are treated as independent of each other while they are not.

*Problem B:* Inaccurate annotations on one annotation layer cause inaccuracies on a dependent annotation layer created and annotated later on.

Both problems are strongly intertwined in the sense that if problem A is not avoided, problem B will occur as well. Hence, both are illustrated using the same example of syntax annotations of the RUEG data. Such annotations exist for all RUEG languages. Syntactic trees following the Universal Dependencies scheme for syntactic annotations (UD; de Marneffe et al. 2021) are available for Turkish, English and Greek. These annotations did not undergo a step of manual correction. For Russian and German, there are manually corrected syntactic annotations following the SUD annotation scheme. The sequence of successive processes that led to the different syntax annotations is visualized in Figure 6. It shows only the section of layers and dependences between them that are relevant here.

Complex interdependence between annotation layers already unfolds in the process product of SUD annotations alone. A wrong syntactic relation or wrong relation label in the automatically preparsed SUD annotations could originate in any of the layers the SUD layer directly or indirectly depends on. An annotator correcting such parses needs to be aware of the existence and the nature of the interdependent structure to know where to look for possible errors leading to an error on the layer that is of current concern.

Newly emerging interdependences are one reason why annotation should be viewed as an iterative process. Including a new annotation layer in the corpus that is based on a now emerged independent layer may cause annotation decisions concerning the independent layer to be reconsidered and annotations to be revised and adjusted accordingly. For instance, during correction of automatically parsed syntactic annotations, an annotator using a top-down perspective may detect a mismatch between the assigned PoS tags of two tokens and the requirements of the syntactic relation they would like to assign for them. What

### 3 Mapping the mappings and then containing them all

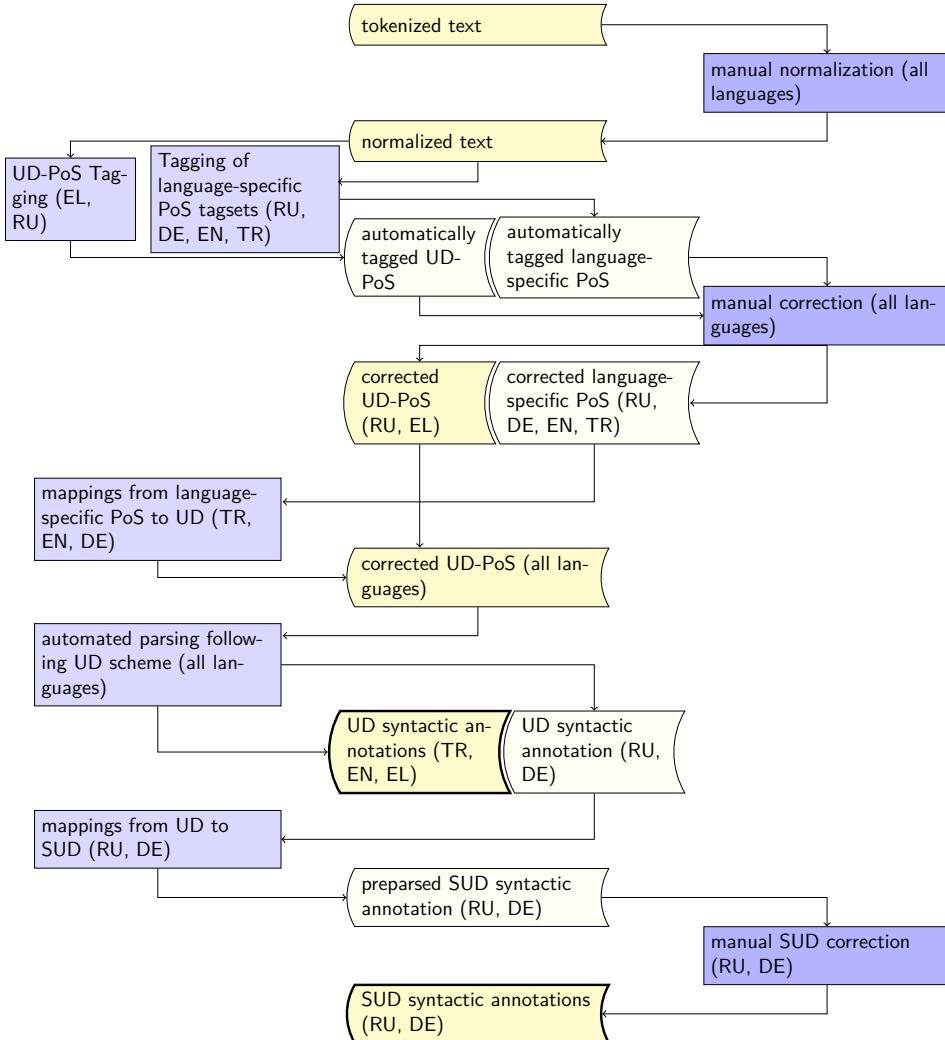


Figure 6: Subprocesses involved in the creation of syntactic dependency annotations for all RUEG languages. Book shaped boxes indicate process products (light yellow – intermediate process product; brighter yellow – process product.) Rectangle shaped boxes represent processes (light violet – automated process; brighter violet – manual process). Arrows indicate *results in* or *is passed over to*.

could be a simple annotation error on the PoS layer might also indicate a general mismatch between the models operationalized in the PoS tagset (or lower-level layers) on the one hand and in the label set for syntactic dependencies on the other hand. The PoS annotation might require revision, the definition of the syntactic relations might need to be widened to describe the data, or there may be unresolved conflicts with other annotation layers, the syntactic annotation indirectly relies on, e. g. a tokenization process that already creates preconditions for applying the “correct” dependence relation.

As formulated in Problem B, inaccurate annotations on the independent layer get inherited by a dependent layer in automatic pre-annotations, which increases the effort of manual post-correction. Otherwise the result is an annotation layer with numerous incorrect annotation values. Unawareness of an interdependence between syntactic relations and PoS however, i. e. problem A, would mean that an annotator does not take the PoS layer into account during the correction of syntactic annotations. Potential mismatches between models are overlooked and necessary model adaptions may be missed. This, again, would result in either incomprehensible or misinterpreted outcomes in analysis. This can lead to false interpretations and assumptions for our linguistic understanding, particularly when looking for emerging phenomena. A lack of awareness of attention towards the interdependence of annotation layers results in an underestimation of the complexity of the applied model(s) and of the problem at hand. Future annotations, that then rely on such annotations, fail to be understood from the beginning. This way, an initial misconception is passed on from annotation to annotation and is mirrored in any follow-up analysis. For instance, if we fail to identify proper nouns in PoS, or ignore PoS in our annotations of syntax, we may then also fail to evaluate the introduction of referents through specific syntactic means.

Ignoring the interdependence of layers replicates itself when more additional annotation layers are added while previous interdependence still goes unnoticed, undocumented and not taken into consideration for the planning of new dependent layers. An alternative approach to documenting interdependence and the consequences arising from that for the underlying models is possible if different models cannot be conceptually reconciled. In this case the independence and incompatibility of the layers has to be made transparent and explicit in the documentation because incompatible layers should consequently not be used together for analyses. A more radical solution is to publish such layers separately in different subcorpora.

#### 4.4 Mapping unknown models

So far, we have discussed different aspects of handling multiple models, such as their compatibility or their interdependence. A special case occurs in the modeling of data from fully or partially *unknown* linguistic models. As RUEG is interested in *emergent* properties, i.e., new linguistic feature distributions that cannot be fully predicted from the sum of the parts of other features, this is a common requirement in the project context.

The linguistic basis for the process we sketch out in this section is the idea that language behaves as a complex dynamic, i.e. emergent, system (Beckner et al. 2009, Ellis & Larsen-Freeman 2009, Massip-Bonet 2013, Croft 2014: among others). A system of this kind is emergent, because it amounts to more than the sums of its parts and their relations, i.e. some effects arise from the interaction of factors or even interactions. It is (mathematically) *complex*, because the behavior of the entire system can be stochastically described, but not fully predicted. It is dynamic, because it moves through different states in an ever-changing and evolving way. Dynamics may include *directional* movement (loss or emergence of cases) or oscillating movement (loss and re-emergence of negation markers), and processes of grammaticalization and de-grammaticalization equally. Language has been modeled as a complex dynamic system both on population (German in Germany) and individual level (a speaker's linguistic system at various stages of acquisition or years of life, see Hiver et al. (2022) for a recent overview).

Some of the underlying assumptions of the dynamic systems/emergentist perspective on language change and language development are that although shifts happen systematically, they are not fully predictable; that once a relevant shift happens to the system, other paths are less likely to be taken (path dependence); that shifts do not usually encompass the entire system at once, but can have *repercussions* on the entire system (dialectics); that shifts in minor aspects of a language may cascade into more widely effective changes (quantity-quality shift); and that amidst a changing process, the system will be highly variable in the relevant aspects due to residing in transitory, superpositioned, and partially unpredictable states. These may later resolve in one direction or another; however, in describing a system from within its transitory state, different possibilities have to be considered at the same time.

These transitory states are what we refer to as partially (or fully) unknown models. Although they might eventually be fully describable (i.e. known), during analysis they require a perspective that is not fully guided by the origin or the presumed target of the development. Rather, we must assume that the abstractions we are looking for would differ from the canonical paradigm of the standard language, but since it is new, the details are unknown.

We might try to solve this by defining target hypotheses (Reznicek et al. 2013, Lüdeling & Hirschmann 2015, Laarmann-Quante et al. 2017), i.e., a normalized layer of the utterance in question, as it is common in learner corpus research. For this, one would note the canonical form next to the form evidenced in the corpus and classify the differences. However, this does not provide a solution for the emergent system, as the target is not the canonical form – rather, this approach would unintentionally result in a deficiency-oriented perspective, in which one presumes that heritage speakers “mean”, i.e., project as their target, the canonical standard language, but cannot fully meet its requirements. If one were to analyze the new paradigm based on target hypotheses, one would inevitably analyze the emergent system as merely a divergence from the state it emerged from, projecting the old logic and inner constraints onto the system.

Alternatively, one could try and analyze the new forms based on *an assumption* of what the new system might look like. This assumption may be highly educated, for instance informed by theories of ongoing language change and the dynamics of bilingualism. However, even an educated guess is *only a first hypothesis*, not a synthesized and integrated theory of the new paradigm. It will likely not match the data in all relevant aspects. If the presumed new paradigm has already been projected onto the data in the form of annotations, one enters a risk of getting trapped in a never-ending cycle of re-annotating: as soon as one finds a counter-example, the entire new system has to be reworked, hence requiring re-annotation of all previous data. However, since one’s initial hypothesis was probably right in some aspects, then, evidence may be gathered again for the initial hypothesis. The hypothesized system bounces back and the annotations have to be reworked again. It is important to note that this is different from an iterative annotation process based on a clearly defined linguistic model, because the limits and degree of certainty of the hypothesized model (the educated guess) are unknown at each step. It follows that the next iteration does not necessarily result in greater certainty or clarity, as is the case in improved annotation guidelines. Rather, the annotation guidelines and the target system oscillate between several equally uncertain systems.

A solution to this problem instead lies in modeling by form (in a corpus-driven manner) when abstractions are not yet sufficiently certain to be included in the annotation, and mapping relevant aspects of the *original*, i.e., canonical forms, in a separate layer. In this new way, the mapping represents the observed data *form* mapped to a presumed target *function*, i.e., the mapping between the new form and the old function is preserved, but no new function is presumed. Unlike this, the *functional* annotation of the presumed function is *more abstract* than in the approaches mentioned above.

Careful consideration of the effect of abstraction at each step is required in order to arrive at precise annotations. This is particularly the case where models themselves are unstable or dynamic. Finding the right timing and the right level of abstraction requires a bird's eye view on all processes and models involved and is usually not achievable ad hoc, but itself an iterative and communicatively demanding task. However, it is the only way to derive an emergent phenomenon from existing data, and, if done carefully, provides deep insights into all aspects of the target paradigm including its dynamic and not yet fully shifted parts. With this, we conclude our discussion of different types of model and interface mappings.

## 5 Discussion

Corpus data is rather unique in linguistic research, as it contains language at full complexity regardless of how narrowly defined the communicative situation of its elicitation. Unlike in an experimental setting, complexity is not reduced in a selection process prior to elicitation, and outcomes are not funneled into one of  $n$  predetermined options. This is why all the modeling and analytical design has to be completed after elicitation. An aspect that we cannot discuss in this contribution, but that also concerns similar modeling decisions, is that of mapping quantitative models to the open-ended data that corpora present. These mappings, too, reflect complexes of explicit or implicit analytical decisions that carry repercussions for the evidential scope<sup>17</sup> of corpus-linguistic research.

Most of the epistemological weight of corpus-linguistic research lies on post hoc differentiation, rather than a priori design, of the data – it is during annotation design and execution, data extraction (i.e., selection), and data analysis that the fitting of all involved linguistic models with the actual data happens. In addition to the linguistic complexities, a range of technical limitations poses further demands.

In an ideal setting, all of the involved models are defined somewhere (in a linguistic theory outline; a technical framework manual; or an annotation guideline). Unpacking this complexity requires manual, additive, and collaborative work, and effectively creates distributed knowledge. This means that even in the best of scenarios, there are no natural carriers of *all* the information regarding *all* the models relevant to a strongly collaborative project, or simply put: no-one knows everything that matters about the data. This is commonly solved by

---

<sup>17</sup>The evidential scope (see e. g. Leonelli 2009, 2019) defines the limits of what can be learned from a specific type of data or methodology.

distributing work between representatives of different fields of expertise, for instance by research field, methodology, or support roles (technicians, statisticians, linguists, etc.).

Although it lifts some of the pressure to understand many models, this division of labor creates further interfaces between those who carry the understanding of some aspects and those who carry the understanding of some other aspects. In that context, setting aside space for a duplicate of each set of epistemes (i.e., two people working on the same model subset) would be helpful, but is not highly realistic under the current conditions of research funding.

Worse perhaps, no-one can know *enough* about the data *a priori* in a corpus project, since it is unpredictably variable even in the narrowest of settings. Expertise thus has to be grown *from within the corpus project*, that is from a growing understanding of the data, the specific requirements of the annotation or analytical model, and the technological options and limitations. These epistemes emerge as a process product of corpus-linguistic research. Even with high discipline and the best of intentions, they can be difficult to document and transfer, because much of the procedural knowledge exists on a perceptual level which can be difficult to notice and describe, and thus it often remains implicit. Additionally, keeping a discipline at documenting can be a challenge due to the high number of interfaces and mental objects involved, and since there is very little incentive within the academic system for it. Luckily, this is slowly beginning to change. In recent years, well-designed and documented resources and research software are more commonly acknowledged as academic output. This is evidenced, for instance, by the increase in published annotation guidelines, the option to publish materials and documentations as data supplements with many research articles, as well as a changed CV structure in some funding agencies and the professionalization of research infrastructure citation (Lavoie 2012, Anzt et al. 2020, Schlauch et al. 2022, among others). In spite of these very positive developments, the fact remains that the kind of procedural knowledge we have described is not yet regarded in the same way as more established formats of scientific publishing, although it is often the first and primary type of knowledge gained in a complex corpus project during compilation (i.e., before final analysis).

While it seems clear that procedural knowledge relevant to a particular project cannot be simply copied onto the next set of problems, the transfer of process-oriented knowledge is crucial to the success of each corpus project, which includes *every collaborator's next project*. Personal continuity is helpful, but it is clear that not every single person working on a corpus project will continue on the same path in the future; and not every new collaborator in a future corpus project would prefer to read through the documentations of every single other,

even similar, corpus project. Even if they were disciplined and motivated to do so, the number of mental objects, interface definitions and so on would easily occur as overwhelming to them as they would have been to the initial group.

(How) can the *understanding* of model fitting then be transferred between projects? Would it require more synthesis of existing procedural knowledge, as we have attempted in this contribution – a theoretical corpus linguistics for a phenomenology and a typology of corpus-specific modeling problems? Or does it suffice to know the linguistic details and the general necessity to make them fit with the quantitative and technical models and frameworks? Should corpus-linguistic infrastructure and modeling be viewed as a type of advisory work, similar to a statistical consultant, or as a type of engineering work akin to research software engineering; or would *all* corpus linguists need to be more involved with the quantitative, technical and also epistemological aspects of their work, resulting in a community-of-practice approach, in which all participants understand the majority of their data in all its aspects and develop the expertise to debate it in multidimensional ways? If so, as a logical consequence their work would require smaller corpora with more time and space for in-depth linguistic and modeling debate.

With its focus on deep annotation, clear theoretical embedding, an integration and comparison of various features, and its strict handling of documentation, quality assurance, and continuous integration, RUEG is a good example for best practices in some aspects of this. It has created an impressively well-documented, impressively well-controlled and rich database which will not only continue to be used in future research in its current state, but can also be enriched with new layers of annotation, expanded, and its materials reused for conceptual replication. However, even at this level of precision and disciplined documentation, there are still manifold interfaces that will in all likelihood remain underspecified and untreated, unless they are attended to in other projects; and it is not highly likely to receive funding for the modeling of legacy data from other projects. This suggests that RUEG likely poses an upper limit to what can be contained within a corpus project without creating insurmountable organizational and communicative overhead.

This tension stands a huge challenge for prospective projects, especially within the context of a strong emphasis of a multifactorial quantitative paradigm in the current debate. Corpus complexity is a direct correlate of the number of models and factor combinations involved: the more factors there are to be controlled for, the more factors will need to be considered in the analytical model. This results in more variability, in this case variability along more dimensions, that need to be accounted for in the data. To linguistically capture the full extent

of the data, a more complex model is required, i.e., the inclusion of more linguistic aspects or sub-models and their interactions. All of these aspects – the factors, the variability along each dimension, each model, the hypotheses concerning with respect to every aspect, the multidimensional results and their repercussions for each of the involved linguistic subdomains – then has to be considered and discussed in the scientific debate, if as many epistemes as possible are to be procured in the process. From the complexity emerging from this interaction of elements arises a big epistemological question: what can reasonably be *understood* and *explained* as *linguistic knowledge*, rather than simply quantitatively *described* in naturalistic data?

Although task-based corpora have been used for many years in some subfields of corpus linguistics, a recent shift has taken the paradigm towards the elicitation of data that is very strictly controlled by a range of factors. This is a reflection of a better understanding of how task and register effects shape natural language and is a generally positive development, as these elicitations have lead to a much better understanding of the multiplicity of interactions of factors and factor combinations in language production (e.g. Alexopoulou et al. 2017, see also Wiese, Labrenz, et al. 2025 [this volume]). This newer tendency, of which RUEG is one example, has also shown that speakers in *any* communicative situation do not produce (quasi-)experimental data. Rather, we see that the full scale of emergent properties of natural language production and the multiplicity of interrelations between linguistic and extra-linguistic factors are present even within such narrowly constrained communicative situations. It is only natural to conclude that corpus linguistics should pay close attention to those emergent properties. With this contribution, we argue for more exchange, debate, and synthesis of corpus-based modeling expertise, as the many interwoven layers of the emergent space require much epistemological clarity to be fully grasped and productively handled.

## Abbreviations

|     |  |
|-----|--|
| PoS | part of speech                           |
| SUD | surface-syntactic universal dependencies |
| UD  | universal dependencies                   |

## Acknowledgements

All authors are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Research Unit *Emerging Grammars*, Projects Pd A

cross-linguistic corpus of comparative data in heritage and majority language use, 394844736, and *Pc Corpus-linguistic methods*, 313607803.

We thank Thomas Krause for his crucial role in developing the conceptual and practical foundations of the corpus infrastructure. We are grateful to our student assistant Esra Uyanik for her support in corpus development and careful copy-editing of this chapter. We express our gratitude to Sabine Zerbian, Maria Martynova, Mareike Keller, Nadine Zürn, and two anonymous reviewers for their insightful feedback which has helped to improve this chapter greatly. We also thank Lea Coy for help with pre-publication formatting.

None of this would have been possible without the inspiring collaboration with and the support of the entire RUEG team and the Mercator Fellows.

Additionally, Anke Lüdeling and Martin Klotz acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412 *Register*, 416591334.

## References

- Alexopoulou, Theodora, Marije Michel, Akira Murakami & Detmar Meurers. 2017. Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning* 67(S1). 180–208. DOI: [10.1111/lang.12232](https://doi.org/10.1111/lang.12232).
- Anzt, Hartwig, Felix Bach, Stephan Druskat, Frank Löfller, Axel Loewe, Bernhard Y Renard, Gunnar Seemann, Alexander Struck, Elke Achhammer, Piush Aggarwal, et al. 2020. An environment for sustainable research software in Germany and beyond: Current state, open challenges, and call for action. *F1000Research* 9. DOI: [10.12688/f1000research.23224.2](https://doi.org/10.12688/f1000research.23224.2).
- Atwell, Eric. 2008. Development of tagsets for part-of-speech tagging. In Anke Lüdeling & Merja Kyö (eds.), *Corpus linguistics. An international handbook*, 501–527. Berlin: De Gruyter Mouton.
- Bauer, Laurie. 2000. Word. In Geert E. Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie/Morphology: Ein internationales Handbuch zur Flexion und Wortbildung/An international handbook on inflection and word formation*, vol. 17, 247–257. De Gruyter Mouton. DOI: [10.1515/978311011286.1.4.247](https://doi.org/10.1515/978311011286.1.4.247).
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, et al. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59. 1–26. DOI: [10.1111/j.1467-9922.2009.00533.x](https://doi.org/10.1111/j.1467-9922.2009.00533.x).

- Biber, Douglas. 2015. Corpus-based and corpus-driven analyses of language variation and use. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 193–224. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199544004.013.0008.
- Boersma, Paul. 2001. Praat: A system for doing phonetics by computer. *Glot International* 5(9/10). 341–345.
- Bubenhofer, Noah. 2018. Wenn „Linguistik“ in „Korpuslinguistik“ bedeutungslos wird: Vier Thesen zur Zukunft der Korpuslinguistik. *Osnabrücker Beiträge zur Sprachtheorie (OBST)* (32). 17–30. DOI: 10.5167/uzh-199167.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198299554.001.0001.
- Croft, William. 2014. Studying language as a complex adaptive system. *English Linguistics* 31(1). 1–21. DOI: 10.9793/elsj.31.1\_1.
- Cysouw, Michael, Mihai Albu & Andreas Dress. 2008. Analyzing feature consistency using dissimilarity matrices. *Sprachtypologie und Universalienforschung* 61(3). 263–279. DOI: 10.1524/stuf.2008.0025.
- de Marneffe, Marie-Catherine, Christopher D. Manning, Joakim Nivre & Daniel Zeman. 2021. Universal dependencies. *Computational Linguistics* 47(2). 255–308. DOI: 10.1162/coli\_a\_00402.
- Dewaele, Jean-Marc. 2019. The vital need for ontological, epistemological and methodological diversity in applied linguistics. In Claire Wright, Lou Harvey & James Simpson (eds.), *Voices and practices in applied linguistics: Diversifying a discipline*, 71–88. York: White Rose University Press. DOI: 10.22559/BAAL1.e.
- Druskat, Stephan, Thomas Krause, Clara Lachenmaier & Bastian Bunzeck. 2023. Hexatomic: An extensible, OS-independent platform for deep multi-layer linguistic annotation of corpora. *Journal of Open Source Software* 8(86). 4825. DOI: 10.21105/joss.04825.
- Ellis, Nick C. & Diane Larsen-Freeman. 2009. *Language as a complex adaptive system*. John Wiley & Sons.
- Fei, Hao, Shengqiong Wu, Yafeng Ren, Fei Li & Donghong Ji. 2021. Better combine them together! Integrating syntactic constituency and dependency representations for semantic role labeling. In Chengqing Zong, Fei Xia, Wenjie Li & Roberto Navigli (eds.), *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 549–559. Stroudsburg, PA: Association for Computational Linguistics. DOI: 10.18653/v1/2021.findings-acl.49.
- Gould, Elliot, Hannah S. Fraser, Timothy H. Parker, Shinichi Nakagawa, Simon C. Griffith, Peter A. Veski, Fiona Fidler, Daniel G. Hamilton, Robin N. Abbey-Lee, Jessica K. Abbott, et al. 2025. Same data, different analysts: Variation in effect

### 3 Mapping the mappings and then containing them all

- sizes due to analytical decisions in ecology and evolutionary biology. *BMC Biology* 23(35). DOI: 10.1186/s12915-024-02101-x.
- Gozli, Davood. 2019. Shifting focus. *Experimental Psychology and Human Agency*. 1–16. DOI: 10.1007/978-3-030-20422-8\_1.
- Hinrichs, Erhard W., Marie Hinrichs & Thomas Zastrow. 2010. WebLicht: Web-based LRT services for German. In Sandra Kübler (ed.), *Proceedings of the ACL 2010 system demonstrations*, 25–29. Uppsala: Association for Computational Linguistics. <http://www.aclweb.org/anthology/P10-4005>.
- Hiver, Phil, Ali H Al-Hoorie & Reid Evans. 2022. Complex dynamic systems theory in language learning: A scoping review of 25 years of research. *Studies in Second Language Acquisition* 44. 913–141. DOI: 10.1017/S0272263121000553.
- Kahane, Sylvain, Bernard Caron, Emmett Strickland & Kim Gerdes. 2021. Annotation guidelines of UD and SUD treebanks for spoken corpora: A proposal. In Daniel Dakota, Kilian Evang & Sandra Kübler (eds.), *Proceedings of the 20th international workshop on treebanks and linguistic theories (TLT, SyntaxFest 2021)*, 35–47. Sofia, Bulgaria: Association for Computational Linguistics. <https://aclanthology.org/2021.tlt-1.4> (24 June, 2025).
- Klotz, Martin, Rahel Gajaneh Hartz, Annika Labrenz, Anke Lüdeling & Anna Shadrova. 2024. Die RUEG-Korpora: Ein Blick auf Design, Aufbau, Infrastruktur und Nachnutzung multilingualer Forschungsdaten. *Zeitschrift für germanistische Linguistik* 52(3). 578–592. DOI: 10.1515/zgl-2024-2026.
- Klotz, Martin & Anna Shadrova. In preparation. *The impact of manual post correction of annotations on corpus analysis*.
- Knobloch, Clemens & Burkhard Schaeder. 2000. Kriterien für die Definition von Wortarten. In Geert E. Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie/Morphology: Ein internationales Handbuch zur Flexion und Wortbildung/An international handbook on inflection and word formation*, vol. 1, 674–692. Berlin: De Gruyter Mouton. DOI: 10.1515/978311011286.1.10.674.
- Krause, Thomas & Martin Klotz. 2023. *Annatto*. Version 0.2.0. <https://github.com/korpling/annatto/>.
- Krause, Thomas & Amir Zeldes. 2014. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1). 118–139. DOI: 10.1093/lhc/fqu057.
- Laarmann-Quante, Ronja, Katrin Ortmann, Anna Ehler, Maurice Vogel & Stefanie Dipper. 2017. Annotating orthographic target hypotheses in a German L1 learner corpus. In Joel Tetreault, Jill Burstein, Claudia Leacock & Helen Yannakoudakis (eds.), *Proceedings of the 12th workshop on innovative use of NLP for building educational applications*, 444–456. Copenhagen: Association for Computational Linguistics. DOI: 10.18653/v1/W17-5051.

- Lavoie, Brian F. 2012. Sustainable research data. *Managing Research Data*. 67–82. DOI: 10.29085/9781856048910.005.
- Leonelli, Sabina. 2009. On the locality of data and claims about phenomena. *Philosophy of Science* 76(5). 737–749. DOI: 10.1086/605804.
- Leonelli, Sabina. 2019. What distinguishes data from models? *European Journal for Philosophy of Science* 9(2). 22. DOI: 10.1007/s13194-018-0246-0.
- Lüdeling, Anke, Artemis Alexiadou, Aria Adli, Karin Donhauser, Malte Dreyer, Markus Egg, Anna Helene Feulner, Natalia Gagarina, Wolfgang Hock, Stefanie Jannedy, Frank Kammerzell, Pia Knoeferle, Thomas Krause, Manfred Krifka, Silvia Kutscher, Beate Lütke, Thomas McFadden, Roland Meyer, Christine Mooshammer, Stefan Müller, Katja Maquate, Muriel Norde, Uli Sauerland, Stephanie Solt, Luka Szucsich, Elisabeth Verhoeven, Richard Waltereit, Anne Wolfsgruber & Lars Erik Zeige. 2022. Register: Language users' knowledge of situational-functional variation. *Register Aspects of Language in Situation* 1(1). 1–58. DOI: 10.18452/24901.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Lüdeling, Anke & Hagen Hirschmann. 2015. Error annotation systems. In Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (eds.), *The Cambridge handbook of learner corpus research*, 135–157. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139649414.007.
- Massip-Bonet, Àngels. 2013. Language as a complex adaptive system: Towards an integrative linguistics. In Àngels Massip-Bonet & Albert Bastardas-Boada (eds.), *Complexity perspectives on language, communication and society*, 35–60. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-32817-6\_4.
- Moisl, Hermann. 2009. Exploratory multivariate analysis. In Anke Lüdeling & Merja Kyö (eds.), *Corpus linguistics: An international handbook*, vol. 2, 874–899. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110213881.2.874.
- Montrul, Silvina. 2015. *The acquisition of heritage languages*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139030502.
- Ogren, Philip V. & Steven J. Bethard. 2009. Building test suites for UIMA components. In Kevin Bretonnel Cohen & Marc Light (eds.), *Proceedings of the workshop on software engineering, testing, and quality assurance for Natural Lan-*

### 3 Mapping the mappings and then containing them all

- guage Processing (SETQA-NLP 2009), 1–4. Boulder, CO.: Association for Computational Linguistics. <https://aclanthology.org/W09-1501/>.
- Özsoy, Onur & Frederic Blum. 2023. Exploring individual variation in Turkish heritage speakers' complex linguistic productions: Evidence from discourse markers. *Applied Psycholinguistics* 44(4). 534–564. DOI: 10.1017/S0142716423000267.
- Pascual y Cabo, Diego & Jason Rothman. 2012. The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics* 33(4). 450–455. DOI: 10.1093/applin/ams037.
- Petrov, Slav, Dipanjan Das & Ryan McDonald. 2012. A universal part-of-speech tagset. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the eight international conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA).
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Reznicek, Marc, Anke Lüdeling & Hagen Hirschmann. 2013. Competing target hypotheses in the Falko corpus. *Automatic Treatment and Analysis of Learner Corpus Data* 59. 101–123. DOI: 10.1075/scl.59.07rez.
- Robins, Robert H. 1986. The Technē Grammatikē of Dionysius Thrax in its historical perspective: The evolution of the traditional European word class system. In Pierre Swiggers & Willy van Hoecke (eds.), *Mot et parties du discours. Word and word classes. Wort und Wortarten*. Leuven: Leuven University Press.
- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2). 155–163. DOI: 10.1177/1367006909339814.
- Schlauch, Tobias, Oliver Bertuch, Oliver Knodel, Guido Juckeland, Stephan Druskat & Michael Meinel. 2022. HERMES: Automated software publication with rich metadata. In *Helmholtz open science briefing: Helmholtz open science forum Forschungssoftware*, 107–116. DOI: 10.5281/zenodo.6241553.
- Schmid, Helmut. 2008. Tokenizing and part-of-speech tagging. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, vol. 1 (HSK), chap. 24, 527–551. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110211429.1.484.
- Schmidt, Thomas & Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *Handbook on corpus phonology*, 402–419. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199571932.013.030.

- Shadrova, Anna. 2025. No three productions alike: Lexical variability, situated dynamics, and path dependence in task-based corpora. *Open Linguistics* 11(1). 20240036. DOI: 10.1515/opli-2024-0036.
- Shadrova, Anna, Joshua Boivin, Rahel Gajaneh Hartz & Anke Lüdeling. In preparation. *Bilingualism modulates complex verb use in a task-based corpus of L1 German*.
- Shadrova, Anna, Anke Lüdeling, Rahel Gajaneh Hartz, Thomas Krause & Martin Klotz. 2025. „Step away from the computer!“: Über die linguistische Datenkategorisierung als Erkenntnisprozess und daraus folgende Herausforderungen bei der Nachnutzung von Annotationen und Annotationstools. *Zeitschrift für Germanistische Linguistik* 53(1). 166–214.
- Stachowiak, Herbert. 1973. *Allgemeine Modelltheorie*. Wien & New York: Springer.
- White, Hayden. 1973. *Metahistory: The historical imagination in nineteenth-century Europe*. John Hopkins University.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagaria, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Shanley E. M. Allen, Mareike Keller & Artemis Alexiadou. 2025. Introduction: Investigating the dynamics of language contact situations. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 1–29. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775157.
- Wiese, Heike, Oliver Bunk, Fynn Dobler, Ulrike Freywald, Sophie Hamm, Banu Hueck, Anne Junghans, Jana Kiobassa, Julia Kostka, Marlen Leisner, Nadine Lestmann, Katharina Mayr, Tiner Özcelik, Charlotte Pauli, Gergana Popova, Ines Rehbein, Nadja Reinhold, Franziska Rohland, Sören Schalowski, Kathleen Schumann, Kristina T. Sommer & Emiel Visser. 2010. *KiDKo: Ein Korpus spontaner Unterhaltungen unter Jugendlichen im multiethnischen und monoethnischen urbanen Raum*. <http://kiezdeutschkorpus.de/en/>.
- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Lin-*

### *3 Mapping the mappings and then containing them all*

- guistic dynamics in heritage speakers: Insights from the RUEG group, 33–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.1577519.
- Zeldes, Amir. 2018. *Multilayer corpus studies*. London: Routledge.
- Zerbian, Sabine, Marlene Böttcher & Yulia Zuban. 2022. Prosody of contrastive adjectives in mono-and bilingual speakers of English and Russian: A corpus study. In *Proceedings 11th international conference on speech prosody. International speech communication association (ISCA) online archive*, 812–816. DOI: 10.21437/SpeechProsody.2022-165.
- Zerbian, Sabine, Yulia Zuban & Martin Klotz. 2024. Intonational features of spontaneous narrations in monolingual and heritage Russian in the U.S.: An exploration of the RUEG corpus. *Languages* 9(1). DOI: 10.3390/languages9010002.
- Zipser, Florian & Laurent Romary. 2010. A model oriented approach to the mapping of annotation formats using standards. In *Workshop on language resource and language technology standards, LREC 2010*. La Valette, Malta. <https://hal.inria.fr/inria-00527799>.



# Chapter 4

## Family language dynamics: Strengthening heritage speakers' linguistic resources

Judith Purkarthofer<sup>a</sup>, Rosemarie Tracy<sup>b</sup>, Sofia Grigoriadou<sup>a</sup>  
& Johanna Tausch<sup>b,c</sup>

<sup>a</sup>University of Duisburg-Essen <sup>b</sup>University of Mannheim <sup>c</sup>Leibniz Institute for  
the German Language

Research on heritage languages is not only empirically and theoretically relevant: It also touches a nerve with families, minority speech communities and educators, teachers and other agents within the majority language educational and health systems. This contribution presents the rationale of a specific transfer project based on the findings of the Research Unit *Emerging Grammars in Language Contact Situations* (RUEG) as well as on multilingualism research in general. The need for differentiated outreach activities is highlighted; first steps of implementation in close cooperation with non-academic partners and their target groups are addressed. Target groups include parents with very rudimentary to excellent majority language skills, professionals providing advice to parents or involved in language fostering as well as instructors in qualifying programs for educators and teachers. Initiatives include workshops, video lectures, interviews and other material, eventually all available online via [www.ruegram.de](http://www.ruegram.de) as well as suggestions for specific citizen science activities already conducted or under way.

### 1 Introduction

Heritage languages play a crucial role in multilingual societies: they are important social facts for families, but also for educational institutions, workplaces



and social interaction in general. At the same time, they contribute to a multilingual language ecology, thereby at times challenging assumptions about language use and development. Attitudes towards majority, minority and minoritised languages influence how parents and children valorise or disregard their own and their friends' languages. To this day, and maybe even more so than 50 years ago – a time when individual heritage speakers were typically immersed (or rather submersed) in majority language classrooms – immigrant families need to defend their wish to pass on their minority languages, even if researchers generally agree on the positive outcomes of bi- and trilingual family language policies (Schwartz & Verschik 2013, Tracy 2014, Tracy et al. 2018, Arnaus Gil et al. 2019). While early research on language transmission (Fishman 1991) was particularly interested in generational transfer in order to evaluate the vitality of heritage languages, recent research has paid more attention to the parallel use of several languages as part of the linguistic repertoire of speakers (i.e., on multilingual speakers' biographies, Pavlenko & Blackledge 2004, Busch 2012, Purkarthofer & Flubacher 2022; on language, globalisation and superdiversity, Arnaut et al. 2016).

Despite its ubiquity, multilingualism within the family and the wish to enable children to grow up with more than one language is still considered a private matter – albeit sometimes heavily policed by society. In Germany, as in many other countries, questions of multilingual upbringing often come with a strong focus on supposedly successful strategies for integrating minorities into the majority society (Tracy 2014, Schroeder 2017, Wiese et al. 2020). This is happening in parallel to and hardly disturbed by research findings indicating that bilingual schooling can be beneficial for children even with regard to test results in maths and sciences. In general, however, there is still little public awareness of what it takes for language acquisition, monolingual or bilingual, to get off the ground.

Hard-to-die ideologies and, above all, discriminatory discourses work against multilingualism in childhood, except for cases of elite bilingualism involving languages with undisputed status, for instance English or French in Germany. Already in her 1998 volume on bilingual education programs, Brisk (1998: 1) pointed out that “[t]he paradox of bilingual education is that when it is employed in private schools for the children of elites throughout the world it is accepted as educationally valid [...]. However, when public schools implemented education for language minority students over the past 50 years, bilingual education became highly controversial.”

This means that families, but also those they turn to for advice – early childhood educators, teachers, paediatricians – often are at a loss with respect to

crucial language-related and educational issues, regardless of the languages involved (cf. Hopp et al. 2010, Edwards 2015, Purkarthofer 2019). From the perspective of some institutions and majority language speakers, the visibility and the maintenance of heritage languages are seen as threats to societal peace and solidarity (Tracy 2014, Wiese et al. 2020). Concerns about societal rift are even voiced by those who one might expect to know better, namely immigrants, in Germany and beyond. Linguist and US-senator Hayakawa, for example, was a stout supporter of the English-Only movement and opponent of bilingual education programs, which he considered “a costly and confusing bureaucratic nightmare”, likely to enhance political disloyalty and destabilization (Hayakawa 1992: 44). To this day, many agents in the educational system and the public in general lack information on what shapes languages, what is needed for acquisition and what brings about language change apart from the presence of “other” languages. In general, it is fair to say that beyond complaints about decreasing standards of politeness, formality, or writing skills, there is little awareness that languages are highly dynamic, hence – somewhat ironically formulated! – constantly “on the move”, with or without language contact or migration. Transmission of research-backed knowledge is the main objective of the transfer project entitled “Family language dynamics: empowering speakers of majority and heritage languages” of the Research Unit *Emerging Grammars in Language Contact Situations* (RUEG; <https://hu.berlin/rueg>). Target groups are parents and educators of mono- and multilingual children as well as children and adolescents themselves. All activities focus on speakers’ multilingual repertoires and relate to RUEG’s main research question “What are the linguistic dynamics in heritage speakers’ repertoires?”

The RUEG transfer activities are organised in five categories (see Figure 1). We will refer to this pentagon and its modules throughout the rest of this chapter. The categories cover:

1. Knowledge about languages and language use (green),
2. Specific research findings illustrating language ‘live’ in language contact situations (blue),
3. Experiences of multilingual speakers (purple),
4. Myths/Discourses about languages and multilingual language use (red) and finally
5. Citizen science activities, researching in one’s own family and community (yellow).

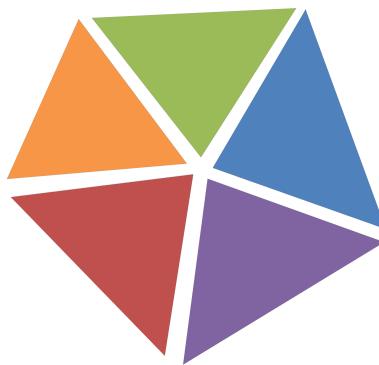


Figure 1: Five modules: (1) knowledge (green), specific research findings (blue), experiences (purple), myths/discourses (red) as well as citizen science (yellow).

Taken together, these components provide research-backed facts for interested parents and educators and invite speakers into meaningful dialogue about their experiences, attitudes and language practices. Our contribution starts in the red area (in Section 2) with myths and misconceptions about the coexistence of two or more languages in individuals and society. We then move to our own experiences (Section 3, purple) to describe the specific research process and methods of the transfer project, before we introduce family language dynamics in its relation to heritage languages more broadly (in blue and green Sections 4 and 5, covering specific results and broader knowledge about family languages). In Section 6 (yellow), we present a range of outreach and citizen science activities backed by linguistic research and developed to strengthen heritage language speakers' resources in the context of the project. Finally, in Section 7, we draw conclusions with regard to audiences, investments, needs and goals.

## 2 Myth busting: Tackling ghosts from the past?

In 1988, parents speaking another language at home besides German were given the following advice in an information leaflet published by a municipal Youth Welfare Department (*Jugendamt*) of a major German city, here translated into English:

If father and mother speak different languages, what should the children learn? Both languages at the same time? Or better one after the other? And in that case which language first? Because every family situation is different,

no definite advice can be given. There's only one thing that can be said; it would be good if the children could feel one language to be their so-called mother tongue, i.e. could feel at home in one language. The advantage of growing up bilingually is often bought at the expense of insecurity in both languages, and this development should be avoided if at all possible. [...] In most cases, the right thing will be for the children to grow up with the language of the country where they live, where they are growing up. Of course they will learn the national language later on anyway playing with their peers and at school, but if they have not always spoken it they won't come to recognize it as their mother tongue. The situation is different if the family intends to move in the near future, if, for example, an Italian father wants to return to Italy with his German wife and child. In such a case, it would of course be good if mother and child got used to the Italian language early on. Often, the foreign father will also wish his children to speak his, the father's language, even when he has left his homeland forever. In such a case, the father should try to put aside his understandable feelings in favour of his child's interests. (translated/cited by Saunders 1988)

While it is unlikely that Youth Welfare Departments still openly voice recommendations like this today, it is more than a fair guess that many childhood educators and teachers throughout the educational system, and even parents themselves, embrace it. Even among those who no longer think that humans need "one" language to "feel at home in" and to avoid linguistic insecurity, there is still a strong sense as to which languages are considered worth acquiring or supporting. The opinion that "other" languages are obstacles to learning majority languages and that their use can be interpreted as signs of failed integration or reluctance to integrate may even be more prevalent nowadays than ever, boosted by the outcomes of large panel studies testing language-related skills (OECD, PISA, etc.). Despite best efforts of research-driven initiatives (like BIVEM), of parents' associations, and of highly motivated educators, many families still feel short of answers four decades after the advice quoted above. They often worry about language-related questions such as how to do "the right thing" and how to enable their children to become successful members of the society they live in without losing the language needed for communicating with family members and friends living in the countries of origin or in other parts of the world. Even people parents consider authorities and turn to, i.e. early childhood educators, teachers, and paediatricians, are often at a loss with respect to providing informed answers, except for those cases where language and multilingualism-related topics have been integrated into curricula of educator and teacher training.

So who, if not linguists with expertise in acquisition, language use, and language change, could and should have an important message to convey to all these stakeholders? Yet, more is needed than expertise and willingness to engage with various interest groups. In Germany, like in many other countries, knowledge transfer (as mentioned in the Higher Education Framework Acts (“Hochschulrahmengesetze”)), is one of three missions, alongside research and teaching. While knowledge and technology transfer have been integral parts of the interface between the Applied Sciences and non-academic institutions and companies, the liberal arts or humanities have only gradually recognized what they have to offer in terms of relevant research findings, thereby also contributing to their departments’ and faculties’ “impact”. Experience with third-mission activities (community engagement, voluntary work, and the like) are often included in the list of desirable job qualifications. Increasing awareness of what linguistics has to offer by way of transfer can also be seen in the rise of funding programs through federal and local ministries of research, foundations, new conferences, academic and parental associations, and private organizations, for instance family networks across languages or for specific heritage languages. For a number of years, federally funded research initiatives – like RUEG – successfully received transfer project funding. This way researchers could engage in cooperations with a partner outside academia, a point to which we will return.

In the following, “transfer” will be used as an umbrella term for activities aiming at sharing knowledge and expertise between researchers and non-academic partners. According to Gonser and Zimmer, these activities typically cover three main aspects of engagement. They distinguish communication (the presentation of results and materials for the public), consultation (the adaptation of scientific knowledge for use in concrete practice contexts) and cooperative actions, where further knowledge is jointly produced via cooperation between researchers and practice partners (Gonser & Zimmer 2020: 19). Gluns (2020: 249) highlights the opportunities inherent in the latter, since in a “functionally diversified society, no actor is in possession of all knowledge relevant to a specific area like migration. Each actor has instead a particular perspective on the topic and can contribute implicit (experience-based) knowledge and/or scientific knowledge” (transl. by the authors).

In a very concrete sense, usefulness of and need for transfer activities in the language domain should be evident. After all, spoken, written, or signed: Whoever wants to reach out to potential audiences or cooperation partners has to use the very medium that is simultaneously part of the challenge and of the solution. This also means that all speakers (writers, signers) have first-hand experience to contribute. As competent speakers of at least one language, they are indispensable sources of data and can provide intuitions about formal well-formedness,

paraphrases, ambiguity, as well as more or less appropriate formal and informal conditions of use in different registers. Regardless of specific sub-disciplines, empirical research with living subjects acutely depends on cooperation, i.e. on more or less collaborative contributions by those investigated. Hence, what used to be called “subjects” or “informants”, more recently “participants”, can quite rightly be considered “consultants”. This especially holds true for cases where linguists crucially depend on competent speakers providing data and reactions to specific questions – e.g. in the case of field work with unwritten languages. Moreover, in applied linguistics, given its concerns with ‘real-world problems’ and the forms and functions of verbal interaction, the move to explicit audience involvement should be a natural step to take. By drawing on the current project as an example, we will proceed to outline challenges and opportunities that such transfer initiatives entail.

### **3 Methodological reflections: Setting out to outreach**

Already before the start of the RUEG transfer project focused on here, RUEG members – individually or jointly – had been contributing talks and workshops on multilingualism/heritage languages at conferences and for special interest groups like childhood educators, teacher and parents’ organisations, speech and language therapists, pediatricians and regional governments and foundations. Our previous involvement in various transfer activities had provided us with a network of partners to build on.

Reaching audiences outside of academia requires long-term investment, and given the often mobile nature of academic employment and changes in administrative staff in schools and other local agents, assuring continuity of tried and tested cooperations is as challenging as it is necessary. The transfer activities described here profited greatly from the fact that several team members already brought various academic and non-academic, local and international networks into the project, including both national and international academics who agreed to act as consultants and cooperation partners and, in some cases, co-authored papers. In order to establish cooperations, persistence and some flexibility in timing are necessary – in particular when it comes to matching plans for events with the partners. The RUEG transfer project’s main practice partners are aim (Akademie für innovative Bildung und Management/Academy for innovative education and management), Heilbronn, and the Servicezentrum der Berliner Volkshochschulen. By choosing to work with them, we opted for highly skilled and rather large partner organisations. Both our partner organizations are deeply invested in education, in the case of VHS Berlin of parents and the general public

and in the case of aim Heilbronn mainly of educators and teachers. Both institutions offer a number of language-related courses and trainings. Our partners' contributions were thus not just a prerequisite for funding on part of the DFG, but also indispensable in terms of reaching our diverse target groups.

Within the overall RUEG context, research findings on the 'dynamics' of Turkish, Russian, Greek, and German as heritage languages and shifting patterns of language use in diaspora settings, both in Germany and in the United States (see the other chapters, this volume) provided data for exemplification. At the same time, our activities were guided by what parents, teachers and language fostering staff themselves considered relevant questions and concerns. In order to tailor activities and materials accordingly, we started with a needs analysis. Questionnaires distributed via our two partner organizations were filled out by about 50 parents enrolled in classes of the Servicezentrum der Berliner Volkshochschulen (Service Centre for Adult Education Berlin) and by about the same number of instructors of courses for early childhood educators and language support staff employed by the Akademie für Innovative Bildung und Management (aim, Academy for Innovative Education and Management) in Heilbronn, our Southern partner.

The questionnaires consisted of closed questions about attitudes towards multilingualism (i.e. 'In which contexts do you use your family language?'), open questions about the language use of the speakers on a daily basis (i.e. 'Which languages do you consider part of your everyday life?'), and they addressed either issues related to language transmission and use within the family in the case of our Berlin partners. Questionnaires targeting instructors revolved around issues they were confronted with in their own classroom contexts.

In the following, we mainly draw on open questions such as – here roughly translated – "How do you feel when you speak your different languages?", "What kind of language-related topics do you discuss in your family?" and "Are there questions about multilingualism that you yourself or participants in your classes would be interested in?". Based on responses given, we identified repeatedly voiced issues as well as topics that we could confidently address on the basis of available evidence. In a more practical section, we also asked questions related to personal learning strategies and preferred learning contexts and materials, including videos, texts, and classes. Overall, responses provided in the questionnaires demonstrated considerable awareness of the crucial role of the family as the first major agent in language socialization. Responses also revealed a high degree of insecurity in how to deal best with language-related issues and conflicts in the family. Therefore, we next turn to family language dynamics in Section 4 and to educational contexts in Section 5. Both times we take as our starting point the information gained through our questionnaires.

## 4 Family language dynamics: Linguistic resources and repertoires

Families, qua primary caregivers, are the most crucial players in early language socialization. They provide the input relevant for the dynamics of language acquisition to unfold in the direction of specific languages. The plural in ‘languages’ is fully intended here since there may well be more than one language involved right from birth. In addition to our biological predisposition and cognitive abilities, caregivers’ input (spoken or signed) supplies what is needed for L1 acquisition to take off. Whether and how development continues beyond early childhood depends on exposure as well as on the minority or majority status of the languages concerned. Sooner or later, multilingual parents tend to raise, or are confronted with, questions of how best to support their children’s development (Lanza 2004). The spectrum of circumstances in which languages are acquired is basically without limits, especially if one considers exposure-related quantitative and qualitative variation: Is there, – to take one of the questions addressed in parent classes – a difference in outcome between children exposed to a specific language for only two to three hours a day vs. a full day a week? We simply don’t know. Questions of such granularity may be irrelevant, anyway. What we do know, however, for a number of languages, is which patterns (e.g. with respect to word order, morphological and phonological subsystems) can get firmly established in language A despite competition from language B. We also know what properties are harder to figure out, even regardless of language contact, and which therefore need more exposure time and more access to specific contrasts in the input, even in the monolingual (cf. Tsimpli 2014, Tsehayé et al. 2025 [this volume]). Research on the combination of different heritage languages within the same majority context, and attention to individual differences in exposure (due to varying family size and differences in family language policy) therefore offer a natural laboratory for investigating both acquisition processes and outcomes. In order to illustrate parental attitudes, we selected written examples from the questionnaires collected in Berlin. We quote the German original and provide a rough English translation. Excerpt (1), mother:

- (1) ich fühle mich total entspannt und es ist mir total lustig hören wie meine Tochter gleichzeitig beide Sprachen lerne  
 ‘I am completely relaxed and I find it completely funny how my daughter learns both languages at the same time.’

Parents adopt very different family language policies (Romaine 1989, Gawlitzek-Maiwald & Tracy 1996, Lanza 2004, De Houwer 2007, Tracy 2008,

Müller et al. 2018), often building on their own experiences of multilingual upbringing (Purkarthofer & Steien 2019). While some families decide on one person-one language strategies or select one language as a common family language used by both parents, others – especially if they feel confident in both languages – may freely alternate. However, initial decisions in favour of a particular policy do not preclude changes later on, especially if children react differently than what parents hoped for (Gawlitzeck-Maiwald & Tracy 1996). An often disheartening experience on part of parents arises when children refuse to speak one of their languages even though they are likely to react appropriately when addressed in it. In one questionnaire, a mother provided a clear description of her feelings in this respect. Excerpt (2), mother:

- (2) Ich bin enttäuscht, dass meine Kinder nicht mehr so gut Türkisch können.  
Sie verstehen manche Wörter nicht  
'I am disappointed that my children cannot speak Turkish so well any longer. They don't understand some words.'

While shifts – temporary or permanent – in children's language preferences may occur in all bilingual settings, minority languages which neither enjoy the status of a language taught in schools nor peer or community support are harder to cultivate. Hence, it is important to reach out to families, for instance by setting up regular meetings involving parents and kindergarten teachers, where games and activities aimed at fostering the majority language can then be replicated in diverse heritage language settings at home (Tracy et al. 2009; for an initiative of bringing parents into classrooms, see cf. Prasad 2017). Another parent, speaking about Italian as a heritage language, considers change inevitable. Excerpt (3), mother:

- (3) In diese Zeit ist es mehr wichtig für mein Kind Deutsch zu sprechen und schreiben zu lernen. Er kennt sehr gut Italienisch, aber ist wichtig für andere italienische Kinder zu besuchen.  
'At this point in time, it is more important for me that my child learns to speak and write German. He knows Italian very well but it is important for him to see other Italian children.'

From this parent's perspective, language use is necessarily tied to interaction, and she recognizes that heritage language maintenance need peer-support. In light of the societal attitudes towards bi- and multilingualism, family languages – heritage languages of immigrants as well as many indigenous minority languages and regional dialects – are threatened (cf. overviews in Baker 2011,

Brehmer & Treffers-Daller 2020). However, there are cases of the successful revitalization of indigenous languages by conscious efforts and investments in the education system and the consultation of families. In Wales, for example, midwives provide first language guidance on the bilingual upbringing of new-born children (Edwards 2015), and in New Zealand and Hawaii, family practices are actively supported by early childhood education, schools and immersion-oriented ‘language nests’ (Hinton 2013). These early childcare facilities operate with a strong language policy to foster the minority language and provide language input for children and parents alike by setting a positive example. This is particularly relevant where parents themselves were no longer in close contact with non-emigrated family members. In other cases, heritage languages are used as the main language of encounter and have served as a mediating lingua franca between family settings and a multilingual community. In recent years, due to war-driven migration to Germany, we have seen an added awareness of i.e. Arabic as a heritage language since speakers and community infrastructure were crucial to deal with needs of recently arrived refugees, at times including unaccompanied minors. Teachers and educators in early childcare are important target groups for transfer activities since parents often consider them authorities with respect to developmental and educational issues. The same goes for medical professionals as key figures who, for instance in regular medical checks, assess the ‘typical’ development of pre-school children. Even though medical professionals rarely have access to up-to-date information about language acquisition and even less so on specific questions of simultaneous bilingualism, their opinions are crucial, for instance when it comes to decisions on speech therapy. Not surprisingly, various projects have shown that there is a considerable risk of over- and under-diagnosing multilingual children and their need for therapy (e.g. the CAMINO and MILA projects by Schulz 2013, Voet Cornelli et al. 2013, Tracy et al. 2018). In the following we provide a brief outline of educational contexts impinging on family language dynamics. Insight into these linguistic landscapes helps us identify the potential for outreach and transfer.

## **5 Family languages in educational contexts**

Among the resources available and relevant for the development of family languages are all kinds of educational institutions (cf. Polinsky & Kagan 2007, Baker 2011, Kasstan et al. 2018, Montrul 2018, Polinsky 2018, Gagarina & Milano 2021). For parents, the availability and goals of heritage language education can be crucial in finding their own position towards multilingual family language practices. In Germany, few heritage languages feature in the mainstream education system:

the bilingual European Schools with different heritage languages in addition to German (e.g., Italian, Portuguese, Turkish, and Russian) started in the 1970s but have since not found a high number of followers (Niedrig 2001). A recent publication addressed Turkish and Russian as particularly important heritage languages in Germany, featuring to some extent in the educational system (Yildiz et al. 2017). With respect to Italian speakers in Germany, Caloi & Torregrossa (2021) also stress the importance of school-based support for heritage language maintenance.

For students whose home languages are not school languages, HL education is organised under regional regulations and often depends on funding from the countries where these languages have majority language status. Teaching is typically focussed on standard majority contexts, e.g. Russian spoken in Russia, aiming at ensuring children's competence in registers which they may not really perceive as "their" heritage due to social or dialectical variation (Wörfel et al. 2020). Therefore, it seems that teachers in heritage language education and in majority language classes as well as in preschool education need to develop a wider understanding of bilingual language acquisition and multilingual learning (Baker 2011, Krifka et al. 2014, Seals & Olsen-Reader 2019). In addition, the use of vernaculars and (often imagined) connections between social groups and linguistic resources raise relevant questions which have not been answered, e.g. for Germany: How are vernaculars present in the school context (Nero & Ahmad 2014)? In which ways are social distinctions addressed with regard to heritage languages (Flores & Rosa 2015)? How can discrimination based on assumed immigrant or language background be addressed (Bonefeld & Dickhäuser 2018)? How should schools respond to political demands requesting constraints on language use (Wiese et al. 2020)? In the questionnaires distributed among lecturers running classes for educators in Heilbronn, we see the same issues and questions addressed. Excerpt (4), lecturer:

- (4) Wie sich der Loyalitätskonflikt minimieren lässt, in dem sich Kinder befinden, deren Eltern sich in Deutschland nicht wohlfühlen. Weil Kinder das spüren, kann es dazu führen, dass sie ihre eigene Sprachentwicklung blockieren, um ihren Eltern nicht in den Rücken zu fallen. Letztlich steht dahinter eine Integrationsfrage. Wie kann Integration besser gelingen, vor allem bei Personen, deren Aufenthaltsstatus nicht geklärt ist.

'How can the loyalty conflict be minimised that affects children whose parents do not feel at home in Germany. Because children can feel that and this in turn can lead to them blocking their language development in order to not turn against their parents. But in principal, this is a question of integration. How can integration work better, in particular for persons whose status of residence is unclear.'

Language skills are indispensable for all school contexts. Regardless of subject, teachers need to be aware of typical language-related ‘hidden’ barriers to comprehension and learning. Beyond this general requirement, insights into types of language acquisition and into properties of heritage languages are necessary for students in teacher training programs in Germany. On the whole, the following players can be identified as immediately relevant for language-related concerns: 1) preschool educators (who are in charge of language fostering and increasingly of language-related diagnostics), 2) teachers of German as a school subject, including German as a second and as a foreign language, 3) teachers of heritage languages studying in university programmes (for example at the University Duisburg Essen, and at Berlin universities where such a program is planned), 4) foreign language teachers of typical school languages (English, French) but also of other subjects, among them increasing numbers of heritage speakers (which is, for instance, the case in Mannheim, where a curriculum for future highschool teachers and teachers of professional schools (“Berufsschulen”) has been established, taking into account the linguistic heterogeneity of both highschool students and their future teachers (cf. Karst et al. 2021), 5) teachers in training programs for other languages, where no institutionalised teacher training exists (e.g., Arabic, Kurdish), and, finally 6) teachers brought in from other countries for complementary schooling (e.g. HL education in Saturday schools) (Wörfel et al. 2020). These target groups are important as they work with children and adolescents, and have ample opportunity and legitimate reasons for reaching out to parents. Their take on speakers and learners of any language, including heritage languages, impacts children, their families and communities (De Korne 2017, Seals 2018). Awareness of acquisition processes in general and of heritage language development in Germany (and in other countries) prepares teachers for their multilingual teaching realities. It also offers an additional basis for reflection to the growing number of future teachers who are themselves heritage language speakers. However, this latter fact does not automatically make them more efficient or insightful teachers since they themselves have often internalized stereotypes which have to be addressed (Thoma & Ofner 2021). Excerpt (5), lecturer:

- (5) Nicht wenige Teilnehmende gehen davon aus, dass mehrsprachig aufwachsende Kinder, vor allem wenn ihre Familiensprache(n) keine der Bildungssprachen des globalen Nordens ist, weniger verstehen.  
Bildungssprachliche Aspekte des (dialogischen) Vorlesens gehören deshalb fast immer dazu.

‘A relevant number of participants assumes that multilingual children whose family languages are not from the Global North understand less.

Therefore, aspects like (dialogically) reading to them are almost always needed.'

Transferring research results to educational settings and to encounters with parents calls for an approach that takes speakers' agency seriously and shows awareness of their multilingual environments. In the following section, we expand on specific strategies for working with educators and parents.

## **6 Family languages in outreach and transfer: Potential momentum**

### **6.1 Starting from an information hub**

In this section, we describe specific proposals to strengthen heritage speakers' resources, firstly through communication and consultation and secondly through citizen science activities. We set out with concrete insights into the website that serves as our information hub and will then go in detail about the categories that structure our teaching and learning material.

Educational resources and a modular curriculum for workshops with parents of children in majority or heritage language education are our main tool to transmit knowledge. In order to illustrate the approach taken in communication and consultation activities, a screenshot from the website can be found in Figure 2.

Highlighted in this screenshot is the link to resources on multilingual speakers' experiences, as in the case of a video with an international researcher and Mercator fellow of the RUEG group who speaks about her own growing-up as a heritage speaker of Portuguese in Germany.<sup>1</sup> She explains how she experienced heritage language education as a child, and she also comments on the fact that her bilingual upbringing helped in the pursuit of her career as a professor of German in Portugal (see Figure 3). The video is recorded in German, with subtitles available in German, Ukrainian, Greek, and English, with more languages to be added.

In order to engage viewers with the contents of the video, activities, e.g. a set of questions or input for discussion for children and adults, are presented. Visitors can access the site themselves, but educators can also make use of the videos and activities suggested to e.g. initiate a discussion during a parents' meeting. Additional resources are provided, such as a chart on languages in the German

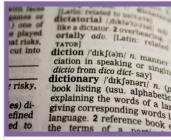
---

<sup>1</sup><https://www.uni-due.de/germanistik/rueg/videodrei.php>

 **For parents**  
RUEGram

Languages play an important role for multilingual families and for their friends. As a parent, you will certainly find the following questions interesting: When and how do I speak in my family language? How do my children, my parents or other family members speak? How can I pass on family languages and how do they change? We are dealing with these and similar questions in our transfer project "Familien und ihre sprachlichen Dynamiken" ("Family language dynamics"). When working with parents, we use their everyday observations of their own languages and combine them with new findings from research.

We publish texts, images, videos and audio material on our open-access website [www.ruegram.de](http://www.ruegram.de). They can be accessed for free by anyone who is interested! We develop these materials together with partners from the field. The materials are tested by our students and are being expanded through Citizen Science Projects. We will provide information in German, English, Greek, Russian and Turkish on the following topics:

|   |   |   |   |  |
|---|---|---|---|--|
|  |  |  |  |  |
| Knowledge about languages, language acquisition and language use                  | Current research findings on language use in contact situations                   | Language experience of multilingual speakers<br><br>Material →                    | Myths and discourses about languages and multilingualism                          | Research in one's own biography, family and environment                            |

Do you have any other questions about language and languages that you want answers to?

Feel free to visit [www.ruegram.de](http://www.ruegram.de) frequently – you'll surely find interesting answers to questions about your everyday multilingual life.

Figure 2: Entry page for parents with links to the five categories coded in the respective colours.

school system that can inform parents about the status of different languages. Alternatively, specific information on heritage language education is available for contexts where parents would want to make decisions about signing their children up for classes. As an entertaining activity for all, a quiz is added with facts and myths about heritage languages (e.g.: "Who is considered a native speaker?", "How many speakers of heritage languages live in Germany?", etc.). The quiz can be taken individually or in a group.

Finally, website links are given to activities that invite participants to think about their own upbringing and to reflect on their experience with schooling in more than one language. These citizen science explorations can, if a speaker wishes to do so, be uploaded and added to the collection of the project (as is described in Section 6.2).

All information, tasks, and discussion prompts are included in at least two or three languages, thereby making the materials suitable for diverse audiences. In particular, the subtitling helps groups to work on the same videos but to access them according to language preference.



Figure 3: Example of a video recording of a heritage language speaker reflecting on her experiences.

## 6.2 Communication and consultation

Our encounters with large numbers of heritage speakers interested in their own family languages encouraged us to widen the range of information offered and to encourage speaker groups to select tools specific to their languages and professional background. This information was put in place through a multilingual website that offers access to different topics.<sup>2</sup> The texts, videos, and audios are accessible via guided paths, and they are framed differently depending on languages chosen, roles (parents, teachers, children, adolescents) and media format. Each group is addressed explicitly through introductory texts, and the language used in descriptions is adapted in length and complexity. Certain activities are specifically designed for children, but many are just framed differently in order to respond to needs in a particular moment in life. Short video clips of multilingual speakers, most of them from the RUEG project group, offer an accessible

<sup>2</sup><https://www.ruegram.de>

way to understand language repertoires in specific contexts and to engage in discussions from there. In general, we cover the five categories already presented in the colourful pentagon above, with some detail following.

### **6.2.1 Knowledge about languages and language use**

What early childhood educators need and interested lay persons, for instance parents, typically want to know are basic facts about language acquisition, typical developmental phases, and the relevance of input. They also need to know that bilingual children may go through phases of intensive language mixing and that this is neither chaotic nor an indication of identity or language confusion. Extensive previous experience with workshops geared to parents, childhood educators and elementary school teachers have shown that they can be demonstrated the linguistic and metalinguistic abilities of very small children (Tracy 2008, Voet Cornelli et al. 2013).

### **6.2.2 Research findings illustrating heritage speakers' language use 'live' in language contact situations**

Illustrations of language contact phenomena that are highly accessible to non-linguists typically include the borrowing of words from the majority language and code-switching as normal bilingual practices. What does the set of RUEG projects contribute to these contents? In addition to what is known from previous research, transfer activities include specific illustrations from language combinations consisting of German or English and the heritage languages Turkish, Greek, Russian in Germany and in the US, including German as a heritage language in the US. Also, families speaking heritage languages are typically interested in learning about how their own language and language use have changed since immigration. Thus they can be shown in what respect the languages in their countries of origin have changed independently of the influence of language contact. Among the changes most easily demonstrable are changes in word order and morphological subsystems as well as the adoption of specific discourse markers, as described in detail in other chapters of this volume.

### **6.2.3 Multilingual speakers' experience**

As long as monolingualism is considered a perfect and the most natural state of the human mind, bi- and multilinguals are likely to continue doubting their own

abilities. Non-experts easily find their own behavioural repertoires deficient, unaware of non-selective lexical access and the natural competition between simultaneously activated resources. Among the many experiences typical for heritage and also dialect speakers is the sense of having been shamed for inappropriate or supposedly flawed language use. Heritage speakers of any age are often faced with negative expectations, underestimation, discrimination, and unfair evaluations based on suspected immigration background (Bonefeld & Dickhäuser 2018). Faced with specific texts, teachers often don't realize that what is heard or read may be based on a particular informal register, and that this does not mean that the producer of this text was incapable of formulations in other styles or registers (Keim & Tracy 2007, Keim 2008).

#### **6.2.4 Myths and discourses about languages and multilingual language use**

Throughout this text, various misconceptions were mentioned. Some reflect opinions no longer supported by the current state of the art of multilingualism research, others were ideological from the start. This module therefore highlights particularly persistent beliefs, such as the claim that bilingualism results in deficits in both languages with monolingualism being the “ideal”, and “more natural” state, that language mixing is a sign of incompetence, that children should not have to deal with more than one language in early childhood, that only one language should be acquired first, that people speaking a language we don't know in our presence must want to hide something from us or talk about us (Wiese et al. 2020). On the other hand, there certainly are advantages of sharing the privilege of a private code, already appreciated by bilingual children who early on speculate on who speaks what (Tracy 2008).

### **6.3 Citizen science**

One important goal of the transfer project was the involvement of interested people of all ages as lay researchers. To qualify as Citizen science (CS), research projects have to involve lay persons in data collection or analysis, or, as Rymes (2020: 5) puts it: “Citizen science is the study of the world by the people who live in it [...].” Golumbic and colleagues identify three key aspects of CS, namely (1) inclusion of citizens in the scientific process through shared activities; (2) contributions to both science and the public, and (3) reciprocity of information in both directions, between researchers and stakeholders outside academia (Golumbic et al. 2017: 6). In their own study about researcher investment in CS, researchers were positive towards the idea that audience members can be educated through

taking part in transfer activities and in particular in citizen science projects. Not surprisingly, they were met with scepticism on the part of other researchers who doubted that untrained members of the public can actually contribute to science. Moreover, engagement with the public was often considered a time-consuming activity of lesser value than doing “actual science”. Robinson et al.’s study summarises many challenges faced by CS projects across disciplines, and scientists have since addressed them, for example by establishing principles CS projects should adhere to (Robinson et al. 2018). These include more than a commitment to making the audience participate in data collection: They call for explicitly acknowledging and crediting work and creative contributions, and to care for scientific quality and credibility. Freitag et al. (2016) published on strategies of CS programs in order to increase data quality and found this to be a topic of considerable concern for research projects dealing with environmental issues on Californian beaches: “While all science can face challenges to its credibility, the specific context of citizen science, along with external assumptions about citizen science, can make establishment of its credibility particularly difficult. Implementing science projects with volunteers, often outside traditional science institutions and typically with limited resources, contributes to the credibility challenge. Preconceptions about citizen science also can be an issue.” (Freitag et al. 2016: 2) In light of this, CS is necessarily embedded in discourses of scientific quality, and throughout the research process reflective practices should be included.

In the humanities, CS has been less of a topic (Shirk & Bonney 2018), with the exception of digital humanities that have, for example, used reading skills of the public for deciphering old texts. For linguistics, advantages are not hard to see, for instance with respect to questions related to sociolinguistics (Rymes & Leone 2014, Svendsen 2018, Rymes 2020), or “the study of the world of language and communication by the people who use it and, as such, have devised ways to understand it that may be more relevant than the ways professional sociolinguists have developed” (Rymes 2020: 5). In our case, CS initiatives create opportunities for speakers of heritage and majority languages to actively engage with scientific results and to reflect on their own environment and communicative practices.

In addition to the activities sketched in Section 6.2, with their focus on research-backed and yet necessarily low-threshold knowledge transfer about language and multilingualism, we successfully piloted CS activities suitable for future roll-out in cooperation with schools and other relevant agents, such as our cooperation partners. In a first step, data collection tools were made available on our website, followed by demonstrating how to use them in university classes, schools, as well as with parents.<sup>3</sup>

---

<sup>3</sup>RUEGram citizen science: [https://www.uni-due.de/germanistik/rueg/citizen\\_science.php](https://www.uni-due.de/germanistik/rueg/citizen_science.php).

Not only do we rely on participants using the website on their own but we also introduce CS tasks in our workshops and teaching. Students are encouraged to provide and work with CS data as part of their course assignments, educators are invited to bring their own collections of pictures to our workshops. Our experience shows that this also promotes discussions on data quality, methodology, and awareness of one's own role and responsibility as researcher. Students and teachers who previously felt rather timid about trying out projects and methods on their own profit greatly from the supervised use of online tools made available. In a playful approach to linguistic research and language repertoires we invite participants to contribute their own experience, expertise and ideas. The topic "Research in my own biography, family, and community" contains six different tasks, as stated below. Each task is introduced by a short video (subtitled in up to six languages) and linked to texts that contain explanations and further material necessary, like an empty silhouette or a short questionnaire.

1. Language portrait: Participants are encouraged to map their language resources onto a body silhouette and to write and reflect on their own language biography, on language use in the family but also in educational and professional settings.
2. My language story – multilingual and multimodal narration: This task asks participants to upload their story as text/audio/video and to talk/write about their multilingual or monolingual experience.
3. My family language biography: Data about language genealogy is collected with the help of structured questions. This task is of particular interest in combination with the (1) portrait, (2) story, or (5) photo elicitation.
4. Lived experience of language: This task asks for a text about one episode or event linked to growing up with two or more languages. An example would be: 'Write about one event where your languages helped you cope with a situation.'
5. Photo elicitation: Participants are asked to take four pictures from their environment that are closely linked to their family language(s). They are also asked to write a short explanation with each picture.
6. Citizen science census: Participants familiarise themselves with current micro-census practices and contribute their own anonymised family language profile. Blueprints for mapping activities in speakers' own environments are discussed.

Individual CS researchers can then either complete tasks for themselves or involve family members for answers. Once drawings are completed, texts written and questionnaire answered, the website provides an upload form where participants enter a self-selected code and upload their data. This way, we encourage engagement with the website and expand the overall data pool for new CS activities. Via the self-selected code, data from the same person can be linked and analysed on the basis of methods developed for language biographical projects (e.g. Purkarthofer 2016) and with regard to language use in specific languages, registers and modes. In contrast to the standardized data elicitation procedure for the RUEG corpus (see Wiese et al. 2025 [this volume]), our CS activities encourage contributions in a multilingual mode and hence language mixing. As the data collection is still ongoing, results cannot be published yet, but an overview of tasks, corresponding instructions, and tools is available.

## 7 Conclusion

This contribution dealt with steps undertaken in a specific transfer project in collaboration with two “practice” partners who were already deeply involved in adult education in two areas of Germany. What the members of the RUEG transfer project and, with them indirectly, the whole research group brought into this cooperation was state-of-the art insight into current research on multilingualism, both from a sociolinguistic and a psycholinguistic perspective, and previous experience with outreach and public relation activities.

In the domain of language, transfer projects are not just compelling but also self-perpetuating because they can draw on readily available or arousable interest of speakers and their experience of themselves as speakers of at least one language. From early childhood on, speakers are interested in language, enjoy language play, and prove highly sensitive to the way they are spoken to and comment on language use (see Tracy 2008, Tracy & Gawlitzek 2023). Lay people can easily understand that building up rich repertoires for expressing themselves presupposes encounters with situations in which these repertoires are needed and that the environment must make an effort to provide these learning opportunities. It clearly works in our favour that communication and language use are among our favourite activities, as very aptly formulated by Levelt (1989: 169): “We, homo sapiens, are fanatic speakers. Most of us talk for several hours a day, and when we are not chatting with others, we are probably talking to ourselves.” For those living with more than one language, talking to themselves in more than one language is part of their multilingual reality and to be expected, as is

occasional competition between resources co-activated. Transfer initiatives in the domain of language appeal to what people know and can do. This differs from other academic areas where researchers convey results of their work to the public. Thus they can make explicit remarkable skills both readily available and ‘by their very nature’ of interest to the individual. For anyone wishing to empower individuals, families, educators, and our already multilingual societies, the importance of this message cannot be overstated.

## Acknowledgements

The support of the overall RUEG team is gratefully acknowledged, this includes our internal reviewers Heike Wiese and Natalia Gagarina as well as the external reviewers. Heartfelt thanks to our research assistants Anne Mölders, Geylan Ahmed Daud, Özge Zar, Serpil Kuzay, Diana Lehl, and Bahar Yilmaz, as well as to Lea Coy for help with pre-publication formatting.

The authors thank all participants in outreach and citizen science activities: as learners and parents, as educators and teachers, as students and creative co-workers and finally as multilingual subjects. The support of other RUEG projects is gratefully acknowledged. Most importantly, without the external cooperation partners in Berlin and in Heilbronn, their investment of time and resources, including access to their target groups, the project would not have been funded by the DFG, and without their excitement and contribution of ideas it would certainly have been less exciting to run.

The research was supported through funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit *Emerging Grammars in Language Contact Situations*, project Pt (grant number: 313607803).

## References

- Arnaus Gil, Laia, Natascha Müller, Marina Hüppop, Meike Poeste, Elena Scalise, Nadine Sette, Abira Sivakumar, Mabel Tirado Espinosa & Katharina Sonja Zimermann. 2019. *Frühkindlicher Trilinguismus: Französisch, Spanisch, Deutsch* (Narr Studienbücher). Tübingen: Narr Francke Attempto.
- Arnaut, Karel, Jan Blommaert, Ben Rampton & Massimiliano Spotti (eds.). 2016. *Language and superdiversity*. New York; London: Routledge.
- Baker, Colin. 2011. *Foundations of Bilingual Education and Bilingualism*. 5th edn. (Bilingual education & bilingualism 79). Bristol, UK; Tonawanda, NY: Multilingual Matters.

- Bonefeld, Meike & Oliver Dickhäuser. 2018. (Biased) Grading of students' performance: Students' names, performance level, and implicit attitudes. *Frontiers in Psychology* 9. 481. DOI: 10.3389/fpsyg.2018.00481.
- Brehmer, Bernhard & Jeanine Treffers-Daller (eds.). 2020. *Lost in transmission: The role of attrition and input in heritage language development* (Studies in Bilingualism (SiBil) 59). Amsterdam: John Benjamins.
- Brisk, María Estela. 1998. Bilingual education: From compensatory to quality schooling by María Estela Brisk. *The Modern Language Journal* 83(4). 601–602.
- Busch, Brigitta. 2012. The linguistic repertoire revisited. *Applied Linguistics* 33(5). 503–523. DOI: 10.1093/applin/ams056. (15 August, 2023).
- Caloi, Irene & Jacopo Torregrossa. 2021. Home and school language practices and their effects on heritage language acquisition: A view from heritage Italians in Germany. *Languages* 6(1). 50.
- De Houwer, Annick. 2007. Parental language input patterns and children's bilingual use. *Applied Psycholinguistics* 28(3). 411–424. DOI: 10 . 1017 / S0142716407070221.
- De Korne, Haley. 2017. "A treasure" and "a legacy": Individual and communal (re)valuing of Isthmus Zapotec in multilingual Mexico. In Mi-Cha Flubacher & Alfonso Del Percio (eds.), *Language, education and neoliberalism*, 37–61. Bristol, Blue Ridge Summit: Multilingual Matters. DOI: 10.21832/9781783098699-005.
- Edwards, Viv. 2015. Diversity lessons from Wales: How marketing bilingualism contributes to language shift and language maintenance. In Almut Küppers & Çiğdem Bozdağ (eds.), *Doing diversity in education through multilingualism, media and mobility*, 18–20. Istanbul: Istanbul Policy Center.
- Fishman, Joshua A. 1991. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages* (Multilingual Matters 76). Clevedon, PA: Multilingual Matters.
- Flores, Nelson & Jonathan Rosa. 2015. Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review* 85(2). 149–171. DOI: 10.17763/0017-8055.85.2.149.
- Freitag, Amy, Ryan Meyer & Liz Whiteman. 2016. Strategies employed by Citizen Science programs to increase the credibility of their data. *Citizen Science: Theory and Practice* 1(1). DOI: 10.5334/cstp.6.
- Gagarina, Natalia & A Milano. 2021. Grammatical development within a context of early bilingual education and language support programs. In Mila Schwartz (ed.), *Handbook of early language education*, 2–34. Cham: Springer. DOI: 10 . 1007/978-3-030-47073-9\_4-9.
- Gawlitzek-Maiwald, Ira & Rosemarie Tracy. 1996. Bilingual bootstrapping. *Linguistics* 34(5). 901–926. DOI: 10.1515/ling.1996.34.5.901.

- Gluns, Danielle. 2020. Transfer in der Fluchtforschung: Perspektiven. In Monika Gonser, Karin Zimmer, Nicola Mühlhäuser & Danielle Gluns (eds.), *Wissensmobilisierung und Transfer in der Fluchtforschung: Kommunikation, Beratung und gemeinsames Forschungshandeln*, 247–260. Münster New York: Waxmann.
- Golumbic, Yaela N., Daniela Orr, Ayelet Baram-Tsabari & Barak Fishbain. 2017. Between vision and reality: A study of scientists' views on citizen science. *Citizen Science: Theory and Practice* 2(1). 6. DOI: 10.5334/cstp.53.
- Gonser, Monika & Karin Zimmer. 2020. Wissensmobilisierung und Transfer in der Fluchtforschung: Begriffsbestimmung und Bestandsaufnahme. In Monika Gonser, Karin Zimmer, Nicola Mühlhäuser & Danielle Gluns (eds.), *Wissensmobilisierung und Transfer in der Fluchtforschung: Kommunikation, Beratung und gemeinsames Forschungshandeln*, 13–31. Münster: Waxmann.
- Hayakawa, Samuel I. 1992. Bilingualism in America: English should be the only language. In Gary Goshgarian (ed.), *Exploring language*, 42–47. Harper Collins.
- Hinton, Leanne (ed.). 2013. *Bringing our languages home: Language revitalization for families*. Berkeley, CA: Heyday.
- Hopp, Holger, Dieter Thoma & Rosemarie Tracy. 2010. Sprachförderkompetenz pädagogischer Fachkräfte: Ein sprachwissenschaftliches Modell. *Zeitschrift für Erziehungswissenschaft* 13(4). 609–629. DOI: 10.1007/s11618-010-0166-z.
- Karst, Karina, Dieter Thoma, Julia Derkau, Jürgen Seifried & Stefan Münzer (eds.). 2021. *Lehrer\*innenbildung im Kontext leistungsbezogener Heterogenität und Mehrsprachigkeit von Schüler\*innen*. Münster New York: Waxmann.
- Kasstan, Jonathan R, Anita Auer & Joseph C. Salmons. 2018. Heritage-language speakers: Theoretical and empirical challenges on sociolinguistic attitudes and prestige. *International Journal of Bilingualism* 22(4). 387–394. DOI: 10.1177/1367006918762151.
- Keim, Inken. 2008. *Die „türkischen Powergirls“: Lebenswelt und kommunikativer Stil einer Migrantinnengruppe in Mannheim*. 2., durchges. Aufl. (Studien zur deutschen Sprache 39). Tübingen: Narr.
- Keim, Inken & Rosemarie Tracy. 2007. Mehrsprachigkeit und Migration. In Siegfried Frech & Karl-Heinz Meier-Braun (eds.), *Die offene Gesellschaft: Zuwanderung und Integration* (Basisthemen politik), 121–144. Schwalbach/Ts: Wochenschau Verlag.
- Krifka, Manfred, Joanna Błaszcak, Annette Leßmöllmann, André Meinunger, Barbara Stiebels, Rosemarie Tracy & Hubert Truckenbrodt (eds.). 2014. *Das mehrsprachige Klassenzimmer: Über die Muttersprachen unserer Schüler* (Springer Spektrum). Berlin Heidelberg: Springer VS. DOI: 10.1007/978-3-34315-5.

- Lanza, Elizabeth. 2004. *Language mixing in infant bilingualism: A sociolinguistic perspective*. Corrected ed. (Oxford Linguistics). Oxford: Oxford University Press.
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation* (ACL-MIT Press Series in Natural Language Processing). Cambridge, MA/London: MIT Press. DOI: 10.7551/mitpress/6393.001.0001.
- Montrul, Silvina. 2018. Heritage language development: Connecting the dots. *International Journal of Bilingualism* 22(5). 530–546. DOI: 10 . 1177 / 1367006916654368.
- Müller, Anja, Petra Schulz & Rosemarie Tracy. 2018. Spracherwerb. In Cora Titz, Sabrina Geyer, Anna Ropeter, Hanna Wagner, Susanne Maria Weber & Marcus Hasselhorn (eds.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln*, 1. Auflage (Bildung durch Sprache und Schrift Band 1), 53–68. Stuttgart: Verlag W. Kohlhammer.
- Nero, Shondel J. & Dohra Ahmad. 2014. *Vernaculars in the classroom: Paradoxes, pedagogy, possibilities* (Routledge Research in Education). New York: Routledge.
- Niedrig, Heike. 2001. Exkurs: Die staatliche Europa-Schule Berlin (SESB). In Ingrid Gogolin, Ursula Neumann & Lutz R. Reuter (eds.), *Schulbildung für Kinder aus Minderheiten in Deutschland 1989-1999* (Interkulturelle Bildungsforschung Band 8), 77–80. Münster: Waxmann.
- Pavlenko, Aneta & Adrian Blackledge (eds.). 2004. *Negotiation of identities in multilingual contexts*. Multilingual Matters. DOI: 10.21832/9781853596483.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10 . 1017 / 9781107252349.
- Polinsky, Maria & Olga Kagan. 2007. Heritage languages: In the “wild” and in the classroom. *Language and Linguistics Compass* 1(5). 368–395. DOI: 10.1111/j.1749-818X.2007.00022.x.
- Prasad, Gail. 2017. *Parents as multilingual experts: Leveraging families’ cultural and linguistic assets in the classroom*. EdCan Network. <https://www.edcan.ca/articles/parents-multilingual-experts/>.
- Purkarthofer, Judith. 2016. *Sprachort Schule*. Klagenfurt: Drava.
- Purkarthofer, Judith. 2019. Austria’s curriculum for heritage language education across languages: A case study in balancing speakers’ needs on the local, national and international level. In C. E. Seals & V. I. Olsen-Reader (eds.), *Embracing multilingualism across educational contexts*, 152–181. Victoria University Press.

- Purkarthofer, Judith & Mi-Cha Flubacher (eds.). 2022. *Speaking subjects in multilingualism research. Speaker-centred and biographical approaches*. Bristol: Multilingual Matters.
- Purkarthofer, Judith & Guri Bordal Steien. 2019. "Prétendre comme si on connaît pas une autre langue que le swahili": Multilingual parents in Norway on change and continuity in their family language policies. *International Journal of the Sociology of Language* 2019(255). 109–131. DOI: 10.1515/ijsl-2018-2005.
- Robinson, Lucy Danielle, Jade Lauren Cawthray, Sarah Elizabeth West, Aletta Bonn & Janice Ansine. 2018. Ten principles of citizen science. In Susanne Hecker, Muki Haklay, Anne Bowser, Zen Makuch, Johannes Vogel & Aletta Bonn (eds.), *Citizen science: Innovation in open science, society and policy*, 27–41. London: UCL Press.
- Romaine, Suzanne. 1989. *Bilingualism* (Language in Society 13). Oxford: Blackwell.
- Rymes, Betsy. 2020. *How we talk about language: Exploring citizen sociolinguistics*. Cambridge: Cambridge University Press.
- Rymes, Betsy & Andrea R. Leone. 2014. Citizen sociolinguistics: A new media methodology for understanding language and social life. *Working Papers in Educational Linguistics* 29(2). 25–43.
- Saunders, Georg. 1988. Good advice? *The Bilingual Family Newsletter* 5(3). [https://web.archive.org/web/20230329172542/https://www.multilingualmatters.com/resources/pdfs/bilingual\\_family/BFN%2005-3.pdf](https://web.archive.org/web/20230329172542/https://www.multilingualmatters.com/resources/pdfs/bilingual_family/BFN%2005-3.pdf).
- Schroeder, Christoph. 2017. Zum Sprachverständnis in der Migrationsdebatte in Deutschland. In Yüksel Ekinci, Elke Montanari, Lirim Selmani & Ludger Hoffmann (eds.), *Grammatik und Variation: Festschrift für Ludger Hoffmann zum 65. Geburtstag*, 361–367. Heidelberg: Synchron, Wissenschaftsverlag der Autoren.
- Schulz, Petra. 2013. Sprachdiagnostik bei mehrsprachigen Kindern. *Sprache · Stimme · Gehör* 37(4). 191–195. DOI: 10.1055/s-0033-1358700.
- Schwartz, Mila & Anna Verschik (eds.). 2013. *Successful family language policy: Parents, children and educators in interaction*. Dordrecht: Springer. DOI: 10.1007/978-94-007-7753-8.
- Seals, Corinne A. 2018. Positive and negative identity practices in heritage language education. *International Journal of Multilingualism* 15(4). 329–348. DOI: 10.1080/14790718.2017.1306065.
- Seals, Corinne A. & Vincent I. Olsen-Reader (eds.). 2019. *Embracing multilingualism across educational contexts*. Victoria University Press.
- Shirk, Jennifer L. & Rick Bonney. 2018. Scientific impacts and innovations of citizen science. In Susanne Hecker, Muki Haklay, Anne Bowser, Zen Makuch,

- Johannes Vogel & Aletta Bonn (eds.), *Citizen science: Innovation in open science, society and policy*, 41–51. London: UCL Press.
- Svendsen, Bente Ailin. 2018. The dynamics of citizen sociolinguistics. *Journal of Sociolinguistics* 22(2). 137–160. DOI: 10.1111/josl.12276.
- Thoma, Dieter & Daniela Ofner. 2021. Bedingungsfaktoren für die Entwicklung von Basiskompetenzen für sprachsensibles Unterrichten bei gymnasialen Lehramststudierenden. In Karina Karst, Dieter Thoma, Julia Derkau, Jürgen Seifried & Stefan Münzer (eds.), *Lehrer\*innenbildung im Kontext leistungsbezogener Heterogenität und Mehrsprachigkeit von Schüler\*innen*, 139–159. Münster: Waxmann.
- Tracy, Rosemarie. 2008. *Wie Kinder Sprachen lernen*. Tübingen: Narr.
- Tracy, Rosemarie. 2014. Mehrsprachigkeit: Vom Störfall zum Glückssfall. In Manfred Krifka, Joanna Błaszcak, Annette Leßmöllmann, André Meingueter, Barbara Stiebels, Rosemarie Tracy & Hubert Truckenbrodt (eds.), *Das mehrsprachige Klassenzimmer: Über die Muttersprachen unserer Schüler* (Springer Spektrum), 13–33. Berlin Heidelberg: Springer VS. DOI: 10.1007/978-3-34315-5.
- Tracy, Rosemarie & Ira Gawlitzek. 2023. *Mehrsprachigkeit und Spracherwerb*. Tübingen: Narr.
- Tracy, Rosemarie, Vytautas Lemke, Maren Krempin & Kerstin Mehler (eds.). 2009. *Offensive Bildung: Sprache macht stark*. 1. Aufl (Frühe Kindheit Sprache & Literacy). Berlin: Cornelsen-Scriptor.
- Tracy, Rosemarie, Petra Schulz & B. Voet Cornelli. 2018. Sprachstandsfeststellung im Elementarbereich. In Cora Titz, Sabrina Geyer, Anna Ropeter, Hanna Wagner, Susanne Maria Weber & Marcus Hasselhorn (eds.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln*, 101–116. Stuttgart: Kohlhammer.
- Tsehaye, Wintai, Rosemarie Tracy & Johanna Tausch. 2025. Inter- and intra-individual variation: How it materializes in Heritage German and why it matters. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 141–177. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775165.
- Tsimpli, Ianthi-Maria. 2014. Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313. DOI: 10.1075/lab.4.3.01tsi.
- Voet Cornelli, Barbara, Petra Schulz & Rosemarie Tracy. 2013. Sprachentwicklungsdiagnostik bei Mehrsprachigkeit: Eine Herausforderung für die pädagogische Praxis. *Monatsschrift Kinderheilkunde* 161(10). 911–917. DOI: 10.1007/s00112-012-2752-z.

- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups. In Shanelley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 33–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.1577515.
- Wiese, Heike, Anke Sennema & Rosemarie Tracy. 2020. *Deutschpflicht auf dem Schulhof? Warum wir Mehrsprachigkeit brauchen*. Berlin: Dudenverlag.
- Wörfel, Till, Almut Küppers & Christoph Schroeder. 2020. Herkunftssprachlicher Unterricht. In Ingrid Gogolin, Antje Hansen, Peter Leseman, Sarah McMonagle & Dominique Rauch (eds.), *Handbuch Mehrsprachigkeit und Bildung*. Springer VS.
- Yıldız, Cemal, Nathalie Topaj, Reyhan Thomas & Insa Gülsow (eds.). 2017. *Die Zukunft der Mehrsprachigkeit im deutschen Bildungssystem: Russisch und Türkisch im Fokus*. Frankfurt am Main: Peter Lang Edition.

## **Part II**

# **Dynamics of grammatical structure**



# Chapter 5

## Inter- and intra-individual variation: How it materializes in Heritage German and why it matters

Wintai Tsehayé<sup>a</sup>, Rosemarie Tracy<sup>a</sup> & Johanna Tausch<sup>a,b</sup>

<sup>a</sup>University of Mannheim <sup>b</sup>Leibniz Institute for the German Language

This contribution focuses on the clausal architecture of Heritage German in the United States. In particular, we identify structural properties of German clauses which prove to be robustly canonical even though there is inter- and intra-individual variation on other levels of the linguistic system. With respect to specific clausal features, group comparisons of heritage speakers in both their languages and with monolingually-raised speakers of German have been conducted. We also contribute to current discussions of the heterogeneity of heritage speaker competence and performance and the relevance of individual speaker profiles by qualitative analyses based on data from what we consider *Tiny Language Islands*, i.e., acquisition scenarios with the heritage language only spoken within the immediate family. These findings matter because they speak to hypotheses concerning interfaces between subsystems of our overall linguistic knowledge and, moreover, to discussions on language learnability under conditions of reduced first language exposure.

### 1 Introduction

The question of what happens to grammars of German under intensive language contact has been addressed many times. Since German-speaking communities form diasporic enclaves in many parts of the world, there exists a rich natural laboratory for investigating the effects of internal and external variables impacting on language maintenance, attrition, cross-linguistic interaction, and change.

Wintai Tsehayé, Rosemarie Tracy & Johanna Tausch. 2025. Inter- and intra-individual variation: How it materializes in Heritage German and why it matters. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 141–177. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775165



In some cases, as in Namibia, German remains a vital minority language, enjoys considerable social prestige, and is supported within the educational system (cf. Shah & Zappen-Thomson 2017, Wiese et al. 2017, 2022, Zimmer 2019). In other parts of the world, and given other status constellations, perspectives for the future of German look rather grim (with respect to German varieties in the US, see Boas (2009a), Boas (2009b) for the decline of Texas German; Hopp & Putnam (2015) for a *moribund* variety, Moundridge Schweitzer German in Kansas). At the same time, upward trends have been noted for Pennsylvania German (PG) communities, with PG gaining prestige among young generations who did not acquire it as their first language (L1) but who positively identify with their German legacy and feel motivated to learn the ancestral language as adolescents or adults (cf. Stolberg 2014).

In addition to the wealth of information already available on heritage German in the context of majority English, we aim at providing further insight by investigating both spoken and written productions of additional heritage speaker populations in systematically varied elicitation contexts.

Our contribution targets a very specific subset of heritage German speakers: the adolescent offspring of first generation German immigrants. They were either born and raised in the US, with German as their L1 and English as a simultaneous first language (2L1), or, in the case of immigration by age two, as an early second language (L2). We therefore contribute to filling a data and description gap between research on first generation German immigrants in English-speaking countries on the one hand (e.g., Tracy & Lattey 2001, Schmid 2011, Tracy 2022, Tracy & Stolberg 2008, Keller 2014), and research on minority German language islands around the world, on the other (e.g., Andersen 2016, Boas 2009a, Boas 2009b, Boas 2010, Clyne 2003, Eichinger et al. 2008, Földes 2016, Fuller 2001, Hopp & Putnam 2015, Huffines 1980, Johannessen & Salmons 2015, Louden 2008, Plewnia & Riehl 2018, Putnam & Salmons 2013, Rosenberg 2005, Stolberg 2014, 2015, among many others).

As part of the Research Unit *Emerging Grammars in Language Contact Situations* (RUEG) and its shared methodology, all heritage speakers (HSs) took part in the same elicitation tasks (see Wiese, Labrenz, et al. 2025 [this volume]), both in speech and in writing, in formal and informal situations, and in both their heritage language (HL) and majority language (ML). This allowed us to tap into their linguistic repertoires, including register-specific resources and preferences for specific structural options according to formality and mode of communicative situations. Repeated elicitations of factually similar reports helped distinguish performance errors from systematically recurring, i.e., more stable phenomena. In various instances, comparing what speakers produced in their minority and

in their majority language helped clarify how they interpreted a specific scene and which formulations in German likely resulted from language contact with English. Moreover, additional data in form of informal chitchats, written sentence corrections, and oral and written sentence completions provided insight into participants' performance in the face of different task demands.

On the basis of biographical information available, we could take language exposure and maintenance conditions into consideration. As opposed to historical language islands, where the minority language may still be used outside the home, a good number of our participants grew up in what we consider a *Tiny Language Island* scenario, with the HL only spoken within the realm of the immediate family, in some cases with only one family member. Potential consequences for HL productions will be addressed in our discussion of HS heterogeneity (Section 6). Since our HS group includes nine sets of siblings, we could take advantage of this more optimally controlled condition due to individuals sharing various linguistic and extra-linguistic background variables.

The main focus of this contribution lies on clausal syntax, specifically on the distribution of finite and non-finite verbs and verb components, on the morphology-syntax interface and on phonological exponents of morphological subsystems. In Section 2, we illustrate critical properties of German clause structure and point out relevant contrasts between German and English. Section 3 turns to findings for historical German language island communities in majority English contexts and formulates research questions. Section 4 provides background on our research methods, consultants, and the corpus, followed by results in Section 5. Section 6, with its focus on Tiny Language Islands, illustrates what individual learner profiles can contribute to our understanding of HS heterogeneity. Our conclusion in Section 7 summarizes findings and relates the spectrum of variation identified in our data to language acquisition in childhood and general theoretical issues.

## 2 Properties of German clause structure

For the clausal architecture at issue in this chapter we draw on a linear model based on a *topological fields* metaphor (Drach 1963, Höhle 1986, Wöllstein 2014). Figure 1 captures the distributional properties at issue and introduces the terminology employed henceforth, including the left-to-right orientation based on written clauses rather than talking about what is perceived first or last.

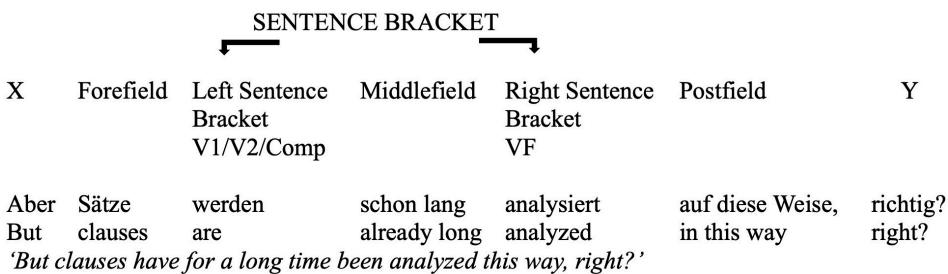


Figure 1: Topological schema of German clauses

In canonical<sup>1</sup> main clauses, the finite verb appears in the left sentence bracket (LSB), resulting in verb second (V2) or verb first (V1) structures, while non-finite parts of the verb phrase (VP), such as infinitival verbs, participles, separable verbal particles, are verb final (VF), i.e., they occur in the right sentence bracket (RSB), as in Table 1a. X and Y in Figure 1 indicate the place of pre- and post-clausal adjuncts. In subordinate clauses, the LSB is occupied by complementizers and relative pronouns.<sup>2</sup> In this case, finite verbs have to join non-finite verbs in the RSB as in Table 1b. In main clauses the forefield is taken up by preposed, topicalized constituents, including subordinate clauses, as in Table 1c. In this case, subjects remain in the middlefield. In yes/no questions (*Bist du hungrig?* ‘Are you hungry?’) and irrealis constructions (*Hättest du das nur früher gesagt!* ‘If you had only said this sooner!’), the preverbal field remains empty, hence the finite verb occurs sentence-initially (V1).

The postfield holds constituents (clauses and adjuncts) extraposed from the middlefield, as in Table 1b–d, or freely adjoined constituents. As a consequence, constituents belonging syntactically to a preceding clause and clarifying afterthoughts are not always easily distinguishable as such, especially when prosodic cues are missing (see various papers in Vinckel-Roisin 2015). The postfield provides the canonical area for subordinate clauses and otherwise *heavy* constituents. Shifting them out of the middlefield reduces the distance between discontinuous parts of verbs, thereby easing memory load (Haider 2010, Imo 2016, Proske 2015). The example in Table 1d illustrates a *heavy* middlefield since the adverbial clause has not been extraposed. At the same time, postponing “lightweights” (focus particles and adverbials, i.e., *auch* ‘also’, *immer* ‘always’,

<sup>1</sup>By canonical we here mean adherence to codified standard language norms (cf. Wiese, Allen, et al. 2025 [this volume]).

<sup>2</sup>We here forgo the discussion of whether relative constituents are placed in the forefield in subordinate clauses (cf. Imo 2016: 214, Wöllstein 2014: 27ff).

*jetzt* ‘now’, e.g., *Sie fährt nach Hause jetzt.* ‘She drives home now.’), is typically considered stylistically infelicitous or highly marked, at best, in written contexts. In informal conversational contexts, however, the right clausal edge is used extensively.

Table 1: German clause structure illustrated

|     | Forefield  | LSB            | Middlefield   | RSB                                      | Postfield  |
|-----|--|----------------|---|--|--|
| (a) | Ich<br>I<br>‘I was walking along the street.’  | lief<br>walked | die Straße<br>the street  | entlang.<br>along                        |  |
| (b) | Ich<br>I<br>‘I was walking along the street when the accident happened.’   | lief<br>walked | die Straße<br>the street  | entlang<br>along                         | als der Unfall<br>passierte.<br>when the<br>accident<br>happened |
| (c) | Als der Unfall<br>passierte<br>When the<br>accident<br>happened<br>‘When the accident happened I was walking down the street.’ | lief<br>walked | ich die Straße<br>I the street                                  | entlang.<br>along.                       |  |
| (d) | Ich<br>I<br>‘I was walking down the street when the accident happened.’  | lief<br>walked | als der Unfall<br>passierte<br>when the<br>accident<br>happened | die Straße entlang.<br>the street along. |  |

Over the last two decades, numerous publications have shown the need for more fine-grained analyses of both clausal peripheries (for critical discussions, cf. Antomo & Steinbach 2010, Freywald 2016, Freywald et al. 2023, Molnár & Winkler 2010, Speyer 2009, Vinckel-Roisin 2015, Winkler 2017, see also Bunk et al. 2025 [this volume]). Since we find similar ordering at the clausal peripheries in our HL data, taking into account what happens in non-heritage varieties of German is crucial. Including monolingually-raised speakers (MSs) in our investigations enables us to identify trends in homeland varieties independently of language contact and guards us against hastily attributing one or the other HL

pattern to transfer from English or to some other “flaw” in HL competence or performance.

At the same time, language contact phenomena<sup>3</sup> are fully expected. After all, German and English share typological features and a considerable number of cognates, the latter being congenial to the emergence of converging forms. Yet, there are also obvious differences between the two languages. Hence, both the extent of structural ‘overlap’ due to parallel clausal patterns in main clauses with simplex verbs (*Bist du hungrig?* ‘Are you hungry?’, *Er schreibt alle Briefe mit der Hand*. ‘He writes all letters by hand.’) and the specific contrasts selectively mentioned below make the two an intriguing but also quite challenging language pair to consider in language-contact scenarios.

The German VP is head-last, while English is head-first across all phrasal constituents. Whereas German allows a full range of V2 effects in main clauses, with almost any constituent able to access the forefield, English is an SVO language, restricting the occurrence of subjects to positions preceding the finite verb. Topicalized constituents then appear in front of subjects (*Ice-cream I only eat in summer. Only in summer I eat ice-cream*). Nevertheless, V2 effects productive in earlier stages of English survive in subject-auxiliary inversion in questions (*Could/Did the man find his key? What have you told him?*) and are tied to special triggers, such as negation (*Under no circumstances/Never again will I eat ice-cream, summer or winter.*). English also residually allows subject-main verb inversion with intransitive verbs and initial locative or directional adverbials (*All of a sudden, across the street jumped a dog.*).

Topicalization and inversion aside, English word order is canonically SVO in both main and subordinate clauses. German, on the other hand, as pointed out above, requires V2 in finite main clauses but asymmetrically places finite verbs in final position (VF) in subordinate clauses. Neither English nor German license pro-drop, i.e., empty subjects in finite clauses, most easily recognized in finite subordinate clauses (\**Ich würde behaupten, \_\_ kommt heute nicht.* ‘I’d claim \_\_ - won’t come today’). In specific contexts, English allows truncating main-clause preverbal subjects (*diary drop*), as in \_\_ *Depends on the weather* or \_\_ *Arrived late, \_\_ missed my train*. In German, on the other hand, all sorts of topicalized constituents, not just subjects, can be dropped from the forefield, given appropriate discourse licensing contexts, such as topics already under discussion. The result is a V1 pattern as in (1), with  $\emptyset$  indicating the position of an inferrable constituent, in this case a topicalized direct object.

---

<sup>3</sup>For comprehensive studies on German-English language contact, see among others (Clyne 2003, Johanson 2002, 2008, Tracy & Lattey 2001, 2010, Stolberg 2014, Keller 2014).

- (1) Willst du mein Buch lesen?  
want you my book read  
‘Do you want to read my book?’  
∅ Kenn ich schon.  
∅ know I already  
‘I already know it.’

In addition to distributional differences, morphological contrasts abound. German possesses rich inflectional paradigms spelling out person, number, and tense in the verb. English, however, – the suppletive copula paradigm apart – only marks person and number in main verbs in present tense third-person contexts. Case surfaces in articles, pronouns, and adjectives within the determiner phrase in German, confounded by syncretism involving grammatical or natural gender and number, and complicated by weak/strong inflections depending on choice of determiner (definite vs. indefinite). As a general rule, only masculine and neuter are marked morphologically for case in the genitive singular and the dative plural. Dative marking in singular masculine nouns, can be considered outdated and appears in highly literary genres (“Dem Manne kann geholfen werden.”, Schiller, Die Räuber, act 5, scence 2). In addition to configurationally determined case marking, some German verbs call for idiosyncratic cases, as the genitive in *jemanden eines Verbrechens beschuldigen*, ‘to accuse someone of a crime’, where case is also marked on the noun.

In comparison, Modern English inflectional morphology is greatly reduced across word classes. Natural gender as well as nominative vs. non-nominative case distinctions are only visible in singular pronouns (e.g., he/him/his; she/her), with their plural unmarked for gender (they/them/their). Articles as well as adjectives do not require agreement inflections, while nouns are only marked for plural and the genitive.

On the basis of the selected typological differences, the question arises, then, which features of German clausal architecture on the one hand and of morphological subsystems on the other can be identified in HS productions and whether findings pointing towards noncanonicity are due to influence from English or to developments independent of language contact (e.g., language-internal dynamics, register sensitivity, task type). We hypothesize that our second generation immigrants will be able to discover most of these properties, even under reduced input conditions, different from what we see in established language islands to which we briefly turn next.

### 3 German from a language contact and *Heritage Language Island* perspective

Given the cross-linguistic similarities and differences mentioned, quite a number of contact phenomena can be expected in bilingual German-English speech and writing (Tracy & Lattey 2001, Schmid 2011, Tracy 2022, for first generation immigrants; Vainikka & Young-Scholten 2011 for L2 acquisition of German by L1 English speakers). With respect to word order, one could, for instance, reasonably expect an increase in postposed constituents due to the stability of English [V<sub>+/−</sub> finite XP (YP)] across clauses. Intensive contact with English might even facilitate the highly noncanonical extraposition of direct objects into the postfield, as demonstrated by Clyne (2003) for heritage German speakers in Australia, see (2).

- (2) (Clyne 2003: 137)

Mummy hat gesagt die Wörter für mich.

Mummy has said the words for me.

‘Mummy told me what to say.’

At the same time, Clyne’s data also underscore the robustness of V2 effects in main clauses and of VF in subordinate clauses, even though there may be complementizer-specific variation. Similarly, both Louden (2008), for Pennsylvania German, and Boas (2009a), Boas (2009b), for German-speaking communities in Texas, found some complementizers in canonical VF clauses, others exclusively with V2 (hence noncanonical), while a third group of complementizers appeared in both positions. Likewise, Hopp & Putnam (2015: 29) report that their consultants consistently produced *dass*+V2-clauses in elicited production tasks (e.g., *dass sie hot ei(n) Car gekauft gestern*<sup>4</sup>, ‘that she has bought a car yesterday’). Speakers also rated these patterns more acceptable than *dass*+VF.<sup>5</sup> On the other hand, the very same persons preferred VF in clauses introduced by *wenn* ‘when’ and *weil* ‘because’, both in their spontaneous productions and when they were asked for well-formedness judgments. Importantly, several researchers stressed

---

<sup>4</sup>This example also illustrates the extraposition of a light-weight constituent, the adverb ‘yesterday’. While a common pattern in spoken German, in English-German bilinguals an increase of such structures enhanced by surface parallelism can be expected.

<sup>5</sup>Stolberg (2015), who also found noncanonical complement clauses with *dass* in her written PG data, points out that this clausal pattern could also be due to verb-projection raising, which is a canonical option in several modern and old Germanic languages and dialects, including standard patterns with infinitives (*dass/weil er hat das nicht wissen können*, ‘that/because he couldn’t know that’).

that some noncanonical patterns were specific to individual speakers (cf. Boas 2009a, Boas 2009b, Clyne 2003, Hopp & Putnam 2015). These findings underscore that it is crucial not to lose sight of speaker idiosyncrasy.

In German-speaking diaspora communities, specific consequences of contact with English surface in the choice of near-homophonous<sup>6</sup> complementizers which partially overlap in meaning, such as German *wenn* (conditional), corresponding to ‘if’, but also to ‘always when’, and English ‘when’ (temporal), corresponding to German *als*, see (3) from one of the elderly participants in Hopp & Putnam (2015). At the same time, clausal syntax is canonical, with the finite verb following the participle in the RSB.

- (3) (German in Kansas, Hopp & Putnam 2015: 196)

wenn mir erscht geheirat henn  
when we first married have  
'when we first got married'

Evidence for the very same crosslinguistic effect involving *when/wenn* can be found in first generation immigrants, both in conversations (4) and in writing (5), (Tracy 2022: 421, spelling of the handwritten original preserved).

- (4) Die mussten immer im im Feld helfen im Sommer  
They had-to always in-the in-the field help in-the summer  
**wenn** sie auch noch in der Schule war'n ne.  
when they also still in the school were PARTICLE.

'They always had to help in the fields in summer when they were still going to school, didn't they?'

- (5) ein Glitzern eine Pracht **when** man von innen rausschauen kann  
a glittering a splendor when one from inside outlook can  
'a glittering, such a splendid view if you can look out at it from inside (the house)'

While choice of *when* is semantically appropriate in (5), there is interference on the orthographic level. Both the evidence from German islands with a long history of language maintenance and from first generation immigrants show how partial congruence of features from various levels (semantics, syntax, phonology, orthography) create “grey zones” (Clyne 1987: 755), favoring convergence. It comes as no surprise, then, that the heritage data under discussion here from

---

<sup>6</sup>See also Johanson (2002: 292) for the selective copying of “relatively homophonous false friends”.

second generation immigrants (RUEG corpus<sup>7</sup>, Lüdeling et al. 2024) points in the same direction (6).

- (6) so *wenn* sie hat gehaltet sie hat die das erste Auto geschlagt  
so when she has stopped she has the the first car hit  
'so when she had stopped she hit the first car' (Language Situations  
(LangSit) narration formal-spoken)

Besides the occurrence of *wenn* instead of *als*, two more features in (6) are noteworthy, both involving the position of the auxiliary *hat* 'has'. Canonically, the linearly first *hat* would occur in the right sentence bracket (VF). Since the follow-up clause is the actual matrix clause, the *wenn*-clause is expected in its forefield, with the subject – the linearly second token of *sie* 'she' – in the midfield. A partially parsed canonical version is shown in (7), where we maintain the calque of 'hit', *schlagen* instead of the expected *getroffen* 'driven into', and ignore gender repair in the article (*die das*).

- (7) [s [s als sie gehalten **hat**] **hat** sie das erste Auto geschlagen]

In our discussion of heterogeneity (Section 6) we will return to the question of how prominent clausal patterns like (6) are within individual HSs.

Against this backcloth of research on German in close contact with US English, our contribution pursues the following questions to see whether both HSs and MSs, if faced with the same communicative challenges and task demands, perform differently in terms of register availability and structural choice.

- RQ1: Are there differences in the clausal syntax in MSs and HSs, and, if so, could this be due to influence from English?
- RQ2: In contrast to early acquired core syntactic phenomena, how does (non)canonical variation manifest itself in the domain of morpho-syntax?
- RQ3: What can we conclude about heterogeneity and individual variation by taking a closer look at Tiny Language Islands where HL input is limited to the family (parents and siblings), and maybe to one parent only?

---

<sup>7</sup>In the RUEG corpus, transcriptions of the spoken data are annotated for pauses “(–)”, prolongations “:”, hesitations *äh/ähm* and other production-related phenomena, such as tongue-clicks. These were not included in our examples since we do not discuss them here. See Zürn et al. (2025 [this volume]) for spoken discourse phenomena, also Böttcher & Zellers 2023, Tracy & Gibbon 2023).

While the first question will be pursued on the basis of published studies (Pashkova et al. 2022, Tsehay 2023, Tsehay et al. 2021), questions two and three will be tackled on the basis of novel data in order to clarify what heterogeneity means along a scale ranging from the predominantly canonical to increases in variation to the emergence of novel grammars.

## 4 Method

Our adolescent HL participants ( $N = 29$ , age 13–18)<sup>8</sup>, all second generation speakers of German, were recruited in the US (Boston, Madison, St. Paul, Minnesota). They grew up with at least one German-speaking parent in the household. Only a few had temporary access to classes in German, attended Saturday Schools, or were engaged in leisure-time activities involving other HSs of German. Crucially, though, they did not grow up in a German-speaking community. Participant metadata was available through a questionnaire on language background, language use, and personality that our consultants filled in and in which they rated their own proficiency<sup>9</sup> with respect to speaking, listening comprehension, reading, and writing. Parallel age groups of MSs of German and MSs of English were recruited in Germany and in the US, respectively.

In addition to the RUEG-wide elicited narratives concerning a fictive accident (cf. Wiese, Labrenz, et al. 2025 [this volume]), we ran an on-site oral sentence completion task as well as a follow-up computer-based online sentence completion and sentence correction task. Both online tasks were conducted with a subset of speakers only. On-site meetings with participants also included a 15-minute informal, relaxed conversation, referred to here as *chitchat*.<sup>10</sup>

Quite apart from narrowly linguistic, system-internal properties, the tasks designed differed along a scale of complexity. While there is an ongoing discussion on what counts as complex (Housen et al. 2012), what matters in our case is *task*

---

<sup>8</sup>These are the numbers and ages for all the heritage German participants. Throughout this chapter, we will provide the numbers for speakers included in the various analyses undertaken.

<sup>9</sup>The two visits to our temporary “labs” on part of our participants were time-consuming and required willingness to engage in different tasks, some of which, especially where they required writing, were not without stress. While many HSs enjoyed taking part (as evident from our informal conversations), they also felt, that they were being tested, causing some of them to do their best and produce highly elaborate clauses. The link for the adolescent participant questionnaires can be accessed via: <https://osf.io/qhupg/>

<sup>10</sup>The data produced in the LangSit narrations are openly accessible via: <https://zenodo.org/record/5808870>. The data of the sentence completion and sentence correction tasks can be accessed via: <https://osf.io/28j57/>. The data of the chitchat is not openly accessible. We included it in our analyses as an additional cue to the options available to participants.

*complexification* via constraints placed on choices left to the participant. Positively framed: We provided opportunities for speakers to display their linguistic resources, including choices associated with different formality settings (formal, informal) and modes (speech, writing). In our sentence completion task, we offered cues nudging our participants in the direction of specific clausal patterns. The following list provides a summary of tasks according to what we consider an increase in complexification.

- A. Chitchat: The informal chitchat consisted of a 15-minute warm-up and familiarization conversation, relatively unconstrained in terms of topics (interests, favorite foods, neighborhood, etc.) and completely unconstrained with respect to syntactic choices. The only constraints were those imposed by general discourse-pragmatic principles governing conversational alignment (Pickering & Garrod 2004), first-encounter topics, politeness, and relevance.
- B. LangSit: Elicited narrations were based on a video shown – without intervening turns by elicitors – in different communicative situations (see Wiese, Allen, et al. 2025 [this volume]). Syntactic preferences were up to the speaker, while narration contents were restricted by the events observed. Specific contents were primed by task demands specified in the instructions of who to report to (police vs. friend). More detail in terms of cause and effect and protagonist identification was called for in spoken and written police reports.<sup>11</sup> Participation in the narration task also differed from the chitchat encounters in that it required willingness to play-act.
- C. Written sentence completion based on the video: Sentence completions confronted the participants with structures provided in the stimuli that required specific continuation patterns and left them no choice, at least not with respect to canonical verb placement. During the written online completion task, there was no time pressure since participants had been sent a link to the stimuli and could take as much time as they wished.
- D. Oral sentence completion of sentences read out to them: In this task, an immediate response in form of clausal continuation was called for, so speakers had very little time for planning of how to pick up the baton.

---

<sup>11</sup>If speakers had reported the same detail (for instance about the way people were dressed, etc.) in informal productions, it would have been at odds with conversational maxims.

- E. Written sentence correction: Sentence corrections gathered in an additional online study requested reactions to stimuli including noncanonical verb placement, auxiliary selection, as well as tense, case, and gender inflection. Again, no time constraint was imposed.

These five task types differ in terms of relative complexity and the high vs. low production constraints placed on participants' choices. The informal chitchat was not analyzed quantitatively, but, as we show below, qualitative properties provide insight into trends identified in the standardized elicitation of the narratives.<sup>12</sup>

On the whole, speaker performance across different tasks with varying demands allowed us to arrive at a comprehensive picture of the spectrum of intra-individual variation as well as of both stable and newly emerging options.

## 5 Results

In the following section, results from quantitative and qualitative studies relating to the issues sketched above are reported. After laying out quantitative findings regarding the clausal architecture across HSs and MSs, we briefly turn to a very different picture emerging from investigations of morphological phenomena, especially with respect to exponents of case and gender. Contrary to what we see in the syntactic domain, there is considerable inter- and intra-individual variation in the realization of various morphological paradigms and with respect to auxiliary choice.

### 5.1 Clause-type optionality

In a study on clause type optionality, operationalized as choice between different clause structures in narrations of the same event, the production of three clause types (i–iii) was investigated in 20 adolescent HSs, 20 adolescent MSs of German, and 20 adolescent MSs of English (Pashkova et al. 2022). Thus, this study inquired into the syntactic (verb placement in main and subordinate clauses) and pragmatic (discourse structure according to register) choices available to MSs and HSs of German in both their HL, German, and their ML, English. In the following, we only focus on the German productions.

---

<sup>12</sup>As our main goal was to elicit either German or English data, interlocutors remained in a monolingual mode and did not engage in codeswitching. With respect to German productions, this might not have corresponded to family language practice and was certainly costly in terms of the monitoring required in order to stick to German. For spontaneous borrowing, calquing, and convergence, see Zürn et al. (2025 [this volume]).

(i) independent main clauses (IMCs):

Ich lief die Straße entlang. Ich sah einen Unfall.

'I walked down the street. I saw an accident.'

(ii) coordinate main clauses (CMCs) with and without subject gapping:

Ich lief die Straße entlang und sah einen Unfall.

'I walked down the street and saw an accident.'

(iii) finite subordinate clauses (SCs):

Während ich die Straße entlanglief, sah ich einen Unfall.

'While I was walking down the street, I saw an accident.'

While all participants produced clauses of the types (i–iii), sometimes more or less canonically, we were concerned with preferred choices and avoidance. Our results show that HSs and MSs of German behave similarly in their production of IMCs across communicative situations, with more IMCs in writing than in speaking, but that they differ in the frequency and distributional patterns of CMCs and SCs. HSs produced more CMCs than MSs, but both participant groups were sensitive to changes in communicative situations, with more CMCs in informal situations and in spoken productions. Additionally, the differences in CMC frequency between speaker groups was greater in formal communicative situations, regardless of the production mode. HSs produced significantly fewer SCs than MSs. However, both speaker groups were sensitive to formality with more SCs in formal settings. A closer look at SC productions across communicative situations showed that MSs distinguished (in)formality in both production modes while HSs only differentiated (in)formality in writing. Even though both speaker groups differed with respect to frequency and distribution of SCs across formality and mode, our results indicate that HSs had mastered the overall clausal architecture of German main and subordinate clauses, and that differences from MSs can be attributed to factors impinging on performance. This shows that HSs are, in fact, sensitive to register but an increase in cognitive load due to typological differences between English and German might make adherence to register sensitivity harder in spoken (online) productions than in written (offline) productions.

## 5.2 Finite subordinate clause distribution

As a follow-up to the previous study, Tsehayé et al. (2021) investigated the distribution of finite subordinate clause types (complement clauses, CompCs; adver-

bial clauses, AdvCs; and relative clauses, RelCs) in 27 adolescent HSs, 32 adolescent MSs of German, and 32 adolescent MSs of English. Again, the reportings in this chapter focus only on the German productions. Quantitative analyses showed that both speaker groups behaved similarly regarding the frequencies of SC types across formality settings (more CompCs in informal situations and more RelCs and AdvCs in formal situations, Figure 2). This is interesting, as differences between speaker groups were expected especially in AdvCs and RelCs due to their relatively late acquisition (Paradis et al. 2017, Vasilyeva et al. 2008, Andreou et al. 2021). With this study, we showed that, while we find differences in overall SC productions across speaker groups, the distribution of SC types is similar.

### 5.3 Clausal peripheries

While the first two studies addressed interface phenomena across clausal boundaries, we now turn to an interface phenomenon within clauses. Once researchers conducted in-depth investigations of the left periphery of German sentences in vernacular speech, a tendency towards V3 patterns in connection with some clausal linkers or framesetters was identified (Sluckin & Bunk 2023, Wiese 2013, Wiese & Müller 2018, Bunk 2020, Walkden 2017, te Velde 2017). At the same time, there is still a recognizable middlefield framed by two sentential brackets, unlike what we see in SVO languages like English, a phenomenon we also find in our data. While V3 remains marginal in group comparisons – with a slight increase in bilinguals (cf. Wiese et al. 2022) – it appears to be a strikingly prominent option for individual speakers, possibly enhanced by cross-linguistic impact from English. Nevertheless, even in these cases, evidence for the clausal bracket, i.e., for the discontinuous placement of finite and non-finite verbs in main clauses prevails, a point we return to in Section 6.

What about the right periphery of German sentences? Once researchers discovered the pervasiveness of *light* constituents in the postfield in informal styles of MSs of German, it became clear that they were on the track of yet another neglected phenomenon located at a clausal edge and at the interface of syntax and discourse pragmatics (cf. Vinckel-Roisin 2015), independently of language contact. The question then arises, whether these trends are visible, if not enhanced, in HSs, where another language is added to the picture. After all, extensive extraposing mentioned so far in research on established German language islands could indeed be a consequence of parallelism and convergence with English (Clyne 2003, Westphal-Fitch 2011).

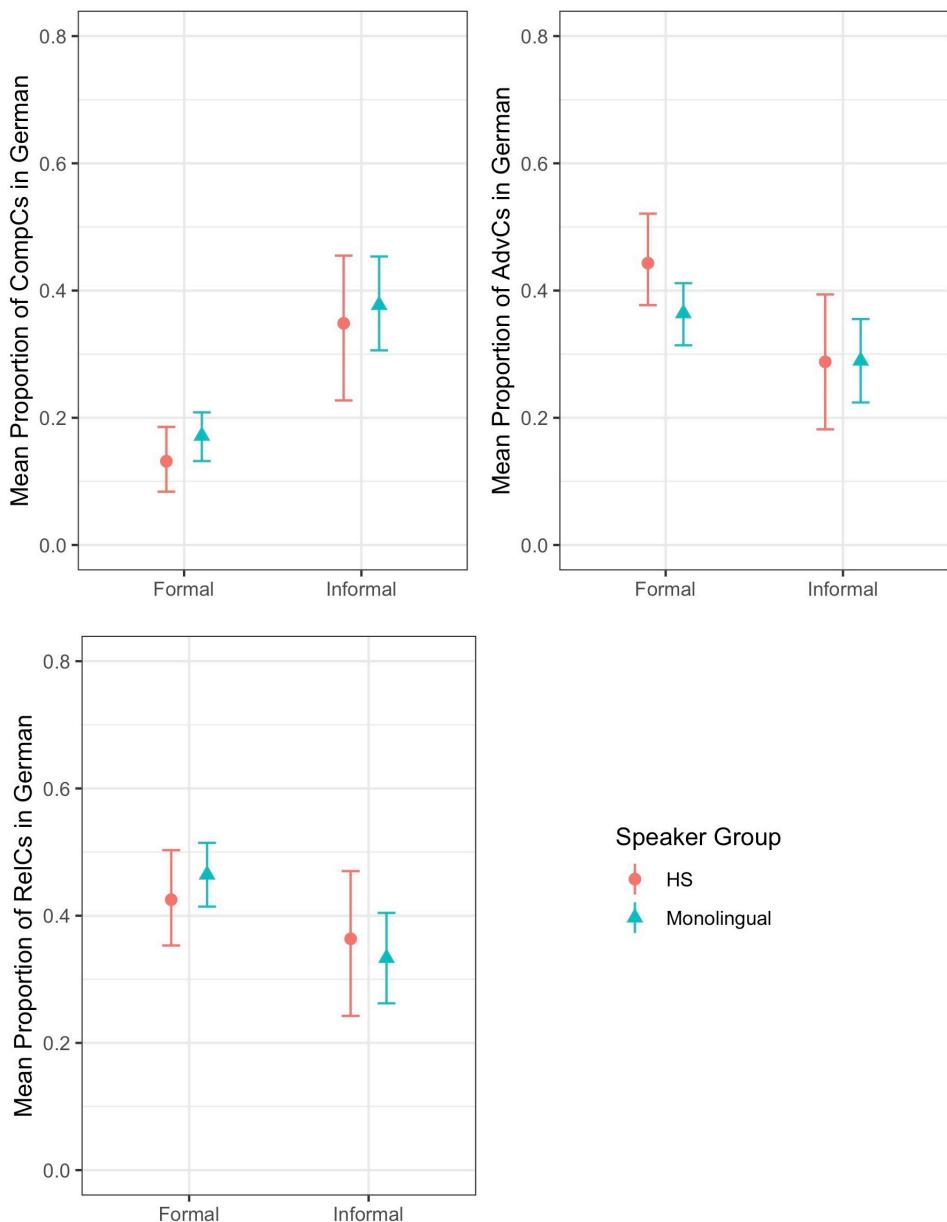


Figure 2: Mean proportions of CompCs, AdvCs, and RelCs across formality levels and speaker group (Tsehaye et al. 2021: 10)

In a study with 29 adolescent HSs and 32 adolescent MSs of German, Tsehayé (2023) focused on non-sentential light-weight constituents (LWCs), i.e., prepositional phrases, adverbial phrases, determiner phrases, etc. that appeared after the clause-final predicate. Analyses showed a similar variational spectrum (i.e., the same constituent types) and overall frequency of LWCs in the postfield in HSs and MSs but different distributional patterns. Both speaker groups behaved similarly across production modes: more LWCs were produced in the spoken mode, indicating that the extraposition of LWCs is still a predominantly spoken phenomenon (Imo 2015, Zifonun 2015, Zifonun et al. 1997). Productions across formality levels revealed that MSs furthermore distinguished between formal and informal situations; they produced more LWCs in the informal communicative situations while HSs did not draw this distinction (see Figure 3).

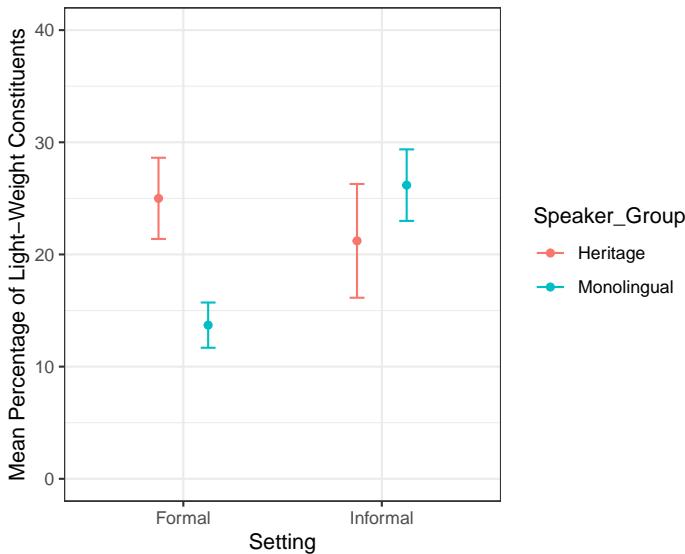


Figure 3: Mean percentage of LWCs across formality levels and speaker group (Tsehayé 2023: 9).

It appears plausible that the missing formality differentiation in HSs can be traced to differences in input conditions between speaker groups and, thus, to diverging access to and awareness of register norms. In order to zoom in on potential language contact effects and transfer phenomena underway in HSs of German, an additional analysis of the extraposition of prepositional phrases (PPs) across speaker groups was performed. While PPs might be ideal extraposition candidates, due to surface parallelism between English and German (Hopp &

Putnam 2015, Westphal-Fitch 2011), the HSs and MSs in this study did not differ with respect to the occurrence of PPs in the postfield. This could be traced back to an increased tendency of PP extrapositions, also present in MSs, which, in turn, results in similarities across speaker groups. Most importantly, with respect to the overall spectrum of constituents extraposed by HSs, we found no extrapositions of direct objects, except for very few instances entirely due to individual speakers (to be discussed in Section 6). Hence, our consultants did not differ from MSs in their overall frequency of extrapositions, regardless of language contact with English. Rather, differences in the distribution of LWCs between speaker groups can be attributed to differences in register awareness.

#### **5.4 Multiple solutions to the same challenge**

In contrast to the (relative) stability of clausal architecture, within and across clausal boundaries, there is considerably more divergence in morphological subsystems. Studies on established language islands typically report leveling of agreement, gender, and case paradigms (Boas 2009a, Boas 2009b, Clyne 2003, Yager et al. 2015, Zimmer 2020). However, Boas (2009a) also stresses that the overall extent of variability in his Texas German corpus, for instance in gender assignment, is due to just some individuals among the group of participants. Clearly, it is important not to lose sight of these individual profiles because they will, in the end, deliver detailed information on the spectrum of choices available. Affected are formally non-transparent exponents due to massive syncretism and distribution requirements across different carriers (articles, quantifiers, wh-constituents, prenominal adjectives, some overtly on the noun as well). Previous research has shown that HSs have difficulty acquiring and/or retaining a canonical and complete paradigm of inflectional morphology (Flores 2020, Montrul 2011, Polinsky 2018a,b). In contrast with basic syntax-internal regularities such as head placement within VPs, the detection and reconstruction of morphological subsystems depends on many – often idiosyncratic – properties of target systems.

Cross-linguistic differences show up in the way morphological paradigms are integrated in the course of acquisition, especially where various levels or subsystems have to be mapped onto each other. This sits well with the Interface Hypothesis (Tsimpili 2014): Core syntactic, syntax-internal grammatical properties are early in L1, 2L1 and in early childhood L2 acquisition, hence thoroughly entrenched and remarkably stable in the long run, as opposed to other subsystems involving interfaces requiring more and differentiated exposure in order to reach target-like states. The latter undergo leveling more easily. The discovery of, for instance, non-transparent case or gender marking in German and the mapping

at the syntax-morphology interface needed for case requires time. Case marking, for instance, is not fully acquired before school age. In addition, both case and gender marking presuppose the emergence of articles as functional categories and carriers of morphological exponents.

First, morphosyntactic aspects raise a number of acquisition hurdles even for MSs of German. With respect to HSs with English as the ML, we expect variation depending on transparency of the system and typological properties. Second, we expect inter-individual variation and differences depending on quantity and quality of exposure. Third, speakers may come up with intra-individual, unique and nevertheless systematic options.

In our heritage German data, individual speakers show an unexpected preference for dative forms and even overgeneralize them in accusative contexts. This differs from early case marking in L1 German (Clahsen 1984, Tracy 1986), where the accusative gets the better of the dative for several years. Overgeneralized datives from one HS participant across all registers can be seen in (8a-d), with masculine and neuter articles in the context of feminine nouns, as in (8b), where a canonical accusative form (*eine*) gets replaced by a noncanonical dative alternative (*einem*).<sup>13</sup> This could be attributed to a tendency towards transparency and salience in HSs (Polinsky 2018a: 165,b, see also Zürn et al. 2025 [this volume]). In the following oral and written set by the same speaker, only two of the marked datives are formally canonical (*diesem Unfall* in 8a and *der Straße* in 8d), with (8d) being semantically infelicitous.<sup>14</sup>

- (8) a. LangSit formal-spoken  
**mein-em erfahrung mit dies-em Unfall**  
 my-DAT experience with this-DAT accident  
 ‘my experience with this accident’
- b. LangSit formal-spoken  
**es gab auch ein-e ein-em Frau**  
 it gave also a-ACC a-DAT woman  
 ‘there was also a woman’
- c. LangSit formal-written  
**der man hatte auch ein-em fussball**  
 the man had also a-DAT soccer ball  
 ‘the man also had a soccer ball’

<sup>13</sup>To enhance readability, we restrict our glosses to the point under discussion, currently case marking.

<sup>14</sup>Canonical alternatives would have been *die Dame auf der Straße gegenüber* ‘the lady on the street across’, ...*auf der gegenüberliegenden Straßenseite* ‘...on the across lying street side’.

- d. LangSit informal-written  
die dame über der straße hatte ein-em Hund  
the lady over the.DAT street had a-DAT dog  
'the lady across the street had a dog'

Despite the overall spectrum of noncanonical inter- and intra-individual variation in case marking, the data show: HSs regularly, albeit sometimes noncanonically, mark DPs for various features. They implicitly know that a functional category is needed as a carrier, and they grasp detail of formal inventory, while the actual mapping and accuracy of choice in various online and offline tasks is a totally different matter.

## **6 Heritage speaker heterogeneity vs. individual systematicity**

Characterizations of HS groups point out their heterogeneity in terms of exposure, proficiency, and performance. While some of the HSs considered here are hardly distinguishable from MSs, especially highly literate ones, others produce strikingly different structures, both with respect to clause structure and morphology. Hence, quantitative analyses over such a highly diverse population blur relevant distinctions and may make us overlook the emergence of new grammatical systems. As a next step, we turn our attention to inter- and intra-individual variation and to qualitative analyses, first, for a subgroup of HSs and then for the productions of an individual speaker.

Our heritage German subcorpus includes nine sets of siblings, making it possible to compare participants who grew up under similar linguistic and extra-linguistic circumstances and were exposed to (presumably) similar input on part of the parents, at least initially (see Aalberse et al. 2019, Bridges & Hoff 2014, Shin 2002 for effects of birth order on HL input and proficiency). For younger siblings, the probability of older siblings and eventually parents speaking the ML at home as well certainly increases. We briefly consider the productions of three siblings (two brothers and one sister: the oldest brother being 18 years old, the sister 17 years, and the youngest brother 14 years<sup>15</sup>). All three were born in the US and, at the time of elicitation, they still lived in the same household with their parents. Their mother was born in Germany and their father in the US, and they reported that their parents both spoke German and English at home. All three

---

<sup>15</sup>The speaker codes of the three siblings in the RUEG corpus, are: USbi74MD (oldest brother), USbi72FD (sister), USbi73MD (youngest brother).

siblings rated themselves native speakers of German and English. In the following, we present selected results across several production tasks. While small, the number of instances of clausal structures is sufficient to infer basic patterns.

Table 2: Total number of clauses, total number of finite SCs in brackets per sibling and LangSit narration, and mean percentages of SCs across productions.

|                  | older<br>brother |       |        | younger<br>brother |
|------------------|------------------|-------|--------|--------------------|
|                  |                  | total | mean % |                    |
| formal-written   | 25               | (9)   | 14 (4) | 11 (1)             |
| informal-written | 7                | (1)   | 11 (1) | 5 (0)              |
| formal-spoken    | 39               | (9)   | 12 (1) | 16 (2)             |
| informal-spoken  | 15               | (6)   | 21 (2) | 11 (0)             |
| total            | 86               | (25)  | 58 (8) | 43 (3)             |
|                  |                  |       | 29.1%  | 13.8%              |
|                  |                  |       |        | 7.0%               |

Table 2 suggests two patterns: Firstly, we find that with decreasing age of the siblings, the narrations become shorter. Secondly, the overall proportion of finite SCs in comparison to matrix clauses decreases as well. Concerning the canonicity of the siblings' clause productions across LangSit narrations, the data show that the older brother and the sister produce exclusively canonical matrix and subordinate clauses. The younger brother produces four noncanonical V3 structures in matrix clauses. Two of these are the result of a preposed SC (9a) which, if placed in the forefield, needs to be immediately followed by the finite verb to maintain canonical V2 structure in the matrix clause. The other two cases occur with an adverbial in the prefield (9b).<sup>16</sup>

In a next step, we analyzed the siblings' syntax in the oral sentence completion task. The results show that the older brother and the sister again did not produce a single structurally noncanonical sentence, hence we do not list them here. In the younger brother's data one noncanonical clause (9c) follows the same clausal pattern as in his LangSit narrations.

- (9) a. LangSit formal-written

- als die Autos gestoppt war-en ein Hund ist weg-gerannt  
 when the cars stopped were-3PL a dog is<sub>V3</sub> away-run  
 'when the cars stopped, a dog ran away'

<sup>16</sup>In order to highlight our relevant point, we added the index V3 to the English gloss.

- b. LangSit informal-spoken  
und dann des auto hinter des erste auto is in des andere  
and then the car behind the first car is<sub>V3</sub> in the other  
rein-gefahren  
in-driven  
'and then the car behind the first car drove into the other one'  
c. Als das erste Auto bremste, ... 'When the first car braked, ...'  
das andere ist hinten rein-gefahren.  
the other is<sub>V3</sub> behind in-driven.  
'...the other-one hit it from behind.'

The siblings also took part in the online written sentence completion task with the very same stimuli administered a few months later. Their written sentence completion mirrors the results of the corresponding oral sentence completion task. Here as well, the older brother and the sister showed no noncanonical verb placement. In the data of the younger brother, however, there were several non-canonical sequels, both with respect to case marking (10a) and word order (10b-c).

- (10) a. Aus der Tasche der Frau mit dem Auto...  
'Out of the bag of the woman with the car...'  
ist ein-em Apfel aus der Tasche gefallen.  
is an-DAT apple out the bag fell.  
'...an apple fell out.'
- b. Nachdem der Ball über die Straße rollte, ...  
'After the ball rolled across the street, ...'  
der Hund ist<sub>V3</sub> zu dem Ball gesprungen.  
the dog is to the ball jumped  
'After the ball rolled across the street, the dog jumped to the ball.'
- c. Als das erste Auto bremste, ...  
'When the first car braked, ...'  
das andere is<sub>V3</sub> hinten rein-gefahren.  
the other is behind in-driven.  
'When the first car braked, the other one hit it from behind.'

Results across the three tasks show the smallest number of noncanonical productions in the older brother, followed by his younger sister. Most noncanonical structures are produced by the youngest brother, and all point in the expected direction: Noncanonical patterns predominantly result from placement of the finite

verb not far enough towards the left periphery, i.e., V3 instead of V2. The same tendency and difference between the siblings also manifest themselves in the sentence correction task – not illustrated here for reasons of space. The crucial point is that even within a single family, i.e., a Tiny Language Island scenario, we see individual differences – in line with what previous research has found on birth order effects on HL productions – pointing in a direction relevant to our discussion on clause structure. Only one of the siblings exhibits a clear shift towards V3 clauses with a complex prefield. While the result is a parallel with English, the main verb and its particle are still separated by other constituents: the remnant of a shrunken middlefield.

In the following, we take the Tiny Island scenario one step further and consider data from a 17-year-old female adolescent HS<sup>17</sup>, whose German input only comes from the mother. In her LangSit narrations, her oral sentence completions, and all throughout the informal chitchat, she considerably – and well beyond what we saw above – diverges from canonical patterns, both in word order and morphological subsystems, including an idiosyncratic spell-out of subject-verb agreement. The data show noncanonical third person singular verbs in contexts calling for plurals (11a–b), with (11a) potentially strengthened by English *were*. The utterances in (11c–d) also show noncanonical first person singular and plural inflections. We will turn to the “?” in the interlinear glosses in (11b) shortly.

- (11) a. LangSit formal-written
 

|                 |        |      |
|-----------------|--------|------|
| all-e           | war    | okay |
| all-PL          | was.SG | okay |
| 'all were okay' |        |      |
- b. LangSit formal-spoken
 

|  |                  |      |     |      |      |     |      |     |         |        |        |   |
|--|------------------|------|-----|------|------|-----|------|-----|---------|--------|--------|---|
| ein mann   | er <sup>18</sup> | frau | und | er   | baby | und | die  | hat | zu      | eine   | straße |   |
| a  | man              | ?    |     | wife | and  | ?   | baby | and | they.PL | has.SG | to     | a |
| gelaufen   |                  |      |     |      |      |     |      |     |         |        |        |   |
| street   | walked           |      |     |      |      |     |      |     |         |        |        |   |
| 'a man, his wife and his baby walked towards a street' |                  |      |     |      |      |     |      |     |         |        |        |   |
- c. Chitchat
 

|                                    |        |            |         |      |       |         |
|------------------------------------|--------|------------|---------|------|-------|---------|
| Ich                                | gern   | geh-t-st   | draußen | mit  | meine | Freunde |
| I                                  | gladly | go-3SG-3SG | outside | with | my    | friends |
| 'I like going out with my friends' |        |            |         |      |       |         |
- d. Chitchat (topic: cooking chili)

---

<sup>17</sup>The speaker code of this participant is USbi77FD.

wir    **nehm-st** Bier  
we.1PL take-3SG beer  
'we take beer'

Taking into account all auxiliaries<sup>19</sup> across elicitation tasks, a clear pattern emerges: *haben* 'have' predominates, irrespective of contextual requirements of both standard German and dialects. The examples in (10a–b) show this kind of leveling of the auxiliary inventory, possibly under the influence of cross-linguistic transfer from English. The noncanonical suffixes in (11c+d) are based on noncanonical person markers: a 2<sup>nd</sup> Ps.Sg. added to an already 3<sup>rd</sup> Ps.Sg.-marked verb in (11c) and added to a noncanonical stem form in (11d). This overgeneralization pattern is quite unexpected and uncommon for L1 acquisition (Aalberse 2009: 26, Aalberse & Stoop 2015: 193).

In (12a), a very fluently produced utterance, we see multiple departures from canonical German, including calques based on English words and collocations. Also, in the same example, word order in the relative clause follows an English pattern, as does the sequel after the complex PP stimulus in the oral sentence completion task in (12b).

- (12) a. Chitchat
- ich weiß nich alle diesen platz auf mein kopf wo ich **habe**  
I know not all this place on my head where I have  
gegangen  
gone  
'I don't remember all the places I have been on top of my head'
- b. Aus der Tasche der Frau mit dem Auto... 'Out of the bag of the woman with the car...'   
Äpfel oder etwas **hat** gefallen und das Mann **hat** gerennen  
apples.PL or something has.SG fallen and the man has run  
'...apples or something fell out and the man ran.'

Despite many clearly noncanonical and not even dialectally licensed features of the young woman's written and spoken productions, a number of relevant properties can be identified. Examples (13a–13d) illustrate the idiosyncratic but consistent creation of what we hypothesize to be a possessive placeholder. Crucial insight in support of this interpretation comes from her written narratives.

<sup>18</sup>We return to the identification of *er* as a potential placeholder later on.

<sup>19</sup>German exhibits auxiliary alternation between *haben* 'have' and *sein* 'be' depending on the property of the main verb. English, on the other hand, always uses 'have' in active clauses.

Just on the basis of spoken versions alone, one could have easily dismissed the syllables preceding *frau*, *baby*, and *ball* as hesitations or filler particles. But as the written version (13b) shows, there is a visible orthographic exponent. Moreover, when we checked the participant's English narratives, we found our interpretation of the kind of concepts she wanted to express corroborated, as shown in (13c) and (13d).

- (13) a. LangSit formal-spoken  
ein mann **er** frau und **er** baby  
a man ? wife and ? baby  
'a man, his wife, and his baby'
- b. LangSit formal-written  
es gibt ein par leute ein mann, **er** frau un kind  
it gives a few people a man ? woman and child  
'There were a few people. A man, his wife and child'
- c. LangSit formal-spoken  
there was a dude **his** wife and **his** baby
- d. LangSit formal-written  
there was a man with **his** wife and baby

The hypothesis that *er* in German fulfills a syntactic dummy function aligns with what we know from research on L1, 2L1, and L2 acquisition in children and on L2 adults on other placeholder phenomena (Daskalaki et al. 2019, Tracy 2011, see also various papers in Blom et al. 2013).

More importantly with respect to our current focus on clause structure: In this HS's productions, regardless of type of communicative context, of register, and of the type of task, word order is predominantly noncanonical, here selectively illustrated with (14a), with only non-finite verbs (here infinitives and participles) placed in final position and at a distance from the finite verb, as shown in (14b).

- (14) a. Chitchat  
aber das ist nicht was ich **will** tun wenn ich **bin** alt  
but this is not what I want do when I am old  
'but this is not what I want to do when I am old'
- b. Chitchat  
dass ich **habe** zu einundzwanzig countries gegangen  
that I have to twenty-one countries gone  
'that I have visited 21 countries'

Finally, this speaker is one of the very few participants who extrapose a direct object (15).

(15) Chitchat

- dass ich gerne esst chili  
that I gladly eat chili  
'that I like eating chilli'

What makes the overall idiosyncratic but also internally systematic spectrum of noncanonical structures in this adolescent speaker particularly valuable is that it is like a fast forward into patterns familiar from established language islands. After all, many older speakers in what once used to be vibrant German speech communities are on their own way towards Tiny Language Islands since they may well be the only HSs left within their family.

## 7 Discussion and conclusion

We hope to have shown that HS research offers a promising testing ground for exploring various intriguing phenomena which have been puzzling language acquisition researchers for a long time. In this last section, we recapitulate our main points.

Our first research question focused on the retention of clausal syntax in the face of language contact. Our findings show, with the exception of a few individuals, predominantly canonical productions across HSs. As we know from early phases of word combining in typically developing German-speaking children, OV[-fin] is acquired before age two (Schulz & Tracy 2018). This may explain why, in the overwhelming number of cases, the direct objects produced by adolescent HSs occur in their canonical position, as predicted by the Interface Hypothesis. Thus, with respect to core syntactic, early acquired features, we find no major transfer effects from the ML.

A different picture emerged when we focused on the morphological features of HSs' productions, addressing research question two. Qualitative analyses showed noncanonical variation in case and gender marking, i.e., phenomena at the interface of morphology, syntax, and semantics.

Concerning our last research question, which focused on inter- and intra-individual variation, our findings fit the general picture emerging from research on traditional language islands and from current HL investigations with respect to the heterogeneity of speaker profiles (Aalberse et al. 2019, Montrul & Polinsky 2021, Polinsky 2018a,b, Wiese et al. 2022 for other languages represented in

the RUEG project). In the majority of the heritage German speakers taking part in our study, core-syntactic properties related to finite and non-finite verb placement in main and subordinate clauses – especially the sentence bracket – proved relatively stable. At the same time, we also saw differences in speakers' output even in cases where they grew up within the same family and were (most likely) exposed to qualitatively similar input. A subset of crucial properties of German clause structure was still detectable in the final case study we selected for illustration: a speaker whose grammatical system had considerably diverged from canonical German patterns, both in the syntax and in morphological spell-out. Importantly, variation proved intra-individually systematic.

We conclude that there are certainly many *dots to connect*, as also formulated in Montrul (2018), between research on language acquisition in early childhood and attempts at reconstructing the grammatical systems available to adolescent and older HSs (Polinsky 2018a). After all, typically developing monolingual and bilingual children also pass along trajectories where they produce noncanonical main and subordinate clauses and noncanonical inflectional morphology.<sup>20</sup> Since these phases may be short or undocumented, they might easily go by unnoticed in cross-sectional investigations without access to individual longitudinal data.

It does not take much imagination to think of what might happen in typically developing L1 children for whom temporarily divergent paths have been documented in longitudinal research, had they been transported into a different ML setting before their grammars had converged on target states. In that case, their intermediate solutions might easily have been there to stay, for instance the regularization of irregular forms amply documented in L1 acquisition for many languages (Yang 2016, Gawlitzeck-Maiwald 1994).

Our research design provided various opportunities for speakers to activate and demonstrate their linguistic resources across tasks with varying complexity: from free conversations to the elicitation of contents to report and the elicitation of particular target patterns, as in the spoken and written sentence completion task. This provided us, in turn, with opportunities for triangulation and guarded us against misconceptions which might have arisen had we only relied on one kind of data or on just one type of situation, or had we not included the productions of MSs of German and productions in the ML of HSs.

Our findings and our conclusions are also relevant beyond research on HSs of German, as shown by other contributions in this volume. Given the heterogeneity of background situations in terms of quantity and quality of exposure,

---

<sup>20</sup>For transitional phases with noncanonical clauses and idiosyncratic patterns in monolingual and bilingual children see Döpke 2000, Fritzenchaft et al. 1990, Gawlitzeck-Maiwald et al. 1994, Müller 1998, Rothweiler 1993, Tracy 2011.

personal preferences, and attitudes, it is quite impressive to what extent HSs, including those raised on Tiny Language Islands, are able to discover and maintain crucial properties of their parental baseline, i.e. of first-generation immigrants. Despite reduced HL exposure and the increasing relevance of the ML, in our case English, core-syntactic features, especially those acquired early, remain stable, even in contexts with reduced exposure to the HL, while interface phenomena are subject to variation and change from one generation to the next. The theoretically most relevant outcome of our study may be that both inter- and intra-individual variation are systematic, and that like any other acquisition scenario, HL acquisition and use contribute to our understanding of natural languages.

## Abbreviations

|       |   |      |                              |
|-------|---|------|------------------------------|
| 2L1   | Simultaneous acquisition of two first languages | ML   | Majority language            |
| AdvC  | Adverbial clause                                | MS   | Monolingually-raised speaker |
| CMC   | Coordinate main clause                          | PG   | Pennsylvania German          |
| CompC | Complement clause                               | RelC | Relative clause              |
| HL    | Heritage language                               | RSB  | Right sentence bracket       |
| HS    | Heritage speaker                                | SC   | (Finite) Subordinate clause  |
| IMC   | Independent main clause                         | SVO  | Subject-Verb-Object          |
| L1    | First language                                  | V1   | Verb-first-position          |
| L2    | Second language                                 | V2   | Verb-second-position         |
| LSB   | Left sentence bracket                           | VL   | Verb-last-position           |
| LWC   | Non-sentential light-weight constituent         |      |                              |

## Acknowledgements

We would like to thank the DFG for funding the research for this chapter (grant number: 394995401). We also owe thanks to two anonymous external reviewers for their much appreciated and encouraging comments on previous version of this chapter, as well as to our internal reviewers Tatiana Pashkova, Vassiliki Rizou, and Annika Labrenz. For helpful criticism on the pre-final version of our chapter we would like to thank our cooperation partner Dafydd Gibbon (University of Bielefeld). Our research would not have been possible without the help of our enthusiastic student research assistants. We would therefore like to thank

Stella Baumann, Lale Bilgili, Joshua Boivin, Ryan Carroll, Franziska Cavar, Anna Kuhn, Nils Picksak, Sam Schirm, Elena Unger, and Nadine Zürn. We additionally thank Lea Coy for help with pre-publication formatting.

## References

- Alberse, Suzanne P. 2009. *Inflectional economy and politeness: Morphology-internal and morphology-external factors in the loss of second person marking in Dutch*. Utrecht: LOT.
- Alberse, Suzanne P., Ad Backus & Pieter Muysken (eds.). 2019. *Heritage languages: A language contact approach* (Studies in Bilingualism 58). Amsterdam: John Benjamins. DOI: 10.1075/sibil.58.
- Alberse, Suzanne P. & Wessel Stoop. 2015. The exceptional loss of the pronoun T. *Journal of Pragmatics* 88. 190–201. DOI: 10.1016/j.pragma.2015.07.003.
- Andersen, Christiane. 2016. Nachfeld im Kontakt: Eine Korpusuntersuchung am Russlanddeutschen in Sibirien. *Göteborger Arbeitspapiere zur Sprachwissenschaft* (6). 1–16.
- Andreou, Maria, Jacopo Torregrossa & Christiane Bongartz. 2021. Sentence repetition task as a measure of language dominance. In Danielle Dionne & Lee-Ann Vidal Covas (eds.), *Proceedings of the 45th annual Boston University Conference on Language Development*, vol. 1, 14–25. Sommerville, MA: Cascadilla Press.
- Antomo, Mailin & Markus Steinbach. 2010. Desintegration und Interpretation: Weil-V2-Sätze an der Schnittstelle zwischen Syntax, Semantik und Pragmatik. *Zeitschrift für Sprachwissenschaft* 29(1). 1–37. DOI: 10.1515/zfsw.2010.001.
- Blom, Elma, Ineke Craats & Josje Verhagen. 2013. *Dummy auxiliaries in first and second language acquisition*. Berlin: De Gruyter. DOI: 10.1515/9781614513476.
- Boas, Hans C. 2009a. Case loss in Texas German: The influence of semantic and pragmatic factors. In Jóhanna Barðdal & Shobhana L. Chelliah (eds.), *The role of semantic, pragmatic, and discourse factors in the development of case* (Studies in Language Companion Series 108), 347–373. Amsterdam: John Benjamins. DOI: 10.1075/slcs.108.18boa.
- Boas, Hans C. 2010. On the equivalence and multifunctionality of discourse markers in language contact situations. In Theo Harden & Elke Hentschel (eds.), *Partikelforschung*, 301–315. Tübingen: Stauffenburg.
- Boas, Hans Christian. 2009b. *The life and death of Texas German*. Durham, NC: Duke University Press.

- Böttcher, Marlene & Margaret Zellers. 2023. Hesitating with and without language heritage: Prosodic aspects of filler particles in the RUEG corpus. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, 2795–2799. Prague: The International Phonetic Association.
- Bridges, Kelly & Erika Hoff. 2014. Older sibling influences on the language environment and language development of toddlers in bilingual homes. *Applied Psycholinguistics* 35(2). 225–241. DOI: 10.1017/S0142716412000379.
- Bunk, Oliver. 2020. „Aber immer alle sagen das“: The status of V3 in German: Use, processing, and syntactic representation. Humboldt-Universität zu Berlin. (Doctoral dissertation). DOI: 10.18452/22085.
- Bunk, Oliver, Shanley E. M. Allen, Sabine Zerbian, Tatiana Pashkova, Yulia Zuban & Erica Conti. 2025. Information packaging and word order dynamics in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 379–412. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775181.
- Clahsen, Harald. 1984. Der Erwerb von Kasusmarkierungen in der deutschen Kindersprache. *Linguistische Berichte* (89). 1–31.
- Clyne, Michael C. 2003. *Dynamics of language contact: English and immigrant languages* (Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511606526.
- Clyne, Michael G. 1987. Constraints on code-switching: How universal are they? *Linguistics* (25). 739–764.
- Daskalaki, Evangelia, Vasiliki Chondrogianni, Elma Blom, Froso Argyri & Jóhanne Paradís. 2019. Input effects across domains: The case of Greek subjects in child heritage language. *Second Language Research* 35(3). 421–445. DOI: 10.1177/0267658318787.
- Döpke, Susanne. 2000. *Cross-linguistic structures in simultaneous language acquisition*. Amsterdam: John Benjamins.
- Drach, Erich. 1963. *Grundgedanken der deutschen Satzlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Eichinger, Ludwig M., Albrecht Plewnia & Claudia M. Riehl (eds.). 2008. *Handbuch der deutschen Sprachminderheiten in Mittel- und Osteuropa*. Tübingen: Narr Francke Attempto Verlag.
- Flores, Cristina. 2020. Attrition and reactivation of a childhood language: The case of returnee heritage speakers. *Language Learning* 70(S1). 85–121. DOI: 10.1111/lang.12350.

- Földes, Csaba. 2016. Ungarndeutsche Sprachvariation und Mehrsprachigkeit: Ein Korpusprojekt auf der Basis von empirischer Feldforschung und Online-Sprachdokumentation. *Sprachtheorie und germanistische Linguistik* 26(2). 167–190.
- Freywald, Ulrike. 2016. Clause integration and verb position in German: Drawing the boundary between subordinating clausal linkers and their paratactic homonyms. *Linguistische Berichte Sonderheft* 21. 181–220.
- Freywald, Ulrike, Heike Wiese, Hans C. Boas, Katharina Brizić, Antje Dammel & Stephan Elspaß. 2023. *Deutsche Sprache der Gegenwart*. Stuttgart: J.B. Metzler. DOI: 10.1007/978-3-476-04921-6.
- Fritzenschaft, Agnes, Ira Gawlitzek & Rosemarie Tracy. 1990. Wege zur komplexen Syntax. *Zeitschrift für Sprachwissenschaft* (9). 52–134.
- Fuller, Janet M. 2001. The principle of pragmatic detachability in borrowing: English-origin discourse markers in Pennsylvania German. *Linguistics* 39(2). 351–369. DOI: 10.1515/ling.2001.014.
- Gawlitzek-Maiwald, Ira. 1994. How do children cope with language variation in the input? The case of German plurals and compounding. In Rosemarie Tracy & Elsa Lattey (eds.), *How tolerant is Universal Grammar?* (Linguistische Arbeiten), 225–266. Tübingen: Niemeyer.
- Gawlitzek-Maiwald, Ira, Rosemarie Tracy & Agnes Fritzenschaft. 1994. Language acquisition and competing linguistic representations: The child as arbiter. In Jürgen M. Meisel (ed.), *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition* (Studies in Theoretical Psycholinguistics 16), 139–179. Dordrecht: Kluwer.
- Haider, Hubert. 2010. *The syntax of German*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511845314.
- Höhle, Tilman N. 1986. Der Begriff „Mittelfeld“: Anmerkungen über die Theorie der topologischen Felder. In Walter E. Weiss, Herbert E. Wiegand & Marga Reis (eds.), *Kontroversen, alte und neue: Akten des VII. Internationalen Germanisten-Kongresses Göttingen 1985*, 329–340. Tübingen: Niemeyer.
- Hopp, Holger & Michael T. Putnam. 2015. Syntactic restructuring in heritage grammars: Word order variation in Moundridge Schweitzer German. *Linguistic Approaches to Bilingualism* 5(2). 180–214. DOI: 10.1075/lab.5.2.02hop.
- Housen, Alex, Folkert Kuiken & Ineke Vedder (eds.). 2012. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Language Learning & Language Teaching 32). Amsterdam: John Benjamins. DOI: 10.1075/llt.32.
- Huffines, Marion Lois. 1980. Pennsylvania German. *International Journal of the Sociology of Language* (25). 43–57.

- Imo, Wolfgang. 2015. Nachträge im Spannungsfeld von Medialität, Situation und interaktionaler Funktion. In Hélène Vinckel-Roisin (ed.), *Das Nachfeld im Deutschen*, 231–254. De Gruyter.
- Imo, Wolfgang. 2016. *Grammatik: Eine Einführung*. Stuttgart: J.B. Metzler.
- Johannessen, Janne Bondi & Joseph C. Salmons. 2015. The study of Germanic heritage languages in the Americas. In Janne Bondi Johannessen & Joseph C. Salmons (eds.), *Germanic heritage languages in North America*, 1–18. Amsterdam: John Benjamins.
- Johanson, Lars. 2002. Contact-induced change in a code-copying framework. In Mari C. Jones & Edith Esch (eds.), *Language change* (Contributions to the Sociology of Language), 285–314. Berlin/New York: Mouton De Gruyter.
- Johanson, Lars. 2008. Remodeling grammar: Copying, conventionalization, grammaticalization. In Peter Siemund & Noemi Kintana (eds.), *Language contact and contact languages*, 61–80. Amsterdam/Philadelphia: John Benjamins.
- Keller, Mareike. 2014. *Phraseme im bilingualen Diskurs: „All of a sudden geht mir ein Licht auf.“*. Frankfurt a.M.: Peter Lang Edition. DOI: 10.3726/978-3-653-04258-0.
- Louden, Mark Laurence. 2008. Synthesis in Pennsylvania German language and culture. In Josef Raab & Jan Wırer (eds.), *Die deutsche Präsenz in den USA*, 671–699. Münster: LitVerlag.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehayé, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Molnár, Valéria & Susanne Winkler. 2010. Edges and gaps: Contrasts at the interfaces. *Lingua* 120(6). 1392–1415.
- Montrul, Silvina. 2011. Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition* 33(2). 163–192.
- Montrul, Silvina. 2018. Heritage language development: Connecting the dots. *International Journal of Bilingualism* 22(5). 530–546. DOI: 10.1177/1367006916654368.
- Montrul, Silvina & Maria Polinsky (eds.). 2021. *The Cambridge handbook of heritage languages and linguistics*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108766340.

- Müller, Natascha. 1998. Transfer in bilingual first language acquisition. *Bilingualism: Language and Cognition* 1(3). 151–171.
- Paradis, Johanne, Brian Rusk, Tamara Sorenson Duncan & Krithika Govindarajan. 2017. Children's second language acquisition of English complex syntax: The role of age, input, and cognitive factors. *Annual Review of Applied Linguistics* 37. 148–167. DOI: 10.1017/S0267190517000022.
- Pashkova, Tatiana, Wintai Tsehaye, Shanley E. M. Allen & Rosemarie Tracy. 2022. Syntactic optionality in heritage language use: Clause type preferences of German heritage speakers in a majority English context. *Heritage Language Journal* 19(1). 1–41. DOI: 10.1163/15507076-12340022.
- Pickering, Martin & Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(2). 169–226.
- Plewnia, Albrecht & Claudia M. Riehl (eds.). 2018. *Handbuch der deutschen Sprachminderheiten in Übersee*. Tübingen: Narr Francke Attempto Verlag.
- Polinsky, Maria. 2018a. Bilingual children and adult heritage speakers: The range of comparison. *International Journal of Bilingualism* 22(5). 547–563. DOI: 10.1177/1367006916656048.
- Polinsky, Maria. 2018b. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Proske, Nadine. 2015. Die Rolle komplexer Nachfeldbesetzungen bei der Einheitenbildung im gesprochenen Deutsch. In Hélène Vinckel-Roisin (ed.), *Das Nachfeld im Deutschen*, 279–298. De Gruyter. DOI: 10.1515/9783110419948-014.
- Putnam, Michael T. & Joseph C. Salmons. 2013. Losing their (passive) voice: Syntactic neutralization in heritage German. *Linguistic Approaches to Bilingualism* 3(2). 233–252. DOI: 10.1075/lab.3.2.05put.
- Rosenberg, Peter. 2005. Language island research: The traditional framework and some sociolinguistic questions. In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect change*, 221–235. Cambridge: Cambridge University Press.
- Rothweiler, Monika. 1993. *Der Erwerb von Nebensätzen im Deutschen: Eine Pilotstudie* (Linguistische Studien 302). Tübingen: Niemeyer.
- Schmid, Monika. 2011. *Language attrition*. Cambridge: Cambridge University Press.
- Schulz, Petra & Rosemarie Tracy. 2018. Revisiting the tolerance of universal grammar. In Bart Hollebrandse, Jaieun Kim, Ana T. Perez-Leroux & Petra Schulz (eds.), *T.O.M. and grammar: Thoughts on mind and grammar*, 129–145. University of Massa.
- Shah, Sheena & Marianne Zappern-Thomson (eds.). 2017. *Heritage language policies around the world*. 1st edition. London: Routledge.

- Shin, Sarah J. 2002. Birth order and the language experience of bilingual children. *TESOL Quarterly* 36(1).
- Sluckin, Benjamin Lowell & Oliver Bunk. 2023. Noncanonical V3 and resumption in Kiezdeutsch. In Karen de Clercq, Liliane Haegeman, Terje Lohndal & Christine Meklenborg (eds.), *Adverbial resumption in verb second languages*, 327–354. Oxford: Oxford University Press. DOI: 10.1093/oso/9780197651148.003.0014.
- Speyer, Augustin. 2009. Doppelte Vorfeldbesetzung im heutigen Deutsch und im Frühneuhochdeutschen. *Linguistische Berichte* (2016). 455–485.
- Stolberg, Doris. 2014. Changing argument structure in (heritage) Pennsylvania German. *Applied Linguistics Review* 5(2). 329–352. DOI: 10.1515/applirev-2014-0015.
- Stolberg, Doris. 2015. *Changes between the lines: Diachronic contact phenomena in written Pennsylvania German*. Berlin/New York: De Gruyter Mouton.
- te Velde, John R. 2017. Temporal adverbs in the Kiezdeutsch left periphery: Combining late merge with deaccentuation for V3. *Studia Linguistica* 71(3). 301–336. DOI: 10.1111/stul.12055.
- Tracy, Rosemarie. 1986. The acquisition of case morphology in German. *Linguistics* 24(1). DOI: 10.1515/ling.1986.24.1.47.
- Tracy, Rosemarie. 2011. Konstruktion, Dekonstruktion und Rekonstruktion: Minimalistische und (trotzdem) konstruktivistische Überlegungen zum Spracherwerb. In Stefan Engelberg, Anke Holler & Kristel Proost (eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik*, 397–428. Berlin: De Gruyter. DOI: 10.1515/9783110262339.397.
- Tracy, Rosemarie. 2022. Gemischtsprachiges Sprechen: Formen, Funktionen, Dynamik. In Csaba Földes & Thorsten Roelcke (eds.), *Handbuch Mehrsprachigkeit*, 399–428. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110623444-018.
- Tracy, Rosemarie & Dafydd Gibbon. 2023. The beat goes on: A case study of timing in heritage German prosody. In Michael Beißwenger, Eva Gredel, Lothar Lemnitzer & Roman Schneider (eds.), *Korpusgestützte Sprachanalyse: Linguistische Grundlagen, Anwendungen und Analysen* (Studien zur Deutschen Sprache 88), 261–283. Tübingen: Narr.
- Tracy, Rosemarie & Elsa Lattey. 2001. Language contact in the individual: A case study based on letters from a German immigrant in New Jersey. In Per Sture Ureland (ed.), *Global Eurolinguistics: European languages in North America – migration, maintenance and death*, 413–433. Tübingen: Niemeyer.
- Tracy, Rosemarie & Elsa Lattey. 2010. “It wasn’t easy but irgendwie äh da hat sich’s rentiert, net?”: A linguistic profile. In Michaela Albl-Mikasa (ed.), *Dimensionen der Zweitsprachenforschung*, 53–73. Tübingen: Narr.

- Tracy, Rosemarie & Doris Stolberg. 2008. Nachbarn auf engstem Raum: Koexistenz, Konkurrenz und Kooperation im mehrsprachigen Kopf. In Ludwig M. Eichinger & Albrecht Plewnia (eds.), *Das Deutsche und seine Nachbarn* (Studien zur deutschen Sprache), 83–108. Tübingen: Narr.
- Tsehaye, Wintai. 2023. Light-weights placed right: Post-field constituents in heritage German. *Frontiers in Psychology* 14. DOI: 10.3389/fpsyg.2023.1122129.
- Tsehaye, Wintai, Tatiana Pashkova, Rosemarie Tracy & Shanley E. M. Allen. 2021. Deconstructing the native speaker: Further evidence from heritage speakers for why this horse should be dead! *Frontiers in Political Science* 12. 1–14. DOI: 10.3389/fpsyg.2021.717352.
- Tsimpli, Ianthi-Maria. 2014. Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313. DOI: 10.1075/lab.4.3.01tsi.
- Vainikka, Anne & Martha Young-Scholten. 2011. *The acquisition of German*. Berlin: De Gruyter.
- Vasilyeva, Marina, Heidi Waterfall & Janellen Huttenlocher. 2008. Emergence of syntax: Commonalities and differences across children. *Developmental Science* 11(1). 84–97. DOI: 10.1111/j.1467-7687.2007.00656.x.
- Vinckel-Roisin, Hélène (ed.). 2015. *Das Nachfeld im Deutschen*. De Gruyter. DOI: 10.1515/9783110419948.
- Walkden, George. 2017. Language contact and V3 in Germanic varieties new and old. *Journal of Comparative Germanic Linguistics* 20(1). 49–81. DOI: 10.1007/s10828-017-9084-2.
- Westphal-Fitch, Gesche. 2011. Changes in frequency as a measure of language change. In Michael T. Putnam (ed.), *Studies on German-language islands* (Studies in Language Companion Series 123), 371–384. Amsterdam: John Benjamins. DOI: 10.1075/slcs.123.14wes.
- Wiese, Heike. 2013. What can new urban dialects tell us about internal language dynamics? The power of language diversity. In Werner Abraham & Elisabeth Leiss (eds.), *Dialektologie in neuem Gewand: Zu Mikro-/Varietätenlinguistik, Sprachenvergleich und Universalgrammatik* (Linguistische Berichte Sonderheft 19), 207–245. Hamburg: Buske.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagariana, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehaye, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.

- Wiese, Heike, Shanley E. M. Allen, Mareike Keller & Artemis Alexiadou. 2025. Introduction: Investigating the dynamics of language contact situations. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 1–29. Berlin: Language Science Press. DOI: 10.5281/zenodo.1577515.
- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 33–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.1577515.
- Wiese, Heike & Hans Georg Müller. 2018. The hidden life of V3: An overlooked word order variant on verb-second. In Mailin Antomo & Sonja Müller (eds.), *Non-canonical verb positioning in main clauses* (Linguistische Berichte Sonderheft 25), 202–223. Hamburg: Buske.
- Wiese, Heike, Horst Simon, Christian Zimmer & Kathleen Schumann. 2017. German in Namibia: A vital speech community and its multilingual dynamics. Péter Maitz & Craig A. Volker (eds.). 221–245.
- Winkler, Julia. 2017. V3-Stellung im Deutschen: Wettbewerb um das Vorfeld. *Linguistische Berichte* (250). 139–168.
- Wöllstein, Angelika. 2014. *Topologisches Satzmodell*. Heidelberg: Winter.
- Yager, Lisa, Nora Hellmold, Hyoun-A Joo, Michael T. Putnam, Eleonora Rossi, Catherine Stafford & Joseph Salmons. 2015. New structural patterns in moribund grammar: Case marking in heritage German. *Frontiers in Psychology* 6. 1–9. DOI: 10.3389/fpsyg.2015.01716.
- Yang, Charles. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.
- Zifonun, Gisela. 2015. Der rechte Rand in der IDS-Grammatik: Evidenzen und Probleme. In Hélène Vinckel-Roisin (ed.), *Das Nachfeld im Deutschen*, 25–52. De Gruyter.
- Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker (eds.). 1997. *Grammatik der deutschen Sprache* (Schriften des Instituts für deutsche Sprache 7). Berlin/New York: De Gruyter.
- Zimmer, Christian. 2019. Deutsch als Minderheitensprache in Afrika. In Joachim Herrgen & Jürgen Erich Schmidt (eds.), *Sprache und Raum: Ein internationales Handbuch der Sprachvariation* (Handbücher zur Sprach- und Kommunikationswissenschaft), 1176–1190. Berlin: De Gruyter Mouton.
- Zimmer, Christian. 2020. Kasus im Namdeutschen. *Zeitschrift für Germanistische Linguistik* 48(2). 298–335.

Zürn, Nadine, Mareike Keller, Rosemarie Tracy & Anke Lüdeling. 2025. Dynamic properties of the heritage speaker lexicon. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 325–375. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775179.



# Chapter 6

## Null subjects in heritage Greek, Russian and Turkish

Onur Özsoy<sup>a,b</sup>, Vasiliki Rizou<sup>b</sup>, Maria Martynova<sup>b</sup>, Natalia Gagarina<sup>a,b</sup>, Luka Szucsich<sup>b</sup> & Artemis Alexiadou<sup>a,b</sup>

<sup>a</sup>Leibniz-Centre General Linguistics <sup>b</sup>Humboldt-Universität zu Berlin

Greek, Russian, and Turkish represent three different types of null subject languages: Greek is characterized as a strict pro-drop language; Russian is claimed to be either a partial pro-drop, or a non-pro-drop language with abundant subject elipsis; Turkish is classified as a pro-drop or topic pro-drop language in which overt subjects are required in unspecified contexts. This chapter reports results from corpus-based studies exploring the question of whether subjects in heritage varieties of these three languages in contact with English and German pattern alike regarding the realization of overt subjects. Moreover, the chapter examines whether heritage speakers of these three languages apply topic drop more widely in one of their majority languages, namely German. We consider different possibilities for our observations reaching from language contact effects, to language internal dynamics and most importantly a typological continuum within the different pro-drop types regarding the stability and dynamicity in their pro-drop systems.

### 1 Introduction

The main aim of this chapter is to examine the phenomenon of null subjects in Greek, Russian, and Turkish majority and heritage language varieties using a comparative approach. The study analyzes data from both monolingual and bilingual populations, where the latter are heritage speakers. Heritage speakers are typically individuals who use their heritage language(s), learned during early childhood in their families, in addition to the majority language(s) spoken in



the host community (cf. among many others Valdés 2005, Rothman 2009, Benmamoun et al. 2013, Guijarro-Fuentes & Schmitz 2015, Polinsky 2015, Montrul 2015). Heritage speakers are considered bilinguals belonging to a nativeness continuum as native speakers of their heritage language, but the proficiency in the languages they speak can vary throughout their lifetime (Benmamoun et al. 2013, Polinsky 2018, Wiese et al. 2022). In this chapter, we adopt this perspective and we do not compare the heritage data to the monolingual data as a measure of accuracy. Rather, we view the heritage and monolingual populations as belonging to distinct language varieties in their own right, in line with Rothman et al.'s (2023) perspective.

Additionally, we aim to gain an understanding of subject realization in the heritage languages Greek, Russian, and Turkish, which differ typologically, and investigate whether they are influenced by contact with the majority languages English and German. The latter are both characterized as non-pro-drop languages with German allowing for the omission of subjects in topic positions (Oppermann 2021, Schäfer 2021, Trutkowski 2016). Homeland varieties of Greek, Russian and Turkish all allow subject omission to varying degrees. Heritage Greek, Russian, and Turkish have been argued to exhibit similar tendencies regarding their pro-drop properties. Notably, overtly realized pronominal subjects have been found to be more frequent in heritage varieties, particularly those in contact with Germanic languages, than in monolingual ones. This has been discussed in various studies such as Argyri & Sorace (2007) and Tsimpli et al. (2004) for heritage Greek; Dubinina & Polinsky (2013), Gagarina (2012), and Ivanova-Sullivan (2014, 2015) for heritage Russian; and Haznedar (2010) and Koban Koç (2016) for heritage Turkish (but see Azar et al. 2020 for results showing that heritage and monolingual speakers of Turkish omit subjects to a similar degree).

While previous studies have mostly focused on the heritage language of these speakers, there are very few studies that investigate the majority language too. This chapter adds to these studies and discusses whether HSs of Greek, Russian, and Turkish in Germany apply topic drop more widely in their majority language German. An expansion of topic drop would represent a creative extension of German grammar which points to a reorganization at the pragmatics-syntax interface.

The investigation of subject realization in heritage Greek, Russian, and Turkish in the US and Germany, and of topic drop in majority German is a promising area of study since it relates to the interface between internal linguistic domains (core grammatical system) and external linguistic domains (information structure and discourse organization). To achieve the above-mentioned objectives, the study uses a multi-factorial analysis of semi-naturalistic and ecologically more

valid data from monolingual and heritage speakers, taking into account differences in formality and mode.

This study also extends previous research in directly addressing the effect on subject realization of animacy, mode, and formality – three factors which are meaningful to subject realization but are not often considered in this field. Some studies investigating languages unrelated to the languages of our study point to animacy playing an important role in subject realization. Animacy is an inherent feature of nominal referents (Corbett 1979, Comrie 1989). Numerous scholars have classified animacy on a hierarchy (Silverstein & Dixon 1986), placing it on a continuum with animate human discourse participants on the one end and inanimate and abstract referents on the other (cited from Corbett 2000: 90), as illustrated below:

- (1) speaker (first person pronouns) > addressee (second person pronouns) > third person > kin > human > animate > inanimate

Li et al. (2012) and Pu (1997) report for the radical pro-drop language Mandarin that animate subjects and objects are more prone to be null subjects than inanimate ones. Similarly, for some creole languages like Tamambo, animacy plays a significant role in subject deletion with animate subjects being more likely to be omitted (Meyerhoff 2009). Finally, it was reported that subject omission in the Austronesian languages Vera'a and Teop is affected by animacy (Vollmer 2019).

Another uncharted territory that this study explores regarding null subjects is their relation to different communicative situations and speakers' repertoires. Previous studies have shown that certain text types such as short messages on social media or on the phone facilitate subject omission and ellipsis more generally. For example, Frick (2017) shows that in German text messages, there are more null pronouns than realized ones. This is atypical for German generally, but it is typical for this specific text genre. We are not aware of a cross-modality and cross-formality study of subject omission, which makes the investigations in this chapter pioneering and exploratory.

However, there is related previous work from the domain of nominal reference which refers to demonstrative-marked noun phrases in Greek, Russian and Turkish mono- and bilingual speakers. Martynova et al. (2024) show that different communicative settings impact the use of definite noun phrases in different ways. For instance, Greek speakers use more double definite structures in informal settings, Turkish speakers use demonstratives more in spoken than in written mode, and Russian speakers use demonstratives more frequently in informal and oral communication settings. These findings imply that effects of communicative situations could be language-specific in the domain of null subjects too.

The chapter is structured as follows: Section 2 provides the theoretical background on the different types of null subject languages this chapter deals with. In this section, we also summarize results of previous research that discusses animacy, mode and formality as possible factors influencing when subjects are omitted. Section 3 introduces the research questions and the corresponding predictions for every language in our study derived from results in the literature. Section 4 describes our methodology, including the design, the participant sample, the corpus annotation scheme and the statistical analysis. Section 5 provides the results reported by the models and grouped according to the different factors accounted for in the analysis, namely typology (i.e., different pro-drop types), animacy, mode and formality. Finally, Section 6 presents a discussion of our results.

## **2 Null subjects**

Languages have been grouped into different types regarding their ability to omit subjects. Leaving aside languages with so-called radical pro-drop or expletive subjects, most authors distinguish between languages readily allowing pro-drop (strict pro-drop) and non-pro-drop languages which by and large disallow the omission of subjects. The former are claimed to require rich verbal morphology (Perlmutter 1971, Holmberg 2005, Roberts & Holmberg 2010). A third class of languages are referred to as partial pro-drop; subject omission in these languages is more limited than in strict pro-drop languages, and may be restricted by (morpho)syntactic, information-structural, or lexical constraints (Holmberg 2005, Roberts & Holmberg 2010, Frascarelli 2018, Madariaga 2018). Lastly, a fourth class of discourse-prominent languages allow for the omission of subjects when they are topics; this process, however, follows different constraints in different languages (Öztürk 2001, 2008, Schäfer 2021). Greek, Russian, Turkish, English and German are classified as different types of null subject languages, as shown in Table 1.<sup>1</sup> In the following subsections, we will discuss the languages in more detail.

### **2.1 Null subjects in English**

English is a language that does not allow null subjects. This means that subjects, both expletive and referential, are obligatorily overtly realized. Spoken English as well as diary English have been argued to contain null subjects, but it is a matter of controversy whether these registers involve truncated structures or are

---

<sup>1</sup>For a different view on language classification according to the pro-drop feature see Dryer (2013).

Table 1: Pro-drop across languages

| Language | Pro-drop type                      |
|----------|------------------------------------|
| Greek    | consistent/strict pro-drop         |
| Russian  | partial pro-drop (or non-pro-drop) |
| Turkish  | topic pro-drop                     |
| English  | non-pro-drop                       |
| German   | topic drop                         |

instances of topic drop; see Haegeman & Ihsane (2001) for discussion and references. A noticeable higher use of null subjects in informal conversational English has also been noted, particularly in the first person and in negated constructions such as *don't know* and *can't say* (Harvie 1998, Nagy et al. 2011). Additionally, English discourse fixed expressions in the first and second person singular, such as *gotcha!* and *wanna go?* permit null subjects (Haegeman 2010).

Research on heritage languages with strict pro-drop properties that come into contact with English has shown that their speakers tend to have more realized subjects compared to their monolingual counterparts, indicating transfer effects on subject realization. For instance, Paradis & Navarro (2003) report that bilingual Spanish-English children are more likely to use overt subjects in heritage Spanish than monolingual Spanish-speaking children. Austin et al. (2017) observe the reverse transfer effects in sequential bilingual Spanish-English children who are more proficient in their heritage language Spanish than in their majority language English. Second-generation heritage speakers of Polish in Toronto also tend to use more realized subjects in Polish compared to first-generation and monolingual speakers of Polish in subject continuity contexts (Chociej 2011). A study on heritage Welsh by Boon (2014) found that heritage speakers omit subject pronouns only with the copula verb 'to be' more often than the baseline, but not with other verbs, which she attributes to "incomplete acquisition" of the pro-drop parameter.

## 2.2 Null subjects in German

German is classified as a topic drop language in certain registers and under certain grammatical constraints (Fries 1988, Trutkowski 2016, Schäfer 2021). Subject drop in German is possible only in the left periphery (pre-field) and is claimed to show restrictions concerning agreement, with first person being especially prone

to subject drop resulting in V1 sentences (Schmitz et al. 2012, Trutkowski 2016, Schäfer 2021). Example 2 shows a sentence with prototypical German topic drop. The subject pronoun *ich* ‘I’ can be omitted because it is discourse-prevalent and the finite auxiliary verb marks number and person even though in this example first and third person singular would both be permitted. The reading for the first person singular is the default interpretation without further context, but in a context with a given third person singular topic or referent, the interpretation of the empty position would shift to that.

Based on a corpus study and two acceptability rating studies, Schäfer (2021) shows that topic drop in German is influenced by a mixture of factors such as the predictability of the omitted constituent in context, verb probability and distinct verbal inflection.

- (2) *pro kann die Verabredung leider nicht einhalten*  
*e can the appointment unfortunately not keep*  
‘Unfortunately, (I) cannot keep the appointment.’ (Schäfer 2021: 116)

Several studies have investigated subject realization in bilingual populations with a null subject heritage language and with German as a majority language (Schmitz et al. 2012, Flores & Rinke 2020, Brehmer et al. 2023). In contrast to what has been found for English as a majority language, these studies suggest that heritage speakers in contact with German do not show transfer effects in their majority language. Brehmer et al. (2023) found in their study that German-Polish bilingual children and their Polish monolingual peers behaved similarly in their narratives and dialogues regarding the realization of subject pronouns.

### 2.3 Null subjects in Greek

Greek is a consistent pro-drop language showing rich verbal morphology, which means that null subjects can be used in all grammatical feature combinations and syntactic environments (Philippaki-Warburton 1987, Alexiadou & Anagnostopoulou 1998). In the generative tradition, this has been associated with overt movement of the verb to the T(ense) category licensing null subjects (Alexiadou & Anagnostopoulou 1998). In Greek, the default strategy is to use null subjects, except in certain situations such as topicalization, focusing, and contrastive contexts where overt subject realization is required (Tsimpili et al. 2004).

Research on null subjects in Greek bilinguals has focused on both situations where both languages allow null subjects as well as situations where the additional language does not allow null subjects. The findings show that bilingual

adults generally exhibit similar results to monolingual groups. For example, a study conducted by Giannakou & Sitaridou (2022) on adult Greek heritage speakers in Argentina found that these participants produced overt and null subjects in Greek similarly to the monolingual group of Greek speakers. There was also no overuse of overt subjects. The only significant difference was that heritage speakers used null subjects more often in topic shift contexts. Daskalaki et al. (2019) focused on the realization of subjects in heritage Greek children residing in Canada and the US who spoke English as their majority language. The results showed patterns of subject realization to be comparable to monolingual Greek-speaking children. According to Andreou (2020), English-Greek, German-Greek and Greek-Albanian bilingual children exhibited overspecification patterns by using full noun phrases when null subjects were the more appropriate choice. These patterns were modulated either by low language experience in the case of English-Greek and German-Greek speakers with Greek as their heritage language or by language dominance in Greek in the case of Greek-Albanian children with Greek as their majority language. Turning to adult Greek-Italian heritage speakers, Di Domenico & Baroncini (2019) claim that the null subject in Greek is the most employed anaphoric device across speaker groups regardless of Greek being the heritage or the majority language. In contrast, L2 Greek speakers with Spanish as their L1 show diverging patterns in subject realization in Greek depending on the manipulated context (contrastive vs. non-contrastive) and their proficiency in Greek, as reported by Margaza & Gavarró (2020). Finally Torregrossa et al. (2021) tested whether Greek and Italian monolingually-raised speakers differ in the use and interpretation of null and overt subject pronouns known as anaphora resolution and found that Greek null subjects are more flexible referring to subject and object antecedent than Italian null subjects. In sum, heritage speakers' performance is similar to monolinguals, but with some inconsistencies.

## **2.4 Null subjects in Russian**

Since Russian exhibits frequent realization of subjects, but may drop them, it is sometimes classified as a partial pro-drop language or as a non-pro-drop language with frequent subject ellipsis (Fehrman & Junghanns 2008, Shushurin 2018, Budennaya 2019, 2020). Realization and omission of subject pronouns are determined by various factors, especially information structural categories, the distance between an antecedent and the null subject, and lexical choices such as the type of connector in subordination contexts (Fougeron & Breillard 2004, Bizzarri 2015, Madariaga 2018, 2022, Pekelis 2018, Budennaya 2020). These factors

often interact with each other, which makes it difficult to determine the weight of each factor, especially since pro-drop is optional in many cases.

For instance, it has been claimed that being part of an A(boutness)-Topic chain facilitates omission of the subject with additional requirements concerning the distance between the null subject and the antecedent (Madariaga 2018, 2022). Madariaga (2022) claims that the subject in the second sentence in (3) may be omitted, since it continues an A-Topic chain, whereas the subject in the third sentence does not, thus has to be overtly realized:

- (3) My<sub>i</sub> idäm na ozero. pro<sub>i</sub> Nadeemsja tam vstretit' Ivanaj. \*(On)<sub>j</sub>  
we.NOM go.1.PL to lake (we) hope.1.PL there see.INF Ivan.ACC he.NOM  
nam obeščal peredat' ključi.  
we.DAT promised.M pass.INF keys.ACC  
'We are going to the lake. We hope to see Ivan there. He has promised us  
to pass the keys.'

For subordinate contexts, Pekelis (2018) shows that certain connectors like *kak* 'how', *čto* 'that', *potomu čto* 'because', and *esli* 'if', sometimes with and without certain correlates in the main clause, may facilitate or prohibit null subjects. For example, the conditional complementizer *esli* 'if' without a correlate *to* 'then' disallows null subjects as shown in (4) (Shushurin 2018: 155).

- (4) [Esli on<sub>i</sub> dejstvitel'no idet po toj ulice] \*(to) pro<sub>i</sub> skoro uvidit  
if he really walks along that street then (he) soon will.see  
stanciju.  
station  
'If he is really walking that street then he'll see the station soon.'

Research on heritage Russian speakers yields inconsistent findings concerning null vs. overt pronominal subjects. Some studies have shown that overt pronominal subjects are more common in heritage Russian varieties in the US than in monolingual ones, which is claimed to indicate transfer from English (Isurin 2011, Dubinina & Polinsky 2013, Ivanova-Sullivan 2014). In contrast, Nagy et al. (2011), examining first vs. second/third generation Russian heritage speakers in Toronto, found no generational differences in the frequency of null subjects, leading the authors to conclude that English had not caused any changes in this heritage Russian population.

## 2.5 Null subjects in Turkish

Turkish is characterized as a discourse prominent language allowing topic pro-drop, meaning that subjects are not always necessary in certain highly specified contexts as in (5), or when there is continuity of the topic (Öztürk 2001, 2008). In these cases, the use of overt pronouns is considered to be pragmatically marked (Enç 1986). The most widely accepted approaches to analyzing null subject in Turkish have followed the formal analysis that null subjects can be identified through the verbal agreement morphology, which specifies person and number (Enç 1986). However, Öztürk (2001) argues that pro-drop in Turkish is guided entirely by pragmatic constraints and, as such, Turkish should be viewed as a topic drop language, rather than a typical pro-drop language. For the purposes of the current study, we do not take a stance on the specific formal analysis of null subjects in Turkish, but it is important to note that the omission of subjects in Turkish is common. We propose the term “topic pro-drop” to unify existing approaches and signify that Turkish is not a typical pro-drop language like Greek or Italian.

- (5) (Ben) su iç-iyor-um.  
*pro (I) water drink-PROG-1SG*  
 ‘I am drinking water.’

Studies on Turkish heritage speakers have produced inconclusive results with regard to subject realization in heritage Turkish varieties. Some studies, such as those conducted by Haznedar (2010) and Koban Koç (2016) on heritage Turkish in contact with majority English, found that pronominal subjects are more frequently realized in heritage varieties compared to the monolingual standard. However, other studies, such as those by Azar et al. (2020) and Dikilitaş (2017), found contrasting results. Azar et al. (2020) observed a slightly higher use of overt pronouns in the repertoire of heritage speakers in contact with Dutch, but there were no significant differences between the groups. They argued that the groups align in their frequency of realization of overt subjects and referents overall. Similarly, a longitudinal case study by Dikilitaş (2017) of a Turkish-English bilingual child revealed no evidence of cross-linguistic influence from majority English, as the use of null and overt subjects did not deviate from those of monolingual Turkish peers. These divergent findings may be attributed to several factors, including sociolinguistic background variables that guide the rate of overt subject use, differences in the communities in which heritage speakers acquire the language, and differences in methodologies, including statistical tests and sample sizes.

### **3 Research questions and hypotheses**

Our review of null subjects in the three languages under investigation demonstrates that there are clear differences in the type of pro-drop languages at hand. While Greek requires null subjects in most contexts which are licensed by agreement features, this is more limited in Russian and Turkish. We also outlined that the two majority contact languages in this study, German and English, are both non-pro-drop languages which only allow null subjects in very restricted contexts and to different degrees. The participants in this study exhibit interesting bilingual language combinations. These combinations, along with the unique cross-linguistic comparisons we can draw from this set of languages, prompt several lines of research. These research lines also capture grammatical and extra-linguistic factors, such as animacy and different communicative situations, which we have introduced. Drawing from the current body of research, we formulate the subsequent research questions and corresponding hypotheses:

RQ1: Is the way in which subjects are expressed in heritage Greek, Russian, and Turkish similar to the way in which monolingual speakers express subjects?

H1: We anticipate that heritage speakers will show a higher use of overtly realized subjects.

There are several plausible explanations for a higher use of overt subjects by heritage speakers. One reason might be the effect of cross-linguistic influence from the dominant languages German and English. Bilinguals experience interaction between their language systems, according to the Non-Autonomous Version of the Separate Development Hypothesis (as described by Hulk & Müller 2000, Müller & Hulk 2001, Tracy 2000, 2014). Throughout a speaker's life, transfer from the dominant language may be used to fill in the gaps caused by reduced input in the heritage language. Heritage speakers may adopt subject realization strategies from their dominant language for their heritage language (Gawlitzek-Maiwald & Tracy 1996, Döpke 2000). At the same time, research has shown that heritage speakers tend to be more explicit, which could be another factor that leads them to use overt subjects more frequently than monolinguals (Pashkova et al. 2020).

RQ2: Are there any similarities or differences in the patterns of subject omission observed among various groups of heritage speakers and, if so, what factors may account for these patterns?

H2.1: It can be assumed that these different pro-drop types allow for varying levels of language-internal change or cross-linguistic influence from the majority languages. Therefore, we might anticipate higher use of overtly realized subjects from heritage Greek to heritage Russian to heritage Turkish. However, the design only includes two majority languages that are both non-pro-drop languages. This makes it unsuitable to test for a true effect of cross-linguistic influence as this would require a comparison with a pro-drop majority language. So this sub-hypothesis is only falsifiable outside of the framework of this study.

Apart from cross-linguistic influence, the typological differences might also indicate how stable or dynamic the pro-drop system of a language is internally, regardless of the majority language influence. A strict pro-drop system as in Greek might lead to a more stable use of null subjects in the heritage grammar too. Following the same logic, pro-drop systems that are guided more by pragmatic and situational factors such as those of Russian and Turkish might be open to more dynamicity in the use of null subjects in the heritage language.

H2.2: Alternatively, in the context of heritage languages, we could argue that there may be converging developments regardless of the pro-drop type of the languages. Since the (non)realization of pronominal subjects also involves the interface with external linguistic domains (information structure and discourse organization), we may expect uniform transfer effects from majority languages following the Interface Hypothesis (Sorace 2011). We expect to find more overtly realized subjects in heritage varieties for several reasons. First, there is great variability in linguistic phenomena at the syntax-discourse interface (Tsimpli et al. 2004, Sorace 2011). Additionally, as the morphological repertoire weakens in heritage languages (Polinsky 2018) and this might affect agreement marking which licenses pro-drop, we expect a higher use of overtly realized subjects across all heritage varieties of Greek, Russian and Turkish. Furthermore, some heritage speakers with limited language exposure may not be able to adequately mark verb agreement, leading them to avoid null subjects and use more overt subjects instead (Isurin 2011, Daskalaki et al. 2019). Therefore, individual variance within each group needs to be considered to avoid false positives resulting from the analysis of aggregated frequencies (Baayen et al. 2008, Winter 2011, Brezina & Meyerhoff 2014, Winter 2019).

RQ3: Do Greek, Russian and Turkish heritage speakers' expressions of null subjects in majority German align with monolingual German speakers' productions?

H3: German belongs to the non-pro-drop languages and typologically differs from the typology of the three heritage varieties, namely Greek, Russian, and Turkish, and thus we expect cross-linguistic influence with extended use of null subjects in majority German productions. Additionally, the heritage speakers might also use the reverse strategy and produce more standard-like and formal expressions of subject use in German to display that they are proficient users of their majority language German. This would lead to less topic drop and more overt subjects.

RQ4: Do intra-linguistic factors, such as animacy, or extra-linguistic factors, such as formality (formal vs. informal) and mode (written vs. spoken), have an impact on subject realization in both monolingual and heritage varieties?

H4: Based on the literature discussed earlier, we expect null subjects to be preferred for animate subjects. However, in light of a lack of prior research on the impact of the external factors formality and mode on null subjects for Greek and Turkish and conflicting results for Russian, we approach these factors in an exploratory way.

## 4 Methodology

Following the *Language Situations* approach (Wiese 2020) described in Section 4.2 below, we collected ecologically-valid semi-spontaneous data through narration tasks to identify emerging trends in the language systems. The sections below provide detailed information on the composition of speakers in the study, corpus annotation and queries as well as on the statistical analysis applied.

### 4.1 Participants

Table 2 presents the overall information about the number of participants grouped by country of elicitation and the number of tokens in the different sub-corpora. Tokens are used in the way they are defined by the TreeTagger tokenization script (Schmid 1997). All participants were recruited from urban areas to minimize the role of dialect. Urban areas facilitate regional dialect levelling

and were therefore preferred (Britain et al. 2010). Monolingual<sup>2</sup> participants were recruited from Athens, Greece; St. Petersburg, Russia; and İzmir and Eskişehir, Turkey. Bilingual participants were recruited from the United States and Germany. Heritage speakers in the US lived in the greater Washington DC area (including Virginia and Maryland), Chicago, and the greater New York City area (including New Jersey). Heritage speakers in Germany were recruited from the Berlin and Brandenburg urban area.

Table 2: Participant and corpus metadata based on the corpus version 1.0

| Country of elicitation | Group            | N  | Tokens |
|------------------------|------------------|----|--------|
| Greece                 | monolinguals     | 64 | 27,931 |
| Russia                 | monolinguals     | 66 | 25,930 |
| Turkey                 | monolinguals     | 64 | 20,947 |
| Germany                | heritage Greek   | 48 | 19,782 |
|                        | heritage Russian | 61 | 32,882 |
|                        | heritage Turkish | 64 | 23,722 |
| USA                    | heritage Greek   | 63 | 18,302 |
|                        | heritage Russian | 60 | 29,214 |
|                        | heritage Turkish | 58 | 18,502 |

Candidates were invited to take part in the study if they were a) either mono- or bilingual speakers, b) born and raised in the country of the majority language or, in case of bilingual speakers, moved there before the age of 48 months,<sup>3</sup> c) bilinguals should use their heritage language with family and friends regularly, d) not diagnosed with hearing or speech disorders, e) exposed to the majority language (English or German) before the age of five.

In addition, for the written production in Greek and Russian, participants were given the option to use Latin script if they were not able to write in Greek or Cyrillic scripts. Prior to the study, participants were informed about their rights, data protection, and the procedure, and asked to sign a consent form in the

<sup>2</sup>We use monolingual as a shorthand to refer to monolingually-raised participants who have not acquired another language other than their first language in childhood (and later potentially learned other languages in school settings) and, crucially, do not use another language than their first language in everyday life.

<sup>3</sup>In exceptional cases participants who moved to the hosting country before the age of six years were admitted to the study.

majority language of the country where the elicitation took place. In cases where the participant was a minor, one of their parents or legal guardians was asked to sign the consent form.

Several considerations were taken to ensure demographic comparability between the participant groups. Two age groups of participants were included: adolescents (aged 14–18) and adults (aged 22–35). Adolescent participants were either currently attending school or had recently graduated. The two different age groups are not relevant for the present study as no effects of age on null subjects were expected. An additional criterion excluded bilingual candidates with extensive formal education in the heritage language, such as those who attended bilingual primary or secondary schools.

From a sociolinguistic and demographic standpoint, Greek, Russian, and Turkish migrant communities in the US and Germany exhibit similar characteristics in their language-related behaviors and experiences within the host countries. These communities often establish tight-knit groups, acting as crucial centers for cultural and linguistic preservation, actively contributing to language maintenance initiatives. Nevertheless, despite endeavors to preserve and pass on linguistic heritage to the younger generations, the heritage language of second, third and fourth generation migrants is observed to undergo substantial transformations (Benmamoun et al. 2013, Polinsky 2018).

## 4.2 RUEG corpus

The approach employed in this research is a modification of the setup introduced by Wiese (2020), known as the *Language situations* paradigm. This paradigm facilitates the generation of semi-spontaneous data, offering comparable naturalistic information in both oral and written forms, as well as in formal and informal situations. In the elicitation process, participants were presented with a short video depicting a fictional minor car accident, and their task was to narrate the incident as if they had witnessed it, addressing either a close friend or a police officer. Consequently, participants engaged in four distinct communication scenarios during a single session. Heritage speakers took part in two sessions separated by at least three days: one in the majority language and one in their heritage language. Monolingually-raised participants only participated in a single session conducted in the majority language of the respective country. To examine the impact of formality and mode on narrations, we simulated formal-spoken, formal-written, informal spoken, and informal-written settings. The formal part of the elicitation occurred in an office setting, where the elicitor and participant faced each other. The elicitor wore formal attire, and used standardized language

and honorifics. Spoken narration involved leaving a voice message on the police department's answering machine, while the written task required typing a witness report on a "police laptop". For the informal part, another elicitor, stylized as talkative and casually dressed, engaged participants in casual conversation before instructing them to narrate the video's contents via a voice message on WhatsApp to a close friend. The written task involved sending a text message about the accident on WhatsApp to the same friend. The entire session was audiorecorded and the data were pseudonymized. Elicitation orders for the communicative situations were balanced. Once transcribed and annotated, the data were released as the RUEG Corpus (Lüdeling et al. 2024), developed within the Research Unit *Emerging Grammars in Language Contact Situations* (RUEG) funded by the German Research Foundation (<https://hu.berlin/rueg>). The multilevel annotated RUEG corpus is accessible through the ANNIS interface (Krause & Zeldes 2014), comprising audio for spoken data and visualization options across six sub-corpora for English, German, Greek, Russian, Turkish, and Kurmanji. Additional sub-corpora feature special annotations such as aspect and tense. The present study utilized data from the RUEG corpus (Lüdeling et al. 2024) including narrations from a total of 548 speakers. This encompassed both monolingually-raised and heritage speakers of Greek, Russian, and Turkish.

### 4.3 Corpus annotation

The RUEG subcorpora for Greek, Russian and Turkish were manually annotated in EXMARaLDA (Schmidt & Wörner 2014) with respect to subject realizations on the three following levels:

- *denotation and animacy of the referents* derived from the context (the full list can be found at <https://osf.io/25tw6>)
- *syntactic realization* (overt vs. null subjects)
- *expectedness* of subject (i.e., whether an overt subject is expected in the given context using the tags *yes*, *no*, *not sure*)<sup>4</sup>

---

<sup>4</sup>The judgments regarding the *expectedness* of subjects were done by at least two native or near-native speakers of each language and cases where the annotators did not agree were further discussed.

### 4.3.1 Animacy of referents

Subjects were identified in relation to a finite verb corresponding to the broad syntactic notion of subjecthood (McCloskey 1997). The category of subject includes noun phrases such as lexical nouns, pronominal expressions, and demonstratives.

Regarding the denotation of the subjects, we considered human referents like *man*, *woman*, *driver*, and animals like *dog* and personal pronouns referring to them as animate subjects while physical objects like *ball*, *cars*, *groceries* were considered as inanimate subjects. Besides the individual referents, we found some conjoint referents in our data as well. We classified conjoints referents that included at least one animate referent as animate (Adamson & Anagnostopoulou 2024) as in the example below:

- (6) para s koljaskoj  
couple with stroller  
'A/the couple with a/the stroller.'

### 4.3.2 Overtness of subjects

The following explains how the layers “syntactic realization” and “expectedness” were operationalized to assess the expression of subject reference in this study. The concept of expectedness was introduced in order to make the present analysis cross-linguistically comparable. Expectedness captures the factors that license and facilitate null subjects which are highly language-dependent as presented in Section 2. By creating the concept of expectedness, we turn the variety of grammatical factors into a binary choice (Yes, an overt subject was expected according to the grammatical rules of the language; or No, an overt subject was not expected according to those extensive criteria). Some details from the guidelines are listed below. The elaborate annotation guidelines that explain the additional annotation layers for this study in depth are available in a PDF-file at <https://osf.io/pvmx5>.

Our working concept for expectedness differs from the widely used notion of accessibility (Chafe 1987, Gundel et al. 1993, Ariel 2001, Arnold 2010). As Allen et al. (2015) point out, accessibility, generally, is seen as a cognitive proxy to describe how likely a speaker is to produce something on the spectrum between a full form like a noun phrase and a null form like subject drop. Accessibility is one of the factors that we take into account when we code for expectedness which takes into account many additional factors such as topicality and language-specific

grammatical factors. Topicality adds to a three-way interaction with accessibility and expectedness. Additionally, as our descriptions of pro-drop in Greek, Russian and Turkish illustrate, there seem to be stricter grammatical constraints between languages that determine whether a subject may be omitted or overt. This is true regardless of accessibility as Turkish speakers can, for example, realize subjects even when they are accessible, but this is not possible in Greek.

#### 4.3.3 Language specific aspects and exclusions

In all languages of investigation, we excluded dative subjects, subjects in imperatives, rhetorical questions, interrogative as well as exclamative sentences, addressings, and impersonal constructions from the annotation.

For Greek, fixed expressions like *ksero go* ‘I don’t know’ were also excluded as the subject is always expected post-verbally and this could influence the results for overt subjects. 21 such cases were observed in the Greek data.

For the annotation of null subjects in Russian, various factors mentioned in Section 2 were taken into account. Particularly, *communicative situation, discourse, distance between antecedent and subject, lexical markers, information structural categories* were considered. For instance, in an informal-spoken communicative situation, as a tendency, we would expect less overt subject realization in general. More specific situations like contrast or topic change were marked as expecting subjects. Besides, subjects within discourse markers like (*ty*) *ponimaeš* ‘(you) understand’ were annotated as not expected ones.

Since Turkish is a topic pro-drop language, as specified earlier, the available number and person marking on the finite verb allow the subject to be omitted. In many instances, speakers also accept realized subjects in non-requiring contexts and positions as grammatical even though these utterances might be perceived as slightly redundant when speakers judge them. However, our data show that these kinds of overt subject realizations are common in different mono- and bilingual speaker’s repertoires. Additionally, Turkish requires subjects to be overtly realized to mark a new or later reoccurring topic, including contrast contexts which also can require overt subjects. Example (7) demonstrates a contrastive sentence where the subjects of the main and the subordinate clause are both overtly realized:

- (7) ben evi temizlerken sen boş bos film izliyorsun.  
 I house clean you empty empty movie watch  
 ‘I’m cleaning the house and you’re idly watching a movie.’

One exception in Turkish, however, is the first person pronoun *ben* in narratives. When a finite verb is marked for 1SG, it is unambiguously clear that it refers to the narrator. In this case, it is often not necessary to overtly realize the pronoun, except when it is in a contrastive or emphatic use.

Additionally, 232 data points that were tagged as syntactically realized or not realized had to be excluded because their *expectedness* status was not clear. This means that the formal grammars of the language do not allow a clear prediction as to whether a subject would be expected or not in a given context. Compared to the 31,539 data points that we accounted for, the number of excluded cases is marginal, i.e., <1%. The exact number of data exclusion based on this case is as follows:

Table 3: Excluded data points grouped by language and syntactic condition.

| Language | N overt subjects | N null subjects |
|----------|------------------|-----------------|
| Greek    | 3                | 1               |
| Russian  | 115              | 51              |
| Turkish  | 32               | 30              |

#### 4.4 Statistical Analysis

To draw cross-linguistic (Greek, Russian and Turkish heritage speaker groups) and also across-variation (monolingual vs. heritage speakers) comparisons, we ran mixed-effects models in R using the lme4 package (Bates et al. 2015). As independent variables, we included *language* (Greek, Russian, Turkish), *country* (Germany, Greece, Russia, Turkey, US), *expectedness*, *animacy*, *formality* (formal vs. informal), and *mode* (spoken vs. written). Further, to capture the individual speaker variability, we specified random effects by participant. In R notation, the model for each language looked like follows:

```
glmer(realization ~ country + language + expectedness + mode +
      formality + animacy + (1|participant), data = ProdropModel,
      family = binomial, control = glmerControl(calc.derivs=FALSE))
```

In words, this code uses the lme4 package function glmer to create a generalized linear mixed model of the binomial distribution family. The formula models Subject realization with its levels Null and Overt as a function of language group

(heritage USA, heritage Germany, monolingual majority homeland), expectedness according to our annotation scheme (levels yes and no), mode with its levels spoken and written, formality with its levels formal and informal, and animacy with its levels animate and inanimate.

The normalization of data points is conducted according to the number of finite verbs. Table 4 reports the number of finite verb tokens per group.

Table 4: Number of finite verbs per group.

| Country of elicitation | Group            | N finite verbs |
|------------------------|------------------|----------------|
| Greece                 | monolinguals     | 4,954          |
| Russia                 | monolinguals     | 3,965          |
| Turkey                 | monolinguals     | 4,609          |
| Germany                | heritage Greek   | 3,494          |
|                        | heritage Russian | 4,624          |
|                        | heritage Turkish | 4,986          |
| USA                    | heritage Greek   | 3,471          |
|                        | heritage Russian | 4,342          |
|                        | heritage Turkish | 4,257          |

## 5 Results

This section provides the results of the statistical analysis on subject realization and subject drop in different conditions in Greek, Russian, and Turkish. To describe the results from different perspectives, we organize this section in the following subsections: descriptive results (Section 5.1) and intra- and extralinguistic factors (Section 5.2) including formality, mode and animacy. Before we address those perspectives one by one, we present general results and model outcomes in the following.

First, we show how the heritage and monolinguals speakers of Greek, Russian and Turkish use subjects along the two variables which both have two levels, namely *realization* (null vs. overt) and *expectedness* (YES = expected, NO = unexpected).

Figure 1 illustrates all three languages and all participant groups for a better comparison across the same scale. The figure represents a faceted plot with a grid for each of the nine speaker groups sorted by different countries of elicitation

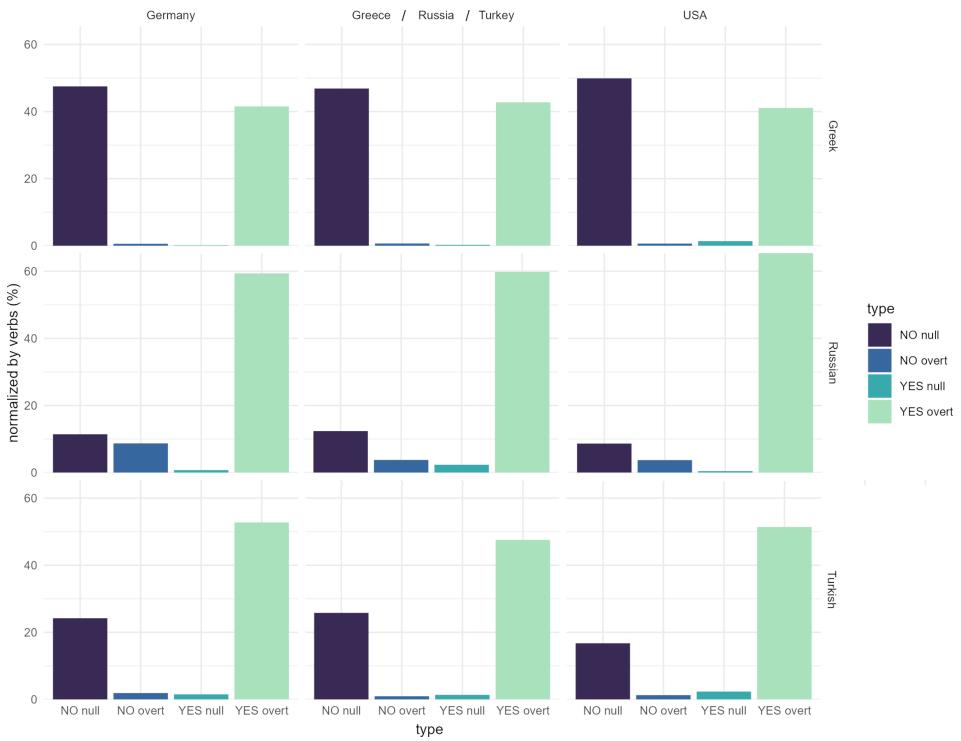


Figure 1: Normalized number of occurrences for four different combinations/categorizations.

from left to right (Germany, respective countries of the monolingual varieties, US) and three languages from top to bottom (Greek, Russian, Turkish). Each row shows one language, and each column reflects one country. For instance, the first grid of the first column represents heritage speakers of Greek in Germany, the second grid in the first column represents heritage speakers of Russian in Germany and the third row in the first column represents heritage speakers of Turkish in Germany. The second column represents monolingual speakers in Greece, Russia and Turkey, respectively. The third column shows heritage speakers of Greek, Russian and Turkish residing in the US, respectively.

Each subplot shows four conditions that are the result of combination of *realization* (null vs. overt) and *expectedness* (YES = expected, NO = unexpected) variables. We list and illustrate each of them below. From left to right, each bar represents the following conditions:

- *NO null*: overt subject is unexpected and it was null
- *NO overt*: overt subject is unexpected, but it was realized
- *YES null*: overt subject is expected, but it was null
- *YES overt*: overt subject is expected and it was realized

The tag “NO overt” applies when an overt subject was observed even though a null subject was expected. In pro-drop languages, this would typically be the case when a referent was just introduced and is just maintained in the discourse. An overt subject in this position would be non-canonical given that it is a pro-drop language. This would be the case if the sentence in Example (8) is a continuation of (10).

- (8) Köpek topu yakaliyor.  
 dog ball catches  
 ‘The dog catches the ball.’

The combination “NO null” was annotated when a canonical case of null subject use was observed. In the given context of the example, this would be the case if the dog is maintained as a referent. We illustrate this in the canonical continuation in (9) as a contrast to non-canonical continuation in (8).

- (9) Topu yakaliyor.  
 ball catches  
 ‘(The dog) catches the ball.’

The combination “YES overt” is used to tag canonically overtly expressed subjects in a language. In pro-drop languages, for example, subjects are often expressed overtly when they are introduced for the first time in discourse such as in Example (10). In this case, the dog is introduced in the narrative. We will use this context for all tag combinations.

- (10) Ondan sonra bir köpek topa doğru koşuyor.  
 after that a dog ball towards runs  
 ‘Then, a dog runs towards the ball.’

Finally, there is the less frequent combination “YES null” where a subject is null even though an overt form was expected. Given our example context, this would be the case if a new referent is introduced in the story but it is not overtly realized as a subject as in Example (11).

- (11) Köpeğe kızıyor.  
dog rant  
'(S)he talks angrily to the dog.'

The *NO null* and *YES overt* conditions represent the canonical way of subject realization, whereas *NO overt* and *YES null* conditions represent the non-canonical subject realization.

The first view on Figure 1 provides the impression that on the one hand, Greek, Russian and Turkish behave differently from each other, and on the other hand, heritage and monolingual speakers of these languages seem to behave similarly to each other.

To address these patterns in a more interpretable way, we present the results for the generalized linear mixed-effects regression models for each language. We fitted three logistic mixed-effects models to predict syntactic realization of subjects (overt vs. null) with *Country*, *Expectedness*, *Mode*, *Formality* and *Animacy* as fixed effects. The model included *Participant* as random effect.

## 5.1 Descriptive results grouped by language

This subsection presents the descriptive results from heritage Greek, Russian, Turkish and majority German.

### 5.1.1 Null subjects in Greek, Russian and Turkish

We first asked whether null subjects in the three languages work in the same way. Starting the exploration of our results based on the typological differences on pro-drop in the three languages under investigation we confirm the claim that Greek clearly belongs to strict pro-drop languages while Russian and Turkish do not. As seen in Figure 1, Greek preserves pro-drop in all varieties explored in this study while Russian and Turkish align in the sense that overt subjects are more prominent in all varieties. Russian speakers across all groups realize subjects to a similar extent in the condition where the subject realization is expected. Specifically, the amount of overt subjects is higher than the amount of null subjects. The same observation holds for Turkish speakers.

### 5.1.2 Topic drop in majority German

We then asked whether each group of heritage speakers produced topic drop in the same way in their majority language, German. An overview of the results from the German data is shown in Figure 2. The black dots that represent the

means almost perfectly mirror each other in all subplots. This points to no differences between the groups with regard to subject realization strategies in the majority language German. We confirmed in a Bayesian model that we report extensively in Özsoy & Burbelko (in preparation) that there are no meaningful differences between the groups. The only strong and important effect of topic drop that we find is based on mode. Specifically, we find that utterances in spoken language are much more likely to include null subjects than utterances in written language.

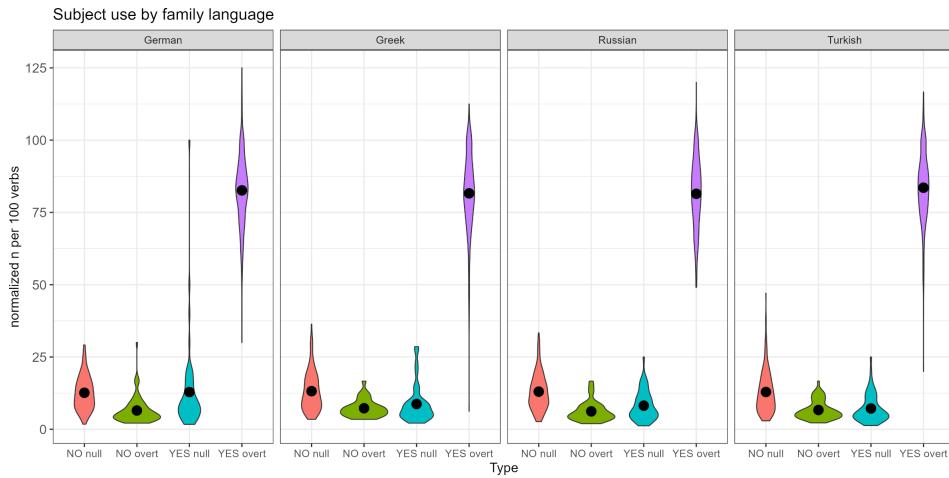


Figure 2: Normalized number of occurrences for four different combinations/categorizations in German data.

## 5.2 Intra- and extralinguistic factors

In order to verify which factors affect the realization of subjects we conducted a regression analysis and the results are shown below.

### 5.2.1 Country

One of the main aims of this study is the typological comparison between Greek, Russian, and Turkish as heritage languages (and partly German as a majority language). Since in our controlled design and study the participant groups live in similar sociodemographic and linguistic environments, this direct typological comparison can be drawn. For each language, there are three levels – Germany and the USA as the host countries of the heritage language, and the respective

homeland countries Greece, Russia and Turkey. We can report the effects of each of these levels per language because we sum-coded these variables which allows us to compare all groups equally.

For Greek, we observe that the effect of Country is statistically significant at the USA level (mean = 0.616, SD = 0.205,  $p < 0.01$ ) and non-significant at the Germany (mean = 0.348, SD = 0.213,  $p > 0.05$ ) and Greece (mean = 0.268, SD = 0.194,  $p > 0.05$ ) levels. The directionality of the effect is similar in all groups. This indicates that heritage speakers of Greek in the USA show a higher use of overt subjects compared to the groups in the other countries.

For Russian, the effect of Country is significant at the Germany (mean = 0.472, SD = 0.099,  $p < 0.01$ ) and Russia levels (mean = -0.599, SD = 0.101,  $p < 0.01$ ), and non-significant at the USA level (mean = 0.127, SD = 0.101,  $p > 0.05$ ). The directionality of the effect is positive for the heritage groups and negative for the Russia group. This indicates that heritage speakers of Russian are more likely to produce overt subjects compared to monolingual speakers who show a reverse effect meaning that they drop subjects more frequently.

For Turkish, the model estimates a significant effect at the Germany (mean = 0.328, SD = 0.115,  $p < 0.001$ ) level and non-significant effects at the Turkey (mean = -0.154, SD = 0.114,  $p > 0.05$ ) and USA (mean = -0.174, SD = 0.119,  $p > 0.05$ ) levels. The directionality of the effect is positive for the Germany level and reversed for the Turkey and USA levels. This means that while heritage speakers of Turkish in Germany are likely to produce more overt subjects, this effect is not present for the other groups.

### **5.2.2 Formality and mode**

To account for a broader repertoire of naturalistic language use, our study tests the effects of Mode and Formality which we utilize to capture register variation. For Greek, the effects are non-significant and non-meaningful (Mode: mean = -0.048, SD = 0.098,  $p > 0.05$ ; Formality: mean = 0.118, SD = 0.097,  $p > 0.05$ ). For Russian (Mode: mean = 0.376, SD = 0.048,  $p < 0.001$ ; Formality: mean = 0.202, SD = 0.045,  $p < 0.001$ ) and Turkish (Mode: mean = -0.114, SD = 0.054,  $p < 0.001$ ; Formality: mean = 0.186, SD = 0.054,  $p < 0.001$ ), the effects for Mode and Formality are all significant, and they partly overlap in their directionality. The effect for Formality indicates that Russian and Turkish speakers are more likely to overtly realize subjects in formal contexts compared to informal ones. The Mode effect goes in different directions for the two languages. Russian speakers are more likely to overtly realize subjects in spoken mode and drop more in

written mode. On the contrary, Turkish speakers drop subjects more in spoken mode and overtly realize the subject more when writing.

### **5.2.3 Animacy**

In addition to situational and discourse-dependent factors such as Mode and Formality, our study tests the grammatical factor of Animacy for the realization of subjects. Animacy is statistically significant for all the languages under investigation (Greek: mean = 0.398, SD = 0.201,  $p < 0.01$ ; Russian: mean = 0.361, SD = 0.099,  $p < 0.01$ ; Turkish: mean = 0.254, SD = 0.116,  $p < 0.01$ ). The effect directionality indicates that inanimate subjects are more likely to be overtly realized and animate subjects are more likely to be null.

## **6 Discussion**

Whether bilinguals drop subjects to similar degrees as other groups has sparked a lively debate in bilingualism research (Tsimpli et al. 2004, Argyri & Sorace 2007, Haznedar 2010, Gagarina 2012, Dubinina & Polinsky 2013, Ivanova-Sullivan 2014, 2015, Koban Koç 2016, Azar et al. 2020). A common observation has been that bilinguals tend to use a higher proportion of overt to null subjects than do monolingual speakers. Our study provides a comparative approach to three typologically different languages concerning the use of null subjects. We conducted an exploration in heritage and monolingual varieties of Greek, Russian, and Turkish to investigate intra- and inter-group convergence or divergence regarding pro-drop. The findings suggest that different patterns are observed depending on the type of pro-drop language in use. In Greek as a strict-pro-drop language, no differences are found between speakers in Germany and Greece, but speakers in the US produced more overt subjects. A more frequent realization of subjects is also observed in heritage Russian and heritage Turkish in Germany, which are partial pro-drop and topic pro-drop languages, respectively. No significant differences were observed in heritage Russian and heritage Turkish in the US. In addition, we controlled for other factors that might affect the production of overt or null subjects, and the most significant one across the languages turned out to be animacy. Furthermore, formality and mode appear to have a significant impact on the frequency of overt subjects in Russian and Turkish as well.

From early on, cross-linguistic influence has been discussed as a key facilitator of more overt subjects in heritage speakers of pro-drop languages with non-pro-drop majority languages (Gawlitzenk-Maiwald & Tracy 1996, Döpke 2000). However, the literature for Greek as a pro-drop heritage language frequently showed

no differences between bilinguals and monolinguals (Di Domenico & Baroncini 2019, Daskalaki et al. 2019, Andreou 2020, Torregrossa et al. 2021, Giannakou & Sitaridou 2022). Our study adds to this literature with the Greek heritage group in Germany producing similar levels of null subjects as monolingually-raised Greek speakers. Heritage speakers of Greek in the US in our sample align with the general trend in the literature towards a higher use of overt subjects (though this effect was not confirmed in our Bayesian re-analysis of the data in Özsoy et al. (2023) which yielded no differences between the Greek groups). Similarly, Russian heritage speaker groups in the US and in Germany differ significantly from the monolingual variety regarding the frequency of overt subjects. This supports findings by Isurin (2011), Dubinina & Polinsky (2013) and contradict Nagy et al. (2011) who found no influence of majority English on subject drop in Russian. As for Turkish, the findings in the literature have been split with Haznedar (2010) and Koban Koç (2016) finding a higher use in overt subjects whereas Azar et al. (2020), Dikilitaş (2017) do not observe any change in heritage Turkishes. This is particularly interesting since Haznedar (2010), Koban Koç (2016) and Dikilitaş (2017) also investigated heritage Turkish speakers with English as a majority language and found mixed results. We find a higher use in Turkish in Germany but not in the US which confirms the mixed nature of the results.

To contextualize these results better, we draw on an explanation that goes beyond cross-linguistic influence and takes the different types of null subject languages in this study, namely, strict pro-drop Greek, partial pro-drop Russian and topic pro-drop Turkish into account (Philippaki-Warburton 1987, Öztürk 2008, Shushurin 2018). According to this perspective, we would find a continuum from more strict to less strict pro-drop languages. On the strict side of the continuum, we expect the pro-drop system to behave more stably under language contact influence in bilinguals. The more we move toward the less strict system, the more dynamicity in the pro-drop system we would expect. Dynamicity could be reflected as a higher use of overt subjects in bilingual speakers. Since the design of our study did not include pro-drop majority languages, we cannot infer whether the dynamics that we observed here are due to language internal change in bilingual grammars or due to cross-linguistic influence from the majority languages. We also were not able to attribute change to factors that we did not control for, such as sociolinguistic background factors.

There are fewer studies that investigate the dynamics in subject use from a perspective that goes from the heritage to the majority language. In this case, we could have expected bi-directional cross-linguistic influence given the interactions in the systems of bilingual grammars (Zhou et al. 2021). However, this does not seem to be the case as our German topic drop results are in line with previous

studies that found stability in heritage bilingual's majority German grammars (Schmitz et al. 2012, Flores & Rinke 2020, Brehmer et al. 2023). This confirms a general trend, in that more often the status of the language, majority or heritage, seems to play a role for linguistic innovation and change (Wiese et al. 2022).

Besides the effects of bilingualism, we explored different extra- and intra-linguistic factors which could influence the realization of pronominal subjects. Outside of the bilingualism literature, animacy has been extensively investigated as an intra-linguistic factor influencing subject omission, with the clear finding that animate subjects are omitted more often than inanimate subjects (Pu 1997, Meyerhoff 2009, Li et al. 2012, Vollmer 2019). However, to our knowledge, the present study is the first systematic investigation regarding the relation between animacy and its impact on the realization of subjects in (heritage) bilinguals. Our findings regarding animacy in Greek, Russian and Turkish indicate that this factor has a significant effect on subject realization. According to the animacy hierarchies given in Corbett (2000), animate subjects are more salient and thus can be easily omitted. Generally, the issue of animacy as a factor is related and overlapping with other concepts that have been found to be meaningful for the realization of subjects such as topicality or cognitive accessibility (Lee 2007, Allen et al. 2015, Laleko & Polinsky 2017). All of these concepts relate to how present the subject referent is in the speaker's mind or the general context.

As for the contexts of formality and mode in our methodology, we explored whether these extra-linguistic factors affect the realization of subjects. These factors are largely uncharted in the framework of bilingual null subject use. In a study with German (monolingual) corpora, Frick (2017) found that null subjects were more frequent in informal brief text messages such as SMS. In our study, the results suggest that the communicative situation does not affect the three languages in the same way. In Greek, neither formality nor mode seem to be significant factors. The picture is different for Russian and Turkish as formality and mode have an impact on subject realization. Specifically, in both languages participants produce more overt subjects in a formal communicative situation. However, the realization patterns diverge regarding mode. While Russian speakers produce more overt subjects in the spoken mode, the probability of overt subjects in the Turkish group is higher in the written mode. Generally, these findings support our suggested continuum from more strict and stable to less strict and dynamic pro-drop languages that align with the typological differences between Greek, Russian and Turkish. Specifically, the strict pro-drop language Greek does not seem to be affected by extra-linguistic factors and relies more on agreement-marking to license null subjects. Leaving aside the typological differences concerning null subjects in Russian and Turkish, being less explicit in the

informal communicative situation can be related to the relaxed tenor of discourse, the reduced attentiveness of the speaker and in general the nonchalant style of informal communicative situations (Stowell & Massam 2017). On the other hand, one would expect that explicitness conforms with written mode so any misunderstanding could be avoided as the paralinguistic features are absent from all communicative situations in written mode. Thus the question why the heritage groups diverge regarding the extra-linguistic factor mode remains an open issue and a desideratum for future research.

Based on our findings from the elicited production task at hand, we could build hypotheses for experimental tasks that focus on the processing and comprehension of pro-drop by speakers of languages that belong to different types of null subject languages. Specifically for the different groups of heritage Greek, Russian and Turkish speakers, we would expect that heritage speakers of Greek are more restricted in their acceptance of non-canonical overt subjects, but Russian and Turkish heritage speakers might allow a more free use of overt subjects. This would allow for testing whether the observation regarding typological differences in dynamicity in null subject use only holds for production or whether it extends to comprehension as well.

A core finding from our study on heritage Greek, Russian and Turkish is the observation that null subject use in the heritage language seems to be more stable or dynamic based on the underlying type of pro-drop language. While our study allows for more cross-linguistic comparisons than previous studies, it still largely relies on languages that can broadly be classified within a pro-drop theory which depends on agreement marking (Rizzi 1986). Therefore, expanding our observation to other languages within the typology of null subject languages would be interesting. For example, it would be interesting to see how radical-pro-drop languages like Mandarin Chinese, which lack agreement marking, behave in this cross-linguistic comparison (Neeleman & Szendrői 2007). Similarly, we would need to explore the role of partial pro-drop languages like Hebrew and Finnish, which allow null subjects in first and second person but not in third (Camacho 2013). This wider cross-linguistic comparison may allow us to test and generalize our hypothesis that languages which require more null subjects will behave more stably, whereas languages that allow overt subjects more freely will show more dynamicity under language contact. For example, the few studies that exist on the acquisition of null subjects by Chinese heritage language speakers indicate that their use of null subjects stays stable (Chou et al. 2020, Zhang 2021), which would be in line with our predictions. This would not only extend the typological diversity in the study of null subjects, but it could also open ways for

the formalization and theorizing regarding the change of the pro-drop parameter in bi- and multilingual grammars of heritage speakers.

To our knowledge, this study is the first that compares subject realization in Greek, Russian and Turkish heritage speakers in Germany and the US and accounts for animacy, formality and mode variation. The study offers insights for the same phenomenon in three heritage varieties in contact with two majority languages, making the outcome comparable for the three populations of interest. So far, studies in the field of heritage languages have rarely used a unified methodology for different languages, which raises questions about the generalizability of the results (Winter & Grice 2021). In contrast, our study offers cross-linguistically comparable data from different heritage communities residing in the same countries and selected according to the same demographic criteria. Its wider implications for bilingualism are manifold. First, since in many languages morphological agreement and syntactic marking are crucial for allowing null subjects, heritage speakers' variability in this domain informs us about the grammatical mechanisms in the bilingual mind. Second, the three languages under scope of this study represent different types of pro-drop languages (strict, partial, topic pro-drop). Combining these different types with two non-pro-drop languages (German or English) in a bilingual's mind gives us insights into diverse outcomes. We can provide evidence of the divergence of heritage languages from the monolingual variety and confirm the unlikeliness of some scenarios in language, for example, that a strict pro-drop language completely abolishes this feature due to language contact. Rather, there are slight adjustments which we labeled dynamicity, but there is not a complete reorganization of the system.

## Acknowledgements

We would like to thank everyone involved in project P10 of the Research Unit *Emerging Grammars in Language Contact Situations* (RUEG, DFG project no. 313607803).

This work has grown out of a big team effort with many student assistants spending hundreds of hours with detailed annotations. For this reason, we would like to thank all of them: Büşra Çiçek, Zeynep Özal, Borbála Sallai, Mariya Burbelko, Alona Prozorova, Ioanna Kolokytha, Panagiota Papavasiliou, Nina Bredereck. We additionally thank Lea Coy for help with pre-publication formatting.

We also thank the anonymous reviewers as well as audiences at the DGfS 2023, ISBPAC 2022, HL@cross, and GAL 2022 conferences for their helpful and very constructive suggestions. The research was supported through funding by the

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit *Emerging Grammars in Language Contact Situations*, project P10 (DFG grant number: 313607803).

## References

- Adamson, Luke James & Elena Anagnostopoulou. 2024. Interpretability and gender features in coordination: Evidence from Greek. In Robert Autry, Gabriela de la Cruz, Luis A. Irizarry Figueroa, Kristina Mihajlovic, Tianyi Ni, Ryan Smith & Heidi Harley (eds.), *Proceedings of the 39th West Coast Conference on Formal Linguistics*, 11–20. Somerville, MA: Cascadilla Proceedings Project.
- Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language & Linguistic Theory* 16(3). 491–539. DOI: [10.1023/A:1006090432389](https://doi.org/10.1023/A:1006090432389).
- Allen, Shanley E. M., Mary Hughes & Barbora Skarabela. 2015. The role of cognitive accessibility in children's referential choice. In Ludovica Serratrice & Shanley E. M. Allen (eds.), *The acquisition of reference*, 123–153. Amsterdam: John Benjamins.
- Andreou, Maria. 2020. The effects of updating and proficiency on overspecification in American-Greek children. *European Journal of Psychological Research* 7(1). 26–32.
- Argyri, Efrosyni & Antonella Sorace. 2007. Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition* 10(1). 79–99. DOI: [10.1017/S1366728906002835](https://doi.org/10.1017/S1366728906002835).
- Ariel, Mira. 2001. Accessibility theory: An overview. In Ted J.M. Sanders, Joost Schilperoord & Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects* (Human Cognitive Processing 8), 29–87. Amsterdam: John Benjamins. DOI: [10.1075/hcp.8.04ari](https://doi.org/10.1075/hcp.8.04ari).
- Arnold, Jennifer E. 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass* 4(4). 187–203. DOI: [10.1111/j.1749-818X.2010.00193.x](https://doi.org/10.1111/j.1749-818X.2010.00193.x).
- Austin, Jennifer, Liliana Sánchez & Silvia Perez-Cortes. 2017. Null subjects in the early acquisition of English by child heritage speakers of Spanish. *Romance Languages and Linguistic Theory* 11. 209–27.
- Azar, Zeynep, Aslı Özyürek & Ad Backus. 2020. Turkish-Dutch bilinguals maintain language-specific reference tracking strategies in elicited narratives. *International Journal of Bilingualism* 24(2). 376–409. DOI: [10.1177/1367006919838](https://doi.org/10.1177/1367006919838).

- Baayen, Rolf Harald, Douglas Davidson & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). Special Issue: Emerging data analysis, 390–412. DOI: 10.1016/j.jml.2007.12.005.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.
- Benmamoun, Elabbas, Silvina Montrul & Maria Polinsky. 2013. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3-4). 129–181. DOI: 10.1515/tl-2013-0009.
- Bizzarri, Camilla. 2015. Russian as a partial pro-drop language: Data and analysis from a new study. *Annali di Ca' Foscari: Serie occidentale* (49). 335–362.
- Boon, Erin Diane. 2014. *Heritage Welsh: A study of heritage language as the outcome of minority language acquisition and bilingualism*. Harvard University. (Doctoral dissertation).
- Brehmer, Bernhard, Aldona Sopata & Massimiliano Canzi. 2023. Same or different? Subject realization in the majority and the heritage language of Polish-German bilingual children. *Linguistics Vanguard* 9(2). 179–189. DOI: 10.1515/lingvan-2022-0061.
- Brezina, Vaclav & Miriam Meyerhoff. 2014. Significant or random? A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics* 19(1). 1–28. DOI: 10.1075/ijcl.19.1.01bre.
- Britain, David, Carmen Llamas & Dominic Watt. 2010. Supralocal regional dialect levelling. *Language and identities*. 193–204.
- Budennaya, Evgeniya. 2019. Russian contrastive pronouns in Polish translations: A corpus-based study [Russkie kontrastivnye mestoimenija v pol'skix perevodax: Opyt korpusnogo issledovaniya]. *Tipologija morfosintakscičeskix parametrov* 2(2). 11–26.
- Budennaya, Evgeniya. 2020. Sub"ekt nastojaščego vremeni v ustnyx i pis'mennyx narrativax [The subject of the present tense in oral and written narratives]. In A. A. Kibrik, Ks. P. Semёnova, D. V. Sičinava, S. G. Tatevosov & A. Ju. Urmančieva (eds.), *VAProsy jazykoznanija: Megasbornik nanostatej. Sb. st. k jubileju V. A. Plungjana [Problems of linguistics: A mega-collection of Nanostatas. Collected articles for the anniversary of V. A. Plungian]*, 648–653.
- Camacho, José. 2013. *Null subjects*. Cambridge: Cambridge University Press.
- Chafe, Wallace L. 1987. Cognitive constraints on information flow. In Russell S. Tomlin (ed.), *Coherence and grounding in discourse: Outcome of a symposium, Eugene, Oregon, June 1984* (Typological Studies in Language 11), 21–51. Amsterdam: John Benjamins. DOI: 10.1075/tsl.11.03cha.

- Chociej, Joanna. 2011. Polish null subjects: English influence on heritage Polish in Toronto. *Ms., University of Toronto*.
- Chou, Chao-Ting Tim, Tsung-Ying Chen & Acrisio Pires. 2020. Acquisition of null objects in Mandarin Chinese by heritage speakers: Syntax-pragmatics interface knowledge without inflectional morphology. *International Journal of Chinese Linguistics* 7(2). 223–269. DOI: 10.1075/ijchl.19016.cho.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. The University of Chicago Press.
- Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15(2). 203–224.
- Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.
- Daskalaki, Evangelia, Vasiliki Chondrogianni, Elma Blom, Froso Argyri & Johanne Paradis. 2019. Input effects across domains: The case of Greek subjects in child heritage language. *Second Language Research* 35(3). 421–445. DOI: 10.1177/0267658318787.
- Di Domenico, Elisa & Ioli Baroncini. 2019. Age of onset and dominance in the choice of subject anaphoric devices: Comparing natives and near-natives of two null-subject languages. *Frontiers in Psychology* 9. 2729. DOI: 10.3389/fpsyg.2018.02729.
- Dikilitaş, Kenan. 2017. Acquisition of pro drop by a Turkish-English bilingual child. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* 32(1). 33–45.
- Döpke, Susanne. 2000. Generation of and retraction from cross-linguistically motivated structures in bilingual first language acquisition. *Bilingualism: Language and Cognition* 3(3). 209–226. DOI: 10.1017/S1366728900000341.
- Dryer, Matthew S. 2013. Expression of pronominal subjects. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/101>.
- Dubinina, Irina & Maria Polinsky. 2013. Russian in the USA. In Michael Moser & Maria Polinsky (eds.), *Slavic languages in migration* (Slavische Sprachgeschichte 6). Wien: Lit Verlag.
- Enç, Mürvet. 1986. Topic switching and pronominal subjects in Turkish. *Studies in Turkish linguistics* 8. 195.
- Fehrman, Dorothee & Uwe Junghanns. 2008. Subjects and scales. *Linguistische Arbeitsberichte* 86. 189–220.
- Flores, Cristina & Esther Rinke. 2020. Factors constraining subject expression in European Portuguese spoken in Hamburg: A bi-generational corpus investigation. *Discours* (26). 3–31. DOI: 10.4000/discours.10648.

- Fougeron, I. I. & Žan Breillard. 2004. Mestoimenie ja i postroenie diskursivnyx svjazej v sovremennom russkom jazyke. [The pronoun *ja* and the construction of discourse relations in modern Russian]. In Tat'jana Mixajlovna Nikolaeva, I.I. Fužeron, V.S. Efimova, R.F. Kasatkina, L.L. Kasatkin, A.V. Vencov & Žan Brejar (eds.), *Verbal'naja i neverbal'naja opory prostranstva mežfrazovyx svjazej* [Verbal and nonverbal supports of interphrase space]. Jazyki slavjanskoj kul'tury.
- Frascarelli, Mara. 2018. The interpretation of pro in consistent and partial null-subject languages: A comparative interface analysis. In Federica Cognola & Jan Casalicchio (eds.), *Null subjects in generative grammar: A synchronic and diachronic perspective*, 211–239. Oxford: Oxford University Press.
- Frick, Karina. 2017. *Elliptische Strukturen in SMS: Eine korpusbasierte Untersuchung des Schweizerdeutschen*. Berlin: De Gruyter. DOI: 10.1515/9783110517859.
- Fries, Norbert. 1988. Über das Null-Topik im Deutschen. *Sprache und Pragmatik: Arbeitsberichte* 3. 19–49.
- Gagarina, Natalia. 2012. Discourse cohesion in the elicited narratives of early Russian-German sequential bilinguals. In Kurt Braumüller & Christoph Gabriel (eds.), *Multilingual individuals and multilingual societies* (Hamburg Studies on Multilingualism 13), 101–120. Amsterdam: John Benjamins. DOI: 10.1075/hsm.13.08gag.
- Gawlitzek-Maiwald, Ira & Rosemarie Tracy. 1996. Bilingual bootstrapping. *Linguistics* 34(5). 901–926. DOI: 10.1515/ling.1996.34.5.901.
- Giannakou, Aretousa & Ioanna Sitaridou. 2022. (In)felicitous use of subjects in Greek and Spanish in monolingual and contact settings. *Glossa: A journal of general linguistics* 7(1). DOI: 10.16995/glossa.5812.
- Guirarro-Fuentes, Pedro & Katrin Schmitz. 2015. The nature and nurture of heritage language acquisition. *Lingua* 164. 239–250. DOI: 10.1016/j.lingua.2015.05.008.
- Gundel, Jeanette K, Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274–307.
- Haegeman, Liliane. 2010. Non-overt subjects in diary contexts. In Joan Mascaró & Marina Nespor (eds.), *Grammar in progress*, 167–174. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110867848.167.
- Haegeman, Liliane & Tabea Ihsane. 2001. Adult null subjects in the non-prop-drop languages: Two diary dialects. *Language Acquisition* 9(4). 329–346. DOI: 10.1207/S15327817LA0904\_03.
- Harvie, Dawn. 1998. Null subject in English: Wonder if it exists? *Cahiers Linguistique Inguistique d'Ottawa*. 15–26.

- Haznedar, Belma. 2010. Transfer at the syntax-pragmatics interface: Pronominal subjects in bilingual Turkish. *Second Language Research* 26(3). 355–378. DOI: 10.1177/02676583103657.
- Holmberg, Anders. 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36(4). 533–564. DOI: 10.1162/002438905774464322.
- Hulk, Aafke & Natascha Müller. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3(3). 227–244. DOI: 10.1017/S1366728900000353.
- Isurin, Ludmila. 2011. *Russian diaspora: Culture, identity, and language change* (Contributions to the Sociology of Language 99). Walter de Gruyter. DOI: 10.1515/9781934078457.
- Ivanova-Sullivan, Tania. 2014. Interpretation of anaphoric subject pronouns in heritage Russian. In Ryan T. Miller, Katherine I. Martin, Chelsea M. Eddington, Ashlie Henery, Nausica Marcos Miguel, Alison M. Tseng, Alba Tuninetti & Daniel Walter (eds.), *Selected proceedings of the 2012 second language research forum: Building bridges between disciplines*, 145–154. Somerville, MA.
- Ivanova-Sullivan, Tania. 2015. *Theoretical and experimental aspects of syntax-discourse interface in heritage grammars*. Leiden: Brill.
- Koban Koç, Didem. 2016. Social variables and Turkish subject pronoun use in New York City: The effect of language contact. *Poznan Studies in Contemporary Linguistics* 52(3). 431–453.
- Krause, Thomas & Amir Zeldes. 2014. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1). 118–139. DOI: 10.1093/llc/fqu057.
- Laleko, Oksana & Maria Polinsky. 2017. Silence is difficult: On missing elements in bilingual grammars. *Zeitschrift für Sprachwissenschaft* 36(1). 135–163. DOI: 10.1515/zfs-2017-0007.
- Lee, Chungmin. 2007. Contrastive (predicate) topic, intonation, and scalar meanings. In Chungmin Lee, Matthew Gordon & Daniel Büring (eds.), *Topic and focus*, 151–175. Springer. DOI: 10.1007/978-1-4020-4796-1\_9.
- Li, Xiaoshi, Xiaoqing Chen & Wen-Hsin Chen. 2012. Variation of subject pronominal expression in Mandarin Chinese. *Sociolinguistic Studies* 6(1). 91. DOI: 10.1558/sols.v6i1.91.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Anniqa Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Heike Wiese, Sabine Zerbian, Yu-

- lia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Madariaga, Nerea. 2018. Diachronic change and the nature of pronominal null subjects. In Federica Cognola & Jan Casalicchio (eds.), *Null subjects in generative grammar: A synchronic and diachronic perspective*, 171–198. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198815853.003.0007.
- Madariaga, Nerea. 2022. Referential null subjects in Russian: A synchronic and diachronic overview. In Gréte Dalmi, Egor Tsedryk & Piotr Ceglowski (eds.), *Null subjects in Slavic and Finno-Ugric: Licensing, structure and typology*, 75–103. Boston/Berlin: de Gruyter Mouton. DOI: 10.1515/9781501513848-003.
- Margaza, Panagiota & Anna Gavarró. 2020. Null/overt subject alternations in L2 Spanish and L2 Greek. *Glossa: A journal of general linguistics* 5(1). DOI: 10.5334/gjgl.969.
- Martynova, Maria, Onur Özsoy, Vasiliki Rizou, Artemis Alexiadou, Natalia Gagarina & Luka Szucsich. 2024. Demonstratives in heritage Greek, Russian, and Turkish in Germany and the US. *International Journal of Bilingualism*. DOI: 10.1177/1367006924126105.
- McCloskey, Jim. 1997. Subjecthood and subject positions. In Liliane Haegeman (ed.), *Elements of grammar: Handbook in generative syntax*, 197–235. Springer.
- Meyerhoff, Miriam. 2009. Replication, transfer, and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change* 21(3). 297–317. DOI: 10.1017/S0954394509990196.
- Montrul, Silvina. 2015. *The acquisition of heritage languages*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO978139030502.
- Müller, Natascha & Aafke Hulk. 2001. Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4(1). 1–21. DOI: 10.1017/S1366728901000116.
- Nagy, Naomi G., Nina Aghdasi, Derek Denis & Alexandra Motut. 2011. Null subjects in heritage languages: Contact effects in a cross-linguistic context. *University of Pennsylvania Working Papers in Linguistics* 17(2). 16.
- Neeleman, Ad & Kriszta Szendrői. 2007. Radical pro drop and the morphology of pronouns. *Linguistic Inquiry* 38(4). 671–714. DOI: 10.1162/ling.2007.38.4.671.
- Oppermann, Sophia. 2021. Subject gap coordination. *Journal of Historical Syntax* 5(16–25). 1–38. DOI: 10.18148/hs/2021.v5i16-25.58.
- Özsoy, Onur & Mariya Burbelko. In preparation. *Topic drop in mono- and bilingual native German speakers*.
- Özsoy, Onur, Maria Martynova, Vasiliki Rizou, Natalia Gagarina, Luka Szucsich & Artemis Alexiadou. 2023. *Pro-drop realization in heritage Greek, Russian and Turkish is modulated by typological differences*. Paper presented at Deutsche

- Gesellschaft für Sprachwissenschaft (DGfS), Universität zu Köln, Cologne, Germany.
- Öztürk, Balkız. 2001. Turkish as a non-pro-drop language. In Eser Erguvanlı Taylan (ed.), *The verb in Turkish* (Linguistik Aktuell/Linguistics Today 44), 239–259. Amsterdam: John Benjamins. DOI: 10.1075/la.44.10ozt.
- Öztürk, Balkız. 2008. Non-configurationality: Free word order and argument drop in Turkish. In Theresa Biberauer (ed.), *The limits of syntactic variation*, 411–440. Amsterdam: John Benjamins.
- Paradis, Johanne & Samuel Navarro. 2003. Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language* 30(2). 371–393. DOI: 10.1017/S0305000903005609.
- Pashkova, Tatiana, Mark Murphy & Shanley E. M. Allen. 2020. *Explicitness in referent introduction in heritage speakers' majority English*. Poster presented at Boston University Conference on Language Development (BUCLD 45). Boston, USA.
- Pekelis, Olga E. 2018. Èllipsis podležaščego glavnoj klauzy v russkom jazyke i ego svjaz's tipologiej korreljatov [Subject ellipsis in the main clause in Russian and the typology of correlatives]. *Voprosy jazykoznanija* (6). 31–59.
- Perlmutter, David M. 1971. *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart & Winston.
- Philippaki-Warburton, Irene. 1987. The theory of empty categories and the pro-drop parameter in modern Greek. *Journal of Linguistics* 23(2). 289–318. DOI: 10.1017/S002226700011282.
- Polinsky, Maria. 2015. Heritage languages and their speakers: State of the field, challenges, perspectives for future work, and methodologies. *Zeitschrift für Fremdsprachwissenschaft* 26. 7–27.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Pu, Ming-Ming. 1997. Zero anaphora and grammatical relations in Mandarin. *Typological Studies in Language* 35. 281–322. DOI: 10.1075/tsl.35.08pu.
- Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17(3). 501–557.
- Roberts, Ian & Anders Holmberg. 2010. Introduction: Parameters in Minimalist theory. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null subjects in Minimalist theory*, 1–57. Cambridge: Cambridge University Press.

- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2). 155–163. DOI: 10.1177/1367006909339814.
- Rothman, Jason, Fatih Bayram, Vincent DeLuca, Grazia Di Pisa, Jon Andoni Duñabeitia, Khadij Gharibi, Jiuzhou Hao, Nadine Kolb, Maki Kubota, Tanja Kupisch, Tim Laméris, Alicia Luque, Brechje van Osch, Sergio Miguel Pereira Soares, Yanina Prystauka, Deniz Tat, Aleksandra Tomić, Toms Voits & Stefanie Wulff. 2023. Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics* 44(3). 316–329. DOI: 10.1017/S0142716422000315.
- Schäfer, Lisa. 2021. Topic drop in German: Empirical support for an information-theoretic account to a long-known omission phenomenon. *Zeitschrift für Sprachwissenschaft* 40(2). 161–197. DOI: 10.1515/zfs-2021-2024.
- Schmid, Helmut. 1997. Probabilistic part-of-speech tagging using decision trees. In Daniel B. Jones & Harold Somers (eds.), *New methods in language processing*, 154–164. London: Routledge.
- Schmidt, Thomas & Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *Handbook on corpus phonology*, 402–419. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199571932.013.030.
- Schmitz, Katrin, Marisa Patuto & Natascha Müller. 2012. The null-subject parameter at the interface between syntax and pragmatics: Evidence from bilingual German–Italian, German–French and Italian–French children. *First Language* 32(1-2). 205–238. DOI: 10.1177/0142723711403880.
- Shushurin, Philip. 2018. Null pronouns in Russian embedded clauses. In Pritty Patel-Grosz, Patrick Georg Grosz & Sarah Zobel (eds.), *Pronouns in embedded contexts at the syntax-semantics interface*, 145–169. Cham: Springer. DOI: 10.1007/978-3-319-56706-8\_5.
- Silverstein, Michael & Robert MW Dixon. 1986. Hierarchy of features and ergativity. *Features and Projections* 25. 163–232.
- Sorace, Antonella. 2011. Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism* 1(1). 1–33. DOI: 10.1075/lab.1.1.01sor.
- Stowell, Tim & Diane Massam. 2017. Introducing register variation and syntactic theory. *Linguistic Variation* 17(2). 149–156. DOI: 10.1075/lv.00001.mas.
- Torregrasso, Jacopo, Maria Andreou, Christiane Bongartz & Ianthi-Maria Tsimpli. 2021. Bilingual acquisition of reference: The role of language experience, executive functions and cross-linguistic effects. *Bilingualism: Language and Cognition* 24(4). 694–706. DOI: 10.1017/S1366728920000826.

- Tracy, Rosemarie. 2000. Language mixing as a challenge for linguistics. In *Cross-linguistic structures in simultaneous bilingualism* (Studies in Bilingualism 21), 11–36. Amsterdam: John Benjamins. DOI: 10.1075/sibil.21.04tra.
- Tracy, Rosemarie. 2014. Mehrsprachigkeit: Vom Störfall zum Glücksfall. In Manfred Krifka, Joanna Błaszcak, Annette Leßmöllmann, André Meinunger, Barbara Stiebels, Rosemarie Tracy & Hubert Truckenbrodt (eds.), *Das mehrsprachige Klassenzimmer: Über die Muttersprachen unserer Schüler* (Springer Spektrum), 13–33. Berlin Heidelberg: Springer VS. DOI: 10.1007/978-3-34315-5.
- Trutkowski, Ewa. 2016. *Topic drop and null subjects in German* (Linguistics & Philosophy 6). Berlin: de Gruyter. DOI: 10.1515/9783110446173.
- Tsimpli, Ianthi-Maria, Antonella Sorace, Caroline Heycock & Francesca Filiaci. 2004. First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism* 8(3). 257–277. DOI: 10.1177/13670069040080030601.
- Valdés, Guadelupe. 2005. Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? *The Modern Language Journal* 89(3). 410–426. DOI: 10.1111/j.1540-4781.2005.00314.x.
- Vollmer, Maria C. 2019. *How radical is pro-drop in Mandarin? A quantitative corpus study on referential choice in Mandarin Chinese*. University of Bamberg. (Doctoral dissertation).
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarcina, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehayre, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Winter, Bodo. 2011. Pseudoreplication in phonetic research. In *17th International Congress of Phonetic Sciences (ICPhS)*, 2137–2140. Hong Kong, China.
- Winter, Bodo. 2019. *Statistics for linguists: An introduction using R*. Routledge.
- Winter, Bodo & Martine Grice. 2021. Independence and generalizability in linguistics. *Linguistics* 59(5). 1251–1277. DOI: 10.1515/ling-2019-0049.
- Zhang, Xinye. 2021. Language variation in mandarin as a heritage language: Subject personal pronouns. *Heritage Language Journal* 18(1). 1–29. DOI: 10.1163/15507076-12340020.

*6 Null subjects in heritage Greek, Russian and Turkish*

Zhou, Jiangling, Ziyin Mai & Virginia Yip. 2021. Bidirectional cross-linguistic influence in object realization in Cantonese–English bilingual children. *Bilingualism: Language and Cognition* 24(1). 96–110. DOI: 10.1017/S1366728920000231.



# Chapter 7

## Dynamics of verbal aspect in heritage Greek, Russian and Turkish

Vasiliki Rizou<sup>a</sup>, Onur Özsoy<sup>b,a</sup>, Maria Martynova<sup>a</sup>, Luka Szucsich<sup>a</sup>, Artemis Alexiadou<sup>b,a</sup> & Natalia Gagarina<sup>b,a</sup>

<sup>a</sup>Humboldt-Universität zu Berlin <sup>b</sup>Leibniz-Centre General Linguistics

Aspect – the grammatical category associated with the internal temporal properties of a situation – is obligatorily marked in the grammar of many languages. Although it is used pervasively, innovative patterns commonly appear in situations of language contact. Thus, aspect provides an ideal window into the dynamic language patterns that occur in heritage grammars. This chapter reviews six studies conducted in the context of the Research Unit *Emerging Grammars* that investigate dynamic patterns in the use of aspect in monolingually-raised speakers of Greek, Russian, and Turkish as well as heritage speakers of each of these languages living in the US and Germany. Each study uses production and/or comprehension tasks to explore one or more potential sources of these dynamic patterns, including cross-linguistic influence, verbal morphology and markedness of the aspectual features, formality (formal vs. informal), mode of expression (spoken vs. written), and participant age. The results show several innovative patterns and some differences in processing in heritage speakers, but largely find that monolingually-raised and heritage speakers produce and process aspect marking in a qualitatively similar way.

### 1 Introduction

In this chapter, we investigate how verbal aspect, either grammatical or lexical, is used in heritage and monolingual varieties of Greek, Russian, and Turkish and how it reflects the dynamic language patterns that occur in heritage grammars.



In particular, we review six studies concerning the production and comprehension of verbal aspect that were conducted as part of the Research Unit *Emerging Grammars in Language Contact Situations: A Comparative Approach*. Each study takes a different perspective on the use of aspect in these varieties (heritage and monolingual) and asks whether aspect use is affected by the language itself (Greek, Russian, Turkish), by the contact language (English, German), and/or by the mode (spoken, written) or formality (formal, informal) of the situation. The studies also cover a number of methods, including elicited narrative, sentence completion tasks, spontaneous speech production, and the Visual World paradigm.

Aspect is one of the most widely explored phenomena across different languages and different groups of speakers, including both monolingually-raised and bilingual speakers (Montrul 2002, Laleko 2010, Montrul & Perpiñán 2011, Cuza et al. 2013). It is described as representing the internal temporal structure of a situation, which is different from the situation's external time, referred to as tense. Researchers typically distinguish between lexical (situation/Aktionsart/inner) and grammatical (viewpoint/outer) aspect (Comrie 1976, Kiyota 2008, Alexiadou et al. 2010). Grammatical aspect is defined as “the different ways of viewing the internal temporal constituency of a situation”, while lexical aspect denotes the inherent semantic category of a verb's eventuality (Comrie 1976). Grammatical aspect is a (morpho-)syntactic category and can be variably marked on verbs, while lexical aspect can be determined by the whole verbal phrase. The studies reviewed in this chapter focus either on grammatical aspect, i.e. Rizou et al. (2024), Alexiadou & Rizou (2023), Rizou et al. (2025) and Özsoy, Büyükyıldırım, et al. (2023), or on the interaction of both Rizou (2021), Gagarina et al. (2020).

A major distinction for grammatical aspect is made between imperfective and perfective (Comrie 1976). Imperfective aspect denotes an event or state as progressive, repeating or continuous. Perfective aspect is used to express that an event or an action is completed, or that is ceased without any reference to its tempus. In other words, perfective aspect indicates that an event is bounded in time. Although aspect has several further distinctions that differ across languages (Gagarina 2000), these are not relevant for the current chapter so will not be elaborated here.

Aspect is an ideal phenomenon to investigate in language development because it is regarded as highly dynamic across the spectrum from early L1 acquisition through heritage language acquisition and L2 acquisition. For example, L1 Greek-speaking children produce perfective aspect as early as 1;1,<sup>1</sup> before the

---

<sup>1</sup>We report child age as years;months. Thus, 1;1 means 1 year and 1 month old.

imperfective (Stephany 1997, Konstantzou et al. 2013). Although imperfective aspect is also used relatively early, the semantics of imperfective aspect are only acquired later by 5;0 to 6;5 (Delidakis 2006, Panitsa 2010). L1 Russian-speaking children recognize aspectual meanings at an early stage and begin to use aspectual forms that resemble the target language from the onset of speech. However, they occasionally generate novel expressions or innovations, wherein aspectual forms deviate from the target language norms, that may persist until approximately the age of 6 years (Cejtin 2000, Gagarina 2007, 2011, Kistanova & Sekerina 2019). Finally, L1 Turkish-speaking children acquire the primary aspect/tense markers quite early, namely the perfective-past marker -(y)DI and the progressive-imperfective marker -(I)yor (Aksu-Koç 1988). These distinct markers serve as a significant focal point for understanding the contrast between perfective and imperfective aspect and are productively used by around the age of 3 (Acarlar et al. 2008).

Innovative aspectual forms are also used regularly in situations of language contact, particularly in languages with concatenative morphology like Greek and Russian, as well as in languages with agglutinating structures like Turkish (Laleko 2008, Montrul 2015, Antonova-Ünlü & Wei 2016, Gagarina et al. 2020). This is illustrated in the following example in (1) taken from Gagarina et al. (2020: 115), where the perfective aspect form combines with the auxiliary *budu* ‘I will (be)’ in the future, a form that is considered ungrammatical in monolingual productions:

- (1) budu \*napisat'  
 will-1SG write-PFV  
 '(I) will write'

Given that aspect appears with dynamic and innovative forms across both L1 and L2 acquisition, it seems to be the ideal phenomenon to explore in heritage language use as well. Heritage speakers are typically described as individuals who are bilingual or multilingual, acquire their heritage language(s) during early childhood within their families, and use their heritage language(s) alongside the majority language(s) or official language(s) of the community they live in (Valdés 2005, Rothman 2009, Benmamoun et al. 2013, Guijarro-Fuentes & Schmitz 2015, Montrul 2015, Polinsky 2015, 2018). Generally, heritage speakers are regarded as bilinguals whose family and majority languages are part of a nativeness continuum (Wiese et al. 2022). However, the level of proficiency in their languages can vary throughout their lives, as discussed by Benmamoun et al. (2013), Polinsky (2018), and Wiese (2020). In this chapter, we follow this perspective and do

not compare heritage speakers' data with monolingually-raised speakers' data to explore accuracy. Instead, we consider the heritage and monolingually-raised populations as distinct language varieties, in line with the view expressed in Rothman et al. (2023).

Due to the informal nature of the acquisition of heritage languages, speakers are more accustomed to the spoken form of the language, while the written form requires formal instruction. Research exploring register variation in heritage speakers suggests that when heritage languages are learned within the family context, without receiving any formal instruction in the heritage language, heritage speakers' repertoires exhibit a conversational and informal style, which is often limited to everyday topics, resulting in register narrowing (Dressler 1991, Chevalier 2004). Therefore, several of the studies that we review specifically focus on whether aspect is used differently in spoken vs. written language, and in formal vs. informal settings.

The six studies reviewed in this chapter were conducted within the projects P1, P3 and P10 of the Research Unit *Emerging Grammars* (RUEG)<sup>2</sup>, aiming to address the overarching question: How do speakers of heritage and monolingual varieties of Greek, Russian, and Turkish use aspect? Specifically, we investigate whether heritage speakers' production and comprehension of aspect diverges from that of monolingually-raised speakers. We account for different sources of variability including cross-linguistic differences, and language transfer effects. The studies below shed light on different levels of the phenomenon.

The structure of this chapter is as follows. In Section 2, we review a comparative study on grammatical aspect in Greek, Russian, and Turkish (Rizou et al. 2024). This study investigates the monolingual varieties of Greek, Russian, and Turkish as well as their heritage variants in Germany and the US regarding the morphological marking of grammatical aspect. By keeping the analysis as parallel as possible in the three languages, it explores how monolingually-raised and heritage speakers perform on perfective and imperfective aspect use, whether effects of cross-linguistic influence from the majority languages are evident, and how the parameters of formality and mode affect their aspect-related productions. Section 3 highlights a study exploring the use of periphrastic aspectual constructions versus the use of aspect on lexical verbs in different formality levels by Greek heritage and monolingually-raised speakers (Alexiadou & Rizou

---

<sup>2</sup><https://hu.berlin/rueg>. P1 dealt with “Nominal morphosyntax and word order in heritage Greek across majority languages”, P3 with “Nominal morphosyntax and word order in heritage Russian across majority languages”, and P10 with “Dynamics of verbal aspect and (pro)nominal reference in language contact”.

2023). In Section 4, we report a study of the accuracy of Greek monolingually-raised and heritage speakers in an elicited production task with items controlling the interaction of grammatical and lexical aspect (Rizou 2021). The same methodology is employed by the study reported in Section 5, in which morphologically novel forms that appear to be challenging in heritage speakers' grammars are analyzed (Rizou et al. 2025). This is followed in Section 6 by a review of a qualitative study by Gagarina et al. (2020), analyzing aspect use in child, adolescent, and adult productions of Russian heritage speakers in Germany and the US and contrasting it with the productions obtained from monolingually-raised Russian-speaking peers. The last study, in Section 7, focuses on Turkish heritage and monolingually-raised speakers' incremental processing of grammatical aspect (Özsoy, Büyükyıldırım, et al. 2023, Büyükyıldırım et al. in preparation). Using the Visual World Paradigm, this eye-tracking study aims to find the underlying factors that influence processing speed and picture-matching accuracy by individual participants. Finally, in Section 8 we discuss the overall findings and their implications for broadening our understanding of the use of aspect in heritage and monolingually-raised speakers. The novelty of these studies lies in the systematic observation and categorization of verbal structures, providing an overview of verbal aspect cross-linguistically by employing different methodologies and controlling for intra- and extra-linguistic factors.

## 2 A cross-linguistic study on grammatical aspect in heritage and monolingual varieties of Greek, Russian and Turkish by Rizou et al. (2024)

The expression of verbal aspect in Greek, Russian, and Turkish varies in its morphological realization, but all three languages distinguish between perfective and imperfective aspect. In Greek, the stem of the morphologically marked aspect type is formed by suffixation, or weak/strong suppletion (Spyropoulos et al. 2017, Revithiadou et al. 2019). Example (2) shows the addition of the suffix /s/ in the perfective, triggering stem allomorphy.<sup>3</sup>

- (2) a. pez-o  
play.IPFV-PRS.1SG  
'play'

<sup>3</sup>In order to keep the aspectual morphological description of the languages under the scope of this study parallel, more details about the Greek verbal morphology are mentioned in Section 5, about Russian acquisition patterns of aspect in Section 6, and about aspect in Turkish in Section 7.

- b. (na) peks-o  
SBJV play.PFV-PRS.1SG  
'to play'

Most verbs in Russian form aspectual pairs and the distinction between perfective and imperfective aspect is primarily marked through morphology: the perfective aspect is derived from the stem of imperfective verbs through processes such as prefixation (as seen in example 3), suffixation or suppletion, and additionally, imperfective verbs may be derived from perfective ones by suffixation leading to secondary imperfectivization (Gagarina 2008a).<sup>4</sup>

- (3) a. delat'  
do/make.IPFV  
'to do/make'  
b. s-delat'  
PFV-do/make  
'to do/make'

In Turkish, the copula in past tense -(y)DI, the marker -DI and the multifunctional evidential marker -*mḷş* are mainly applied to mark the perfective aspect, while for progressive-imperfective the verbal marker -(I)yor is used. Regarding morphological markedness, the perfective aspect is considered morphologically marked in Greek and Russian, while the imperfective aspect is the marked one in Turkish as seen in example (4).

- (4) a. bitir-iyor-du-m  
finish-PROG-PST-1SG  
'I was finishing.'  
b. bitir-di-m  
go-PST.PFV-1SG  
'I finished.'

The aim of our investigation was to analyze how grammatical aspect is expressed in typologically different languages, namely Greek, Russian, and Turkish. By employing a consistent analytical framework for all three languages, the study explored the usage of perfective and imperfective aspect by both monolingually-raised and heritage speakers of the respective languages in the US and

---

<sup>4</sup>There are verbs that do not form aspectual pairs, such as *perfectiva tantum*, as well as *imperfectiva tantum* (Gagarina 2007).

Germany. An intriguing angle of this investigation is the examination of verbal aspect in heritage Greek, Russian and Turkish in contact with majority English, a language that marks aspect, and with majority German, which does not. The study seeks to understand whether and to what extent the use of aspect in heritage Greek, Russian, and Turkish is influenced by the language contact situation. Furthermore, the study addresses the question of whether the notion of morphological markedness impacts the production of grammatical aspect in heritage Greek, Russian and, Turkish, and if so, how. Additionally, the research delves into the impact of formality and mode on grammatical aspect productions. What distinguishes this study from previous research is its emphasis on cross-linguistically comparable and ecologically-valid data obtained from monolingual and heritage varieties of Greek, Russian, and Turkish through narration tasks.

The data analyzed in Rizou et al. (2024) were collected according to the *language situations* approach by Wiese (2020). This methodology provides comparable naturalistic data in both oral and written modes and in formal and informal registers. During the elicitation, the participants were shown a short video of a staged minor car accident, and their task was to narrate what happened to either a close friend or a police officer, imagining that they witnessed the incident. In that way, participants took part in four different communicative settings within one experimental session. Heritage speakers took part in two sessions at least three days apart, one in their majority and one in their heritage language, while monolingually-raised participants took part only in one session in the majority language of the respective country of elicitation. To test how formality and mode affect narrations, we simulated formal-spoken, formal-written, informal-spoken, and informal-written settings. The elicitation orders were balanced. Table 1 presents the overall information about the number of participants grouped by country of elicitation, their mean chronological age<sup>5</sup> and their age of onset to bilingualism, and the number of tokens in the RUEG subcorpora of Greek, Russian, and Turkish.<sup>6</sup>

Verbs were tagged automatically as either perfective or imperfective and corrected manually by native speakers of Greek, Russian, and Turkish. The data

---

<sup>5</sup>Within each group, two age groups were tested, namely adults (age range 25–35) and adolescents (age range 14–18). In the present study, this further categorization was not taken into account, and all participants were analyzed together as the age variable was not a significant factor for grammatical aspect.

<sup>6</sup>A token is often used to mean something like a graphemic word but is technically defined as the smallest unit in a corpus, independent of what this smallest unit might be linguistically (Schmid 2008, Sauer & Lüdeling 2016).

Table 1: Participants information

| Country | Group            | Participants | Tokens | Verbs | M age | M AoO |
|---------|------------------|--------------|--------|-------|-------|-------|
| Greece  | monolinguals     | 64           | 27,931 | 4,954 | 21.4  | -     |
| Russia  | monolinguals     | 66           | 25,930 | 3,965 | 21.0  | -     |
| Turkey  | monolinguals     | 64           | 20,947 | 4,609 | 22.2  | -     |
| Germany | heritage Greek   | 48           | 19,782 | 3,494 | 22.6  | 1.8   |
|         | heritage Russian | 61           | 32,882 | 4,624 | 21.1  | 1.3   |
|         | heritage Turkish | 64           | 23,722 | 4,986 | 21.5  | 2.5   |
| US      | heritage Greek   | 63           | 18,302 | 4,371 | 23.0  | 1.3   |
|         | heritage Russian | 60           | 29,214 | 4,342 | 22.2  | 3.3   |
|         | heritage Turkish | 58           | 18,502 | 4,257 | 22.0  | 2.5   |

extracted from the corpus were processed and analyzed in R using the *tidyverse* and *brms* packages (Schmidt & Wörner 2014, Wickham et al. 2019, R Core Team 2021). The observations of aspect were categorized as a binary dependent variable with two levels: imperfective and perfective. This resulted in a total of 26,788 data points, with 8,829 from Greek, 11,626 from Russian, and 6,333 from Turkish. To draw cross-linguistic (Greek, Russian and Turkish heritage speaker groups) and cross-variational (monolingually-raised vs. heritage speakers) comparisons, we applied methods from a Bayesian statistical framework (Gelman et al. 2013, Kruschke 2014, Kruschke & Vanpaemel 2015). Bayesian regression models with 89 credibility intervals (CIs) were constructed for each language, incorporating random slopes for participants to accommodate individual differences. The independent variables in these models included Country of elicitation (Germany, US, and the native countries Greece, Russia, and Turkey), Formality (formal vs. informal), Mode (spoken vs. written), and Tense (past vs. present). Tense was included as one of the variables because of the interdependence between tense and grammatical aspect. Employing a uniform model structure across all languages ensures comparability in interpreting the outcomes. The availability of both data and analysis code in an open repository ensures the reproducibility of the analysis.

Applying descriptive statistics, we observed a different distribution of perfective and imperfective in the three languages. Greek, Turkish, and only two groups of Russian speakers showed a higher frequency of perfective aspect, whereas Russian heritage speakers in Germany did not align with this pattern. The results of the Bayesian statistical analysis of the estimates revealed that there were no

meaningful effects for the Greek-speaking groups besides the increased production of perfective forms by the Greek heritage speakers in the US as compared to the grand mean [0.13, 0.31]. Russian heritage speakers in Germany opted for imperfective forms more likely than other Russian speaker groups [-0.22, -0.09]. In addition, all Russian-speaking groups tended to produce more imperfective forms in the spoken communicative situation compared to the written one [-0.19, -0.13]. For the Turkish-speaking groups, no meaningful effects of Country were observed, but at the level of formality, we found a small effect indicating that speakers produce more imperfective in the formal compared to the informal condition [-0.13, -0.04]. For mode, Turkish speakers tended to produce more imperfective forms in the spoken compared to the written mode [-0.12, -0.02]. Finally, a common meaningful effect in all three languages was achieved by the Tense factor. Specifically, more verbs bearing the perfective aspect were produced in the past tense than in the present, which can be explained by a task effect initiated by the stimulus question “What happened?”.

All in all, this cross-linguistic study offers an assessment of speakers’ use of grammatical aspect. Although the comparison involves languages that all encode perfective and imperfective aspect grammatically, we observed divergent results for different groups of speakers. This could be driven by language contact or typological differences. The majority languages English and German in contact with heritage languages Greek, Russian, and Turkish seem to affect the use of perfective and imperfective in different ways. Additionally, we found effects of mode and formality. Providing a better overview of the distribution of grammatical aspect among heritage speakers of different languages and different varieties of the same language is the most important contribution to the overall research question.

### **3 Periphrastic constructions in Greek by Alexiadou & Rizou (2023)**

Heritage varieties of Greek are claimed to show interesting dynamics in terms of aspectual formation and use. Several studies investigating verbal aspect show that this feature is challenging for Greek heritage speakers. The first scholar to identify discrepancies in the aspect category within heritage Greek was Seaman (1972), who noted a general simplification of verb forms and the application of periphrastic constructions among Greek heritage speakers residing in the US. Corroborating these findings, the study by Alexiadou & Rizou (2023) that is reviewed in this section investigates more systematically two groups of Greek heritage

speakers, namely one in the US and one in Germany, regarding the production of analytic forms. According to Haspelmath (2000), a periphrasis refers to the linguistic phenomenon where a specific idea is conveyed using a combination of words rather than a single term. An example for Greek is shown in (5), where the light verb like *do* and a noun are used instead of a lexical verb. In Greek, there is a morphological syncretism in light verbs meaning that the perfective and imperfective verbal stems are identical in contrast to full verbs.

- (5) kano katathesi vs. katathetho  
do testify vs. testify  
'conduct a testimony' vs. 'testify'

In the study reported here, the use of periphrastic structures employing light and/or auxiliary verbs to convey past events was explored. These constructions stand in opposition to the synthetic representation of such events, which frequently entails morpho-phonological alterations within the verbal system. Finally, this study controlled for extra-linguistic factors such as mode and formality.

The goal of the study was to detect whether the repertoires of different groups of heritage speakers differed in comparison to monolingually-raised speakers regarding the use of periphrastic constructions. Boon (2014) observed that heritage Welsh speakers resort to different strategies, such as code-mixing, in case they cannot retrieve morpho-phonological elements or lexical clusters. This has also been observed for heritage Greek by Alexiadou & Lohndal (2018). Specifically, analytic structures were assembled from Greek light verbs that lack certain morphological features, and a borrowed infinitival verbal form from the majority language. Alexiadou & Rizou (2023) investigated morphologically deviant patterns in heritage speakers' repertoires compared to monolinguals' productions, as well as emerging novel patterns and periphrastic constructions. Finally, the study explored the question of why heritage speakers make use of periphrastic constructions. What the authors proposed is that monolingually-raised Greek speakers utilize periphrasis to indicate different formality levels and distinguish between oral and written communicative settings, while heritage speakers tend to adopt forms linked with informal communicative settings, and this preference is influenced by the morpho-phonological intricacies of the formality of the communicative setting,<sup>7</sup> leading to a tendency of register levelling overall.

---

<sup>7</sup>The Greek lexicon encompasses components that are delineated along the [+/-learned] spectrum, a trait that directly corresponds to the high and low register and the diglossic context prevalent in the 19th and 20th century.

Methodologically, the narration task in the frame of *language situations* by Wiese (2020) was employed.<sup>8</sup> Two distinct age groups of Greek heritage speakers, namely adolescents and adults, residing in Germany and the US were tested. Additionally, a group comprising monolingually-raised speakers of similar age was included in the study. Their meta-linguistic data (self-ratings, mean age of onset, current input and literacy practices in the heritage language, hours and years of education in the heritage language, parents' generation) were collected in a form of a questionnaire and correlated with the use of periphrastic constructions.

The results showed a general preference for periphrastic constructions by Greek heritage speakers. These analytic constructions were analyzed in terms of appropriateness of use. This means that each construction was evaluated in terms of grammaticality and felicitousness in each context regarding the perfective/imperfective aspectual value. Only a few ambiguous and infelicitous cases were detected in heritage speakers' grammar. The most important finding is that heritage speakers use periphrasis to indicate the perfective aspect instead of using lexical verbs that require morpho-phonological changes as seen in example (6). As the light verb *do* in past tense has one identical stem for perfective and imperfective, speakers prefer to use this instead of lexical verbs that require morpho-phonological changes to form the perfective aspect.

- (6) ekanan bam vs. sigkrustikan  
did bam vs. collided  
'crashed' (attribution with the sound) vs. 'collided'

Another crucial tendency observed in the data is that heritage speakers utilized analytic forms interchangeably in the different formality settings, unlike monolingually-raised speakers who used them mainly in the informal setting. A common finding in both heritage and monolingually-raised speakers' narrations was that they both made use of analytic forms in the spoken mode across formality settings as an indicator of colloquial speech.

In sum, this study revealed some alternative strategies used by Greek heritage speakers for producing grammatical aspect. The investigation of analytic forms suggests a wider use of periphrastic constructions, pointing to a reorganization of heritage speakers' linguistic repertoire. Furthermore, the exploration of different communication settings highlights an emerging pattern, which is the use of periphrastic constructions across formality variation by heritage speakers in the

---

<sup>8</sup>More information can be found in Section 2. Detailed user guidelines and all experimental materials are available in an open-access repository: <https://osf.io/qhupg/>.

US. The contribution to the overarching research question is that it brings to the fore a strategy systematically used by Greek heritage speakers.

#### 4 Verbal aspect in heritage and monolingually-raised Greek speakers: Data from narration and production tasks by Rizou (2021)

Grammatical aspect in Greek can be either perfective or imperfective, as mentioned before. The perfective aspect in Greek denotes either completion or instantaneity of events, and it is marked on the verb stem, as we show below. The imperfective aspect is unmarked on the verb stem and corresponds to either continuous (example 7) or habitual (example 8) interpretation. Imperfective aspect has the continuous or the habitual interpretation depending on the context. The use of lexical cues like adverbs or adverbial phrases clarifies the relevant interpretation as shown in examples (7) and (8) (Moser 1994). As there is no aspectual distinction for the present tense in indicative mood in Greek, we illustrate this distinction by using past tense.

- (7) I Maria magireve      gia tris    ores.  
The Mary cook.IPFV.PST for three hours.  
'Mary was cooking for three hours.'
- (8) I Maria magireve      tis Kiriakes.  
The Mary cook.IPFV.PST on Sundays.  
'Mary used to cook on Sundays.'

The aim of the study by Rizou (2021) was to provide qualitative insights from the narration described in Section 2 and to measure the accuracy rate of Greek heritage speakers in an elicited production task as described below. Furthermore, the study investigated whether heritage speakers rely on the inherent semantics of the verbs as predicted by the Aspect Hypothesis by Andersen & Shirai (1994) and combine the perfective aspect with telic complements and the imperfective with atelic in semi-spontaneous production tasks. Additionally, the differences between two age groups, namely adults and adolescents, within each speaker group are also explored in this study.

The methodology employed for the narration task conforms with the *language situations* approach by Wiese (2020) as described in Section 2. To elicit verbal forms marked for grammatical aspect, an oral controlled production task in the form of sentence completion created by Agathopoulou & Papadopoulou (2009)

was used. This task consisted of thirty items in total, mapping the three conditions of grammatical aspect, namely ten sentences denoting the perfective aspect, ten the imperfective continuous interpretation, and ten the imperfective habitual.<sup>9</sup> The task was performed orally, meaning that the elicitor read the sentences out loud, leaving a 2-second pause in the place where the verb should be produced. Participants had to repeat the whole sentence with the answer, which was recorded.

Three groups of Greek speakers took part in this study, namely Greek monolingually-raised speakers and Greek heritage speakers in Germany and in the US; see Table 1. A further categorization was conducted for age groups, namely adults and adolescents within the three groups.<sup>10</sup>

The results from the qualitative analysis of the narrative data revealed patterns of perfectivization of verbs, mainly with atelic events and the verb class of activity verbs as shown in example (9) from the original publication. This pattern was observed mostly in the narrations of heritage speakers rather than in those of monolingually-raised speakers, and it was found in both formal and informal communication settings.

- (9) Itan ena zevgari pu \*perpatisan me ena karotsi.  
there was a couple who walk.PST.PVF with a stroller  
'There was a couple with a stroller that was walking.'

Moving to the controlled elicited production task, a one-way multivariate analysis of variance test (MANOVA) showed no differences between the three independent groups. The outcome revealed a significant difference between the monolingually-raised and heritage speakers in the US across aspectual conditions. A weak three-way interaction between the groups and the age groups was reported, and so adults and adolescents in the respective groups were analyzed separately. It was found that heritage adults in the US performed more accurately in the three conditions of aspect compared to adolescent heritage speakers in the US. In contrast, Greek adult heritage speakers in Germany performed worse compared to adolescent heritage speakers in Germany. There was no statistical difference between the age groups in the different conditions of aspect.

<sup>9</sup>There are 9 strong verbs that require an exponent /s/ to mark the aspectual feature, 8 verbs that exhibit weak allomorphy requiring the exponent /s/, and 4 verbs that exhibit only weak allomorphy with vowel alternation patterns. Finally, there is one verb with strong suppletion, requiring a different stem altogether. 22 out of 30 verbs belong to the first conjugation class, and 8 of them belong to the second conjugation class.

<sup>10</sup>Further information concerning the number of participants in each age group can be found in Rizou (2021: 41–42).

Summarizing, the overall contribution of the study is the exploration of different age groups, namely adult and adolescent heritage speakers' performance in an offline production task. Via the particular task, which has previously been used for different populations such as L2 speakers and bilingual children, the author obtained similar results to previous studies: the most challenging condition is the imperfective habitual one, for Greek heritage speakers both in Germany and in the US. The most important contribution, though, is the slight reverse pattern observed within and between the groups, namely the performance of adult and adolescent heritage speakers. This finding calls for future research on other factors affecting performance on verbal aspect (grammatical and lexical).

## 5 Novel forms for the expression of aspect in heritage Greek by Rizou et al. (2025)

Greek has a rich verbal morphology; each verb is marked for tense, aspect, voice and agreement. Recent analyses of the verbs in the first and second conjugation classes (CC) have been conducted by Spyropoulos et al. (2017) and Revithiadou et al. (2019). The first CC comprises verbs in which the stress falls on the root of IPFV [-PAST] forms like *graf-o* 'write', while the second CC includes all verbs that exhibit non-root stress for example *agap-ao/o* 'love' and *poth-o* 'desire' Holton et al. (1997: 126–136).

Focusing on the aspectual feature, the first pattern for the verbs of the first CC, in which the root ends either in a vowel or in a consonant in imperfective (e.g. *idri-o* 'establish'), is the addition of the exponent /s/ at the right edge of the root (e.g. (*na*) *idri-s-o* 'to establish'). These verbs are called regular/strong verbs in the literature (Spyropoulos et al. 2017, Revithiadou et al. 2019).

The second pattern concerns the irregular/weak roots, which undergo reshaping due to phonological adjustments, suppletion, and omission or addition of a consonant and syllables. There are two allomorphic vowel patterns of the same root, one for [-]PFV environments (e.g. *sern-o* 'drag') and one for [+PFV, (e.g. (*na*) *sir-o* 'to drag').<sup>11</sup>

The third pattern requires suppletion under which verbs undergo a complete stem change or strong suppletion, (e.g. *tro-o* 'eat' for [-]PFV and (*na*) *fa-o* 'to eat' for [+PFV environments), unlike the verbs that require weak suppletion. The verbs of the second CC exhibit two allomorphic stems in [-]PFV (e.g. *pul(a)-o* 'sell') and [+PFV aspect (e.g. (*na*) *puli-s-o* 'to sell').

---

<sup>11</sup>The /-n/ does not belong to the root and thus, the root vowel changes into /i/ when the coronal /-n/ disappears.

The study by Rizou et al. (2025) explored whether the productions of heritage speakers of Greek differed from those of monolingually-raised speakers regarding the encoding of several morpho-phonological features given the concatenate verbal morphology and, in more complex cases, the addition of prefixes. Specifically, the goal of the study was to reveal whether there were mismatches in verbal features such as aspect, tense, voice, *phi*-features (person and number), and whether there were any other asymmetries in morpho-phonological alternations like vocalic elements or stress placement.

Previous studies have shown that certain morphological features seem to be open to dynamic patterns in bilingual/heritage grammar across a variety of languages. One of the first researchers to mention differences in Russian heritage speakers' verbal morphology was Romanova (2008). Slabakova (2014) and Mikhaylova (2018) proposed the Bottleneck Hypothesis for L2 learners and heritage speakers, proposing that these speakers face difficulties with morphological features in languages with rich inflection. Furthermore, Perez-Cortes et al. (2019) and Fernández-Dobao & Herschensohn (2021) report differences in mood and tense for Spanish heritage speakers while Jóhannsdóttir (2023) reports an over-standardization of tense features in heritage Icelandic, meaning that the speakers resort to default strategies. It has been observed by Scontras et al. (2015) that heritage speakers often alter the structure of the verbal morphosyntactic system, resulting in the simplification of intricate morpho-phonemic rules.

The methodology employed in Rizou et al. (2025) is the same as that used by Rizou (2021), and was described in Section 4. In the elicited production task, there were twenty verbs of the first CC and ten of the second. Both strong verbs and verbs that require weak allomorphy were included in the experimental conditions. The participants who took part in this study were Greek monolingually-raised speakers recruited in Athens, Greece, and Greek heritage speakers recruited in Germany and in the US. They performed an elicited production task with fixed conditions for the production of perfective and imperfective aspect. More details can be found in the aforementioned publication.

The highlight of the analysis is that heritage speakers in the US produced more innovative verbal forms compared to heritage speakers in Germany. In Greek heritage speakers' productions, numerous patterns were detected pointing to a systematic restructuring of stem change, conjugation class change, stress and /s/ suffixation misplacement, and erroneous or dropped augment /e/ as seen in examples (10) categorized in the second CC. Another tendency was detected regarding the voice feature, indicating that heritage speakers make use of the productive rule proposed by Oikonomou & Alexiadou (2022) of employing the

non-active voice as opposed to the typical active voice, suggesting a potential generalization of the voice feature for the sake of morphological simplification.

(10)

- (11) \*(è)-for-e                  vs. fòra-g-e  
      augment-wear-3SG vs wear-IPVF-3SG  
      'was wearing'
- (12) pònak\*-s-e                  vs ponù-s-e  
      hurt-vocalic element-3SG vs hurt-IPFV-3SG  
      'was hurting'

Overall, morphological analyses of verbal features are rare in Greek heritage linguistics. This study contributes to filling this gap through a detailed analysis of dynamic verbal features in Greek heritage grammars. Insights on the features encoded in verbs are provided in this qualitative analysis, shedding light on this unexplored field and pointing the way to further studies on other phenomena that are encoded with different features in different languages.

## 6 Aspect acquisition and use in heritage Russian in Germany and the US by Gagarina et al. (2020)

Aspect production in adult heritage Russian speakers is often characterized by systematic patterns that differ from standard Russian (Pereltsvaig 2004, Laleko 2008). Besides, studies on child heritage Russian report innovations in the use and formation of perfective and imperfective verbs (Gagarina 2011, Antonova-Ünlü & Wei 2016, Kistanova & Sekerina 2019). One explanation for this dynamic linguistic behavior might be the missing sustainability hypothesis proposed by Gagarina et al. (2020); see Swain & Lapkin (1995) for a similar analysis. Specifically, in language contact situations, the aspectual system seems to undergo restructuring due to a lower amount of input in the heritage language in the daily communication of young children, even though the use of aspect is acquired early in Russian (Cejtin 2000, Gagarina 2008b, Kistanova & Sekerina 2019). Additionally, the dominance shift from the heritage language to the majority language, which usually occurs when children enter monolingual education institutions, further contributes to the dynamicity of the aspectual category in the linguistic system of an asymmetrically bilingual child.

The aim of Gagarina et al. (2020) was to investigate the productions of aspectual forms in heritage child and adult Russian in Germany and the US by contrasting them to the typical aspectual patterns of monolingual child Russian. More specifically, the study dealt with the question of whether the observed changes in the aspectual system of heritage Russian are unique for heritage speakers or whether they are typical for the acquisitional patterns of aspect in monolingual acquisition. Furthermore, the study attempted to gain insight into the degree of dynamicity of the aspectual system of heritage Russian, i.e., the degree of change in a given time period. To achieve this goal, the study applied the usage-based theory of language acquisition (Tomasello 2003) as well as frequency accounts of (first) language acquisition (Gülzow & Gagarina 2011, Lieven 2010).

The investigation was carried out on longitudinal spontaneous production data collected from two heritage bilingual children in Germany (Russian-German) in Germany and one heritage bilingual child (Russian-English) in the US (age range: 1;6–4;0 years) on the one hand, as well as on semi-spontaneous production data elicited from 128 adolescent and adult heritage Russian speakers (age range: 15–35 years) in Germany and the US on the other hand. Productions of perfective and imperfective verb forms of heritage speakers were compared with the productions of child monolingual speakers of Russian. Monolingual data were taken from studies by Bar-Shalom (2002), Kistanova & Sekerina (2019), Bondarko (2011), and Gagarina (2007). Adult and adolescent heritage data were collected according to the *language situations* approach by Wiese (2020), described earlier in Section 2. In the present study, only Russian heritage data were used, which were drawn from the 0.2.0 version of the RUEG-RU subcorpus (Wiese et al. 2019).<sup>12</sup>

The findings indicate that in young children at the age of around two years, aspect production is very similar in heritage and monolingual Russian. However, with increasing age starting already from three to four years, aspect use of heritage speakers occasionally shows dynamic patterns and innovations, which are not or rarely attested in productions of monolingual peers. These patterns often represent speakers' strong preference for perfective aspectual forms in analytical constructions, and also for the substitution of perfective aspectual forms by imperfective ones in denoting concrete resultative events. In example (13), the verb form *pojti* 'go' is perfective, which is an ungrammatical use in standard Russian. Instead, an imperfective form *idti* 'go' is canonically used here. In (14), the imperfective use of 'brake' is unexpected in the given context, in which the

---

<sup>12</sup>Please note that the number of heritage Russian speakers reported in Rizou et al. (2024) slightly differs from the numbers in the current study due to the different corpus versions used.

speaker was looking at the car driver braking suddenly to prevent an accident. The expected perfective verb form is *za-tormozil* ‘brake’.

- (13) ne xoču \***pojti** v školu  
not want go.PFV to school  
'(I) don't want to go to school'
- (14) voditel' #**tormozil**  
driver brake.IMPF.PST  
'The car driver was braking'

In adolescent and adult heritage speakers of Russian, those patterns consistently appear in language productions.

Taken together, the results of the study demonstrate that acquisitional patterns in bilingual and monolingually-raised children converge until the age of three, and then rare innovations and dynamic patterns emerge in productions of heritage speakers. Those patterns include the use of perfective aspectual forms and modal verbs in analytical constructions, as well as substitutions of perfective forms by imperfective ones. Such linguistic behavior in heritage speakers might be explained by the *missing sustainability hypothesis*, suggesting that the aspectual system undergoes restructuring caused by a low quantity of input in the heritage language during the daily communication to young children in situations of language contact.

## 7 Processing of verbal aspect in a Visual World eye-tracking paradigm by Özsoy, Büyükyıldırım, et al. (2023)

The study by Özsoy, Büyükyıldırım, et al. (2023) reviewed in this section was conducted with Turkish-speaking groups, and can be thought of as a conceptual replication of two processing studies by Minor et al. (2022) and Minor et al. (2023). Minor et al. (2022) investigated the processing of grammatical aspect in monolingual Russian speakers. Minor et al. (2023) extended this research to monolingual speakers of English and Spanish, comparing the results with the original data from Minor et al. (2022). Minor and colleagues performed an initial picture selection task and a Visual World eye-tracking study focusing on Russian aspectual prefixes, revealing that participants already demonstrated a preference prior to the verb's completion for target images corresponding to either ongoing

or completed events. This suggests incremental processing of grammatical aspect information on a word-internal scale.

The study reviewed here, Özsoy, Büyükyıldırım, et al. (2023), extended the aforementioned aspect processing paradigm to bilingual heritage speakers of Turkish. In doing so, this study aims to fill the gap in the literature on heritage Turkish concerning the comprehension of grammatical aspect and contributes to a paradigm shift towards online psycholinguistic investigations of heritage languages (Bayram et al. 2021). Specifically, the researchers aimed to investigate the extent to which the preference for the perfective aspect in processing completed events and the imperfective aspect for ongoing events would be displayed by Turkish heritage speakers in Germany and Turkish monolingually-raised speakers. They tested this distinction using two different suffixes in item pairs such as in (15) and (16).

- (15) Temiz gömleğini ütule-di      yaşılı amca özene bezene.  
clean shirt      iron-PST.PFV old uncle properly  
'The elderly uncle carefully ironed his clean shirt.'
- (16) Temiz gömleğini ütülü-yor-du      yaşılı amca özene bezene.  
clean shirt      iron-IMPFV.PROG-PST old uncle properly  
'The elderly uncle was carefully ironing his clean shirt.'

One of the questions that the study explored is whether aspectual distinctions would be recognized by heritage speakers of Turkish. These speakers have shown reduced sensitivity to TAM morphology in previous studies (Arslan et al. 2017, Coşkun Kunduz 2018), which could lead to a reduced effect (i.e., slower processing of aspectual distinctions) in the processing patterns of heritage speakers of Turkish when compared to monolingually-raised speakers. Another related research question is whether aspectual distinctions would be processed incrementally as the sentence unfolds. Additionally, the study explored driving factors of individual differences in processing by examining individual cognitive skills (processing speed) and language proficiency (C-test scores).

A statistical analysis of the eye-tracking results focused on eye gaze preferences and the accuracy of picture selection. With regard to the overarching research question of this review chapter, the results indicate that heritage speakers recognize aspectual distinctions between the perfective and the imperfective aspect. The results also suggested that heritage speakers and monolingual speakers of Turkish indeed processed aspectual distinctions incrementally (see also Büyükyıldırım et al. in preparation). However, the researchers found that incremental processing is delayed in heritage speakers compared to monolingually-

raised speakers. The effect of the incremental processing of aspectual distinctions arises shortly after the offset of the verb and most closely patterns with the findings on Spanish, which seems to be due to similar morphological characteristics. Both Turkish and Spanish mark grammatical aspect using suffixes after the verb stem.

Additionally, within the group of heritage speakers, the findings showed small significant effects of individual cognitive abilities (processing speed) and language proficiency, indicating that speakers with higher cognitive abilities and language proficiency recognized aspectual distinctions faster. An effect of language proficiency was also observed in the results of the picture selection task, showing that speakers with higher C-test scores reached higher accuracy.

Overall, this study replicates the findings of Minor et al. (2022, 2023) for Turkish, which is typologically different from languages in the Indo-European language family, such as English, Russian, and Spanish, which were investigated by the aforementioned researchers. Considering the aim to extend (aspect) processing studies to bilingual populations (Bayram et al. 2021), this study took a successful first step. Turkish heritage speakers were able to process aspectual marking and its distinctions incrementally. These findings contrast with previous studies, such as those by Karaca et al. (2024) and Özsoy, Çiçek, et al. (2023), which found that heritage speakers show no or only marginal effects in processing grammatical factors such as case and aspect. This new study demonstrates that heritage speakers do engage in grammatical processing and this is modulated by individual differences factors. The authors offer valuable insights into the interplay of linguistic and cognitive factors in the processing of aspectual distinctions among Turkish heritage speakers, shedding light on potential effects of heritage language bilingualism and accounting for individual differences.

## 8 Discussion

This chapter provides reviews of six studies focusing on aspect in heritage and monolingual varieties of Greek, Russian, and Turkish that were conducted as part of the Research Unit *Emerging Grammars*. The cross-linguistic study evaluates the use of perfective and imperfective aspectual forms among the aforementioned population in narrative tasks, while the rest of the studies provide different foci on the exploration of grammatical aspect in each language separately analyzed with different methods. Some of the studies investigate the influence of formality and mode variation on the expression of aspect forms as well. Analyzing large-scale data for each language with different statistical methods, we

reveal interesting insights into the dynamicity of phenomena related to grammatical aspect that undergo systematic changes. Different factors that affect the acquisition, the preservation, and the production of aspect are analyzed in each study, offering different perspectives regarding the phenomenon of aspect.

Aspect was one of the first phenomena addressed in heritage language research (Montrul 2002, Kagan 2005, Laleko 2008). However, no previous research has used a comparative approach to explore the realization of aspect in heritage speakers of typologically distinct languages. The comparative approach in Rizou et al. (2024) reveals that heritage speakers are influenced by the morphological marking of aspect in the majority language in language contact situations. We expected the notion of morphological markedness as described by Comrie (1976) to affect heritage speakers' preferences with regard to grammatical aspect. On the one hand, we expected the speakers of the heritage languages in contact with majority German to opt for the simpler and morphologically unmarked forms, as German does not have grammatical aspect; on the other hand, we expected heritage speakers to opt for the marked forms due to contact with majority English, which marks grammatical aspect. These predictions were confirmed for Russian heritage speakers in Germany and for Greek heritage speakers residing in the US, respectively. The findings speak in favor of the markedness account (Comrie 1976), although it is not uniform for all languages under the scope of this study. For Turkish, no group effects were found, which is probably connected to the typologically different way of aspect marking, which is provided by the prominent markers *-iyor* and *-(y)di*. As the study by Özsoy, Büyükyıldırım, et al. (2023) revealed, the processing of the aspectual distinctions begins shortly after the offset of the verb and is incremental. Although heritage speakers perform slower compared to monolingually-raised speakers, this does not seem to be related to the salient aspectual markers in Turkish and rather might be explained by language use effects.

Another factor considered in the studies by Rizou et al. (2024) and Alexiadou & Rizou (2023) is the communicative setting. The factors of formality and mode variation affect the realization of grammatical aspect in different ways in the languages discussed in this chapter. By mixing various levels of formality and mode, one can get varied results across different languages, as demonstrated in the study by Pescuma et al. (2023). Thus, in the study by Rizou et al. (2024), mode effects were detected in both Turkish and Russian speakers, who tended to produce more imperfectives in the spoken mode. Additionally, formality effects were reported only for Turkish, manifested in more imperfective forms in the formal setting. In the study by Alexiadou & Rizou (2023) it was found that both heritage and monolingually-raised speakers clearly favor periphrastic constructions in

the oral mode, but there was a tendency by Greek heritage speakers in the US to use these analytic structures more often than monolingually-raised speakers. Since there are no studies correlating the different formality levels and modes with the phenomenon of grammatical aspect, we approached mode and formality variation in an exploratory way. The non-uniformity of the results regarding the factors of formality and mode in the different languages is probably due to the fact that extra-linguistic situational-functional parameters which influence intra-individual variation (cf. Lüdeling et al. 2022) do not uniformly determine the distribution of linguistic means across different language communities, even if the phenomena involved are comparable.

A further factor taken into consideration in the studies investigating verbal aspect is the age factor. In the study by Rizou (2021), different communities of Greek heritage speakers exhibit a divergent performance on production tasks concerning the phenomenon of verbal aspect. The Greek adolescent heritage group in Germany performed more accurately in the task compared to the adult group. However, the opposite pattern was detected in the US: Greek adult heritage speakers performed more accurately compared to the adolescent heritage group. The explanation of this finding might be hidden in community characteristics such as past and current input and literacy practices, which seem to be different in heritage Greek communities in Germany and the US.

The study by Gagarina et al. (2020) provides an overview of the use of aspect in L1 and heritage Russian, spanning from the initial occurrence of verb form production in child L1 to adulthood. The findings indicate a resemblance in the early acquisition and production of aspectual forms between monolinguals and bilinguals, while notable differences in aspect use between the two populations emerge systematically around the age of three to four. Differences in aspect usage from monolingually-raised speakers, such as the substitution of perfective with imperfective in analytical future constructions, as well as the use of imperfective in resultative contexts by heritage Russian speakers, are attributed to the lack of a sustainable input. In their adulthood, even fluent and proficient Russian heritage speakers overlook formal restrictions aligning with the prototypical semantics of aspect. These systematic patterns are in line with previous literature reporting a spread of non-canonical patterns in the aspectual system of Russian heritage speakers (Polinsky 2000).

The study by Rizou et al. (2025) highlights the challenges that heritage speakers face in encoding all verbal features within and outside the stem boundaries in a language with concatenate morphology, such as Greek. The features most susceptible to change seem to be the  $\phi$ -features and aspect. At the same time,

stem alternations that involve an allomorphic vocalic element and morpho-phonological adjustments seem to pose difficulties for heritage speakers' productions even in experiments without time constraints. Thus, heritage speakers resort to morphologically novel patterns with simplified and even default verbal features. This is in line with the theory of representational economy proposed by Scontras et al. (2015). Another strategy followed by heritage speakers is revealed in the study by Alexiadou & Rizou (2023), which is a preference for analytic verbal forms. This periphrasis contains a light verb and either a noun or a code-switched phrase. Heritage speakers tend to resort to morphologically less complex verbs, as is the case with light verbs. Thus, they avoid difficulties with morpho-phonological changes that come with aspect formation of lexical verbs.

The aspectual preference observed in our speakers is not determined solely by one factor; rather, it is influenced by a complex interaction of both language-internal and language-external factors in addition to other background factors. The findings suggest that internal grammatical properties of languages in a contact situation, such as the typology of each language, affect bilingual speakers' selection of aspectual forms in language contact situations.

This review chapter confirms the necessity of adopting a multifaceted approach to elucidate phenomena of the core grammatical system, like grammatical aspect. The methodologies and analyses outlined here underscore the importance of conducting further research to comprehensively explore grammatical aspect across various languages and populations in diverse situational contexts. By conducting experiments that account for the aforementioned factors, researchers can assess whether and to which degree these factors impact the processing and production of grammatical aspect.

## Acknowledgements

We would like to thank everyone involved in projects P1, P3 and P10 of RUEG. This work has grown out of a big team effort, with many student assistants spending hundreds of hours with detailed annotations. For this reason, we would like to thank all of them: Foteini-Maria Karkaletsou, Nikolaos Tsokanos, Ioanna Kolokytha, Iro Malta, Katerina Chatzitheoxari, Olga Kritharidou, Foteini Iliadou, Tsampika Psilou, Panagiota Papavasileiou, Jannis Spiekermann, Olga Buchmüller, Andrei Koniaevel, Anastasia Rozova, Tony Müller, Yelizaveta Vlasova, Alexandra König, Gvantsa Rukhiashvili, Nina Bredereck, Daria Alkhimchenkova, Zeynep Özal, and Büşra Çiçek. We also thank Kalliopi Katsika, Cem Keskin, Kateryna

Iefremenko, and two anonymous reviewers as well as audiences at the WILA 14 and the ISB 14 conferences for their helpful and constructive suggestions. We also thank the RUEG Team and RUEG's Mercator Fellows for their insightful comments. Finally, we thank Lea Coy for help with pre-publication formatting.

The research was supported through funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit *Emerging Grammars in Language Contact Situations*, projects P1 (no. 394836232), P3 (no. 394839191) and P10 (no. 313607803).

## References

- Acarlar, Funda, Ayhan Aksu-Koç, Aylin C. Küntay, İlknur Maviş, Hatice Sofu, Seyhun Topbaş & Figen Turan. 2008. Adapting MB-CDI to Turkish: The first phase. In Sila Ay, Özgür Aydin, Iclal Ergenc, Seda Gökmen, Selcuk Issever & Dilek Pecenek (eds.), *Essays on Turkish linguistics: Proceedings of the 14th International Conference on Turkish Linguistics* (Turcologica 79), 313–320. Wiesbaden: Harrassowitz.
- Agathopoulou, Eleni & Despina Papadopoulou. 2009. Morphological dissociations in the L2 acquisition of an inflectionally rich language. *EuroSLA Yearbook* 9(1). 107–131. DOI: 10.1075/eurosla.9.07aga.
- Aksu-Koç, Ayhan. 1988. *The acquisition of aspect and modality: The case of past reference in Turkish* (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511554353.
- Alexiadou, Artemis, Giana Iordăchioia & Elena Soare. 2010. Number/aspect interactions in the syntax of nominalizations: A distributed approach. *Journal of Linguistics* 46(3). 537–574. DOI: 10.1017/S002226710000058.
- Alexiadou, Artemis & Terje Lohndal. 2018. Units of language mixing: A cross-linguistic perspective. *Frontiers in Psychology* 9. 1719. DOI: 10.3389/fpsyg.2018.01719.
- Alexiadou, Artemis & Vasiliki Rizou. 2023. The use of periphrasis for the expression of aspect by Greek heritage speakers: A case study of register variation narrowing. *Register Studies* 5(1). 82–110. DOI: 10.1075/rs.20022.ale.
- Andersen, Roger W. & Yasuhiro Shirai. 1994. Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition* 16(2). 133–156.
- Antonova-Ünlü, Elena & Li Wei. 2016. Aspect acquisition in Russian as the weaker language: Evidence from a Turkish–Russian child. *International Journal of Bilingualism* 20(2). 210–228. DOI: 10.1177/1367006914552.

- Arslan, Seçkin, Doerte De Kok & Roelien Bastiaanse. 2017. Processing grammatical evidentiality and time reference in Turkish heritage and monolingual speakers. *Bilingualism: Language and Cognition* 20(3). 457–472.
- Bar-Shalom, Eva. 2002. Tense and aspect in early child Russian. *Language Acquisition* 10(4). 321–337. DOI: 10.1207/S15327817LA1004\_2.
- Bayram, Fatih, Grazia Di Pisa, Jason Rothman & Roumyana Slabakova. 2021. Current trends and emerging methodologies in charting heritage language grammars. In Silvina Montrul & Maria Polinsky (eds.), *The Cambridge handbook of heritage languages and linguistics* (Cambridge Handbooks in Language and Linguistics), 545–578. Cambridge: Cambridge University Press. DOI: 10.1017/9781108766340.025.
- Benmamoun, Elabbas, Silvina Montrul & Maria Polinsky. 2013. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3-4). 129–181. DOI: 10.1515/tl-2013-0009.
- Bondarko, Aleksandr V. 2011. *Kategorizacija v sisteme grammatiki [Categorisation of grammatical systems]*. Moscow: Jazyki slavjanskich kul'tur.
- Boon, Erin Diane. 2014. *Heritage Welsh: A study of heritage language as the outcome of minority language acquisition and bilingualism*. Harvard University. (Doctoral dissertation).
- Büyükyıldırım, Nisa, Özce Özceçelik, Serkan Uygun & Onur Özsoy. In preparation. *Verb aspect processing and production in mono- and bilingual Turkish speakers*.
- Cejtlin, Stella N. 2000. *Jazyk i rebenok: Lingvistika detskoj reči* [The language and the child: Language acquisition]. Vlados.
- Chevalier, Joan F. 2004. Heritage language literacy: Theory and practice. *Heritage Language Journal* 2(1). 26–44.
- Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Coşkun Kunduz, Aylin. 2018. *The second language acquisition of inflectional morphology in Turkish*. Boğaziçi University. (Doctoral dissertation).
- Cuza, Alejandro, Rocío Pérez-Tattam, Elizabeth Barajas, Lauren Miller & Claudia Sadowski. 2013. The development of tense and aspect morphology in child and adult heritage speakers. In John W. Schwieter (ed.), *Innovative research and practices in second language acquisition and bilingualism* (Language Learning & Language Teaching 38), 193–220. Amsterdam: John Benjamins. DOI: 10.1075/llt.38.12cuz.
- Delidakis, Sophia. 2006. *The acquisition of tense and aspect in child Greek*. University of Reading. (Doctoral dissertation).

- Dressler, Wolfgang U. 1991. The sociolinguistic and patholinguistic attrition of Breton phonology, morphology, and morphonology. In Herbert W. Seliger & Robert M. Vago (eds.), *First language attrition*, 99–112. Cambridge: Cambridge University Press.
- Fernández-Dobao, Ana & Julia Herschensohn. 2021. Acquisition of Spanish verbal morphology by child bilinguals: Overregularization by heritage speakers and second language learners. *Bilingualism: Language and Cognition* 24(1). 56–68. DOI: 10.1017/S1366728920000310.
- Gagarina, Natalia. 2000. The acquisition of aspectuality by Russian children: The early stages. *ZAS Papers in Linguistics* 15. 232–246.
- Gagarina, Natalia. 2007. Stanovlenie funkciona'l'no-semantičeskogo polja aspektual'nosti v detskoj reči [The functional-semantic field of aspect in language acquisition]. In Yana Akhapkina, Maria Voeikova & Stella N. Cejtin (eds.), *Semantičeskie kategorii v detskoj reči* [Semantic categories in language acquisition], 18–40. Nestor-Istorija.
- Gagarina, Natalia. 2008a. Anaphoric pronominal reference in Russian and German narratives: Bilingual and monolingual settings. *Zeitschrift für Slawistik* 53(3). 326–338.
- Gagarina, Natalia. 2008b. *First language acquisition of verb categories in Russian*. St. Petersburg: Nauka.
- Gagarina, Natalia. 2011. Acquisition and loss of L1 in a Russian-German bilingual child: A case study. In S. Cejtin (ed.), *Monolingual and bilingual path to language*, 137–163. Moscow: Jazyki slavjanskoj kul'tury.
- Gagarina, Natalia, Maria Martynova, Irina A. Sekerina & Luka Szucsich. 2020. From first verbs to adulthood: Aspect in heritage Russian in the US and Germany. In Yevgeny Golovko, Elena Gorbova, Petr Kocharov, Viktor Khrakovskiy & Oksana Chuikova (eds.), *Vzaimodejstvie aspekta so smežnymi kategorijami* [Interaction of aspect with related categories], 112–117. St. Petersburg: RGPU imeni A. I. Gercena.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari & Donald B. Rubin. 2013. *Bayesian data analysis*. CRC press.
- Guijarro-Fuentes, Pedro & Katrin Schmitz. 2015. The nature and nurture of heritage language acquisition. *Lingua* 164. 239–250. DOI: 10.1016/j.lingua.2015.05.008.
- Gülzow, Insa & Natalia Gagarina. 2011. *Frequency effects in language acquisition: Defining the limits of frequency as an explanatory concept* (Studies on Language Acquisition 32). Walter de Gruyter. DOI: 10.1515/9783110977905.

- Haspelmath, Martin. 2000. Periphrasis. In G. E. Booij, C. Lehmann & J. Mugdan (eds.), *Morphology: A handbook on inflection and word formation*, 654–664. Berlin: de Gruyter. DOI: 10.1515/978311011286.1.9.654.
- Holton, David, Peter Mackridge & Irene Philippaki-Warburton. 1997. *Greek: A comprehensive grammar of the modern language*. Psychology Press.
- Jóhannsdóttir, Kristín M. 2023. “And the dog is sleeping too”: The use of the progressive in North American Icelandic. In Birna Arnþjörnsdóttir, Höskuldur Þráinsson & Úlfar Bragason (eds.), *Icelandic Heritage in North America*, 252–270. Manitoba: Univ. of Manitoba Press.
- Kagan, Olga. 2005. In support of a proficiency-based definition of heritage language learners: The case of Russian. *International Journal of Bilingual Education and Bilingualism* 8(2-3). 213–221. DOI: 10.1080/13670050508668608.
- Karaca, Figen, Susanne Brouwer, Sharon Unsworth & Falk Huettig. 2024. Morphosyntactic predictive processing in adult heritage speakers: Effects of cue availability and spoken and written language experience. *Language, Cognition and Neuroscience* 39(1). 118–135. DOI: 10.1080/23273798.2023.2254424.
- Kistanova, Ekaterina V. & Irina A. Sekerina. 2019. Acquisition of the grammatical categories of Russian verbs in a heritage Russian-English child: A case study. In Megan M. Brown & Brady Dailey (eds.), *Proceedings of the 43rd Boston University Conference on Language Development*, 324–337. Cascadilla Press.
- Kiyota, Masaru. 2008. *Situation aspect and viewpoint aspect: From Salish to Japanese*. University of British Columbia. (Doctoral dissertation).
- Konstantzou, Katerina, Angeliek Van Hout, Spyridoula Varlokosta & Maria Vlasopoulos. 2013. Perfective–imperfective: Development of aspectual distinctions in Greek specific language impairment. *Linguistic Variation* 13(2). 187–216. DOI: 10.1075/lv.13.2.03kon.
- Kruschke, John. 2014. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Kruschke, John & Wolf Vanpaemel. 2015. Bayesian estimation in hierarchical models. In Jerome R. Busemeyer, Zheng Wang, James T. Townsend & Ami Eidels (eds.), *The Oxford handbook of computational and mathematical psychology*, 279–299. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199957996.013.13.
- Laleko, Oksana. 2008. Compositional telicity and heritage Russian aspect. In *Proceedings of the thirty-eighth Western Conference on Linguistics (WECOL)*, 150–160. Davis, CA: Department of Linguistics, University of California, Davis.
- Laleko, Oksana. 2010. *The syntax-pragmatics interface in language loss: Covert restructuring of aspect in heritage Russian*. University of Minnesota. (Doctoral dissertation). <https://hdl.handle.net/11299/92219>.

- Lieven, Elena. 2010. Input and first language acquisition: Evaluating the role of frequency. *Lingua* 120(11). 2546–2556. DOI: 10.1016/j.lingua.2010.06.005.
- Lüdeling, Anke, Artemis Alexiadou, Aria Adli, Karin Donhauser, Malte Dreyer, Markus Egg, Anna Helene Feulner, Natalia Gagarina, Wolfgang Hock, Stefanie Jannedy, Frank Kammerzell, Pia Knoeferle, Thomas Krause, Manfred Krifka, Silvia Kutscher, Beate Lütke, Thomas McFadden, Roland Meyer, Christine Mooshammer, Stefan Müller, Katja Maquate, Muriel Norde, Uli Sauerland, Stephanie Solt, Luka Szucsich, Elisabeth Verhoeven, Richard Waltereit, Anne Wolfsgruber & Lars Erik Zeige. 2022. Register: Language users' knowledge of situational-functional variation. *Register Aspects of Language in Situation* 1(1). 1–58. DOI: 10.18452/24901.
- Mikhaylova, Anna. 2018. Morphological bottleneck: The case of Russian heritage speakers. *Journal of Language Contact* 11(2). 268–303.
- Minor, Serge, Natalia Mitrofanova, Gustavo Guajardo, Myrte Vos & Gillian Ramchand. 2023. Aspect processing across languages: A visual world eye-tracking study. *Frontiers in Language Sciences* 1. 1052205. DOI: 10.3389/flang.2022.1052205.
- Minor, Serge, Natalia Mitrofanova & Gillian Ramchand. 2022. Fine-grained time course of verb aspect processing. *PLoS ONE* 17(2). e0264132. DOI: 10.1371/journal.pone.0264132.
- Montrul, Silvina. 2002. Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition* 5(1). 39–68. DOI: 10.1017/S1366728902000135.
- Montrul, Silvina. 2015. *The acquisition of heritage languages*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139030502.
- Montrul, Silvina & Silvia Perpiñán. 2011. Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. *Heritage Language Journal* 8(1). 90–133. DOI: 10.46538/hlj.8.1.5.
- Moser, Amalia. 1994. The interaction of lexical and grammatical aspect in Modern Greek. In Irene Philippaki-Warburton, Katerina Nicolaïdis & Maria Sifianou (eds.), *Themes in Greek linguistics*, 137–144. Amsterdam: John Benjamins. DOI: 10.1075/cilt.117.20mos.
- Oikonomou, Despina & Artemis Alexiadou. 2022. Voice syncretism crosslinguistically: The view from minimalism. *Philosophies* 7(1). 19. DOI: 10.3390/philosophies7010019.
- Özsoy, Onur, Nisa Büyükyıldırım, Özce Özceçelik, Ulaş Aşkın & Serkan Uygun. 2023. Verb aspect processing and production in mono- and bilingual Turkish.

- In *Words in the World conference*. Poster. <https://wordsintheworld.ca/wow-conference/conference-schedule-wow2023/>.
- Özsoy, Onur, Büsra Çiçek, Zeynep Özal, Natalia Gagarina & Irina A. Sekerina. 2023. Turkish-German heritage speakers' predictive use of case: Webcam-based vs. in-lab eye-tracking. *Frontiers in Psychology* 14. 1155585. DOI: 10.3389/fpsyg.2023.1155585.
- Panitsa, Georgia. 2010. *Aspects of Aspect: The acquisition of viewpoint and situation aspect in Modern Greek*. University College London. (Doctoral dissertation).
- Pereltsvaig, Asya. 2004. Immigrant Russian: Factors in the restructuring of the aspectual system under attrition. In Corey Yoquelet (ed.), *Proceedings of the 29th Annual Meeting of the Berkeley Linguistics Society*, 107–119. Berkeley: Berkeley Linguistics Society.
- Perez-Cortes, Silvia, Michael T Putnam & Liliana Sánchez. 2019. Differential access: Asymmetries in accessing features and building representations in heritage language grammars. *Languages* 4(4). 81. DOI: 10.3390/languages4040081.
- Pescuma, Valentina N, Dina Serova, Julia Lukassek, Antje Sauermann, Roland Schäfer, Aria Adli, Felix Bildhauer, Markus Egg, Kristina Hülk, Aine Ito, et al. 2023. Situating language register across the ages, languages, modalities, and cultural aspects: Evidence from complementary methods. *Frontiers in Psychology* 13. 964658. DOI: 10.3389/fpsyg.2022.964658.
- Polinsky, Maria. 2000. A composite linguistic profile of a speaker of Russian in the US. In Olga Kagan & Benjamin Rifkin (eds.), *The learning and teaching of Slavic languages and cultures*, 437–465. Bloomington, IN: Slavica.
- Polinsky, Maria. 2015. Heritage languages and their speakers: State of the field, challenges, perspectives for future work, and methodologies. *Zeitschrift für Fremdsprachwissenschaft* 26. 7–27.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- R Core Team. 2021. *R: A Language and environment for statistical computing*. Version 4.3.0. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Revithiadou, Anthi, Giorgos Markopoulos & Vassilios Spyropoulos. 2019. Changing shape according to strength: Evidence from root allomorphy in Greek. *The Linguistic Review* 36(3). 553–574. DOI: 10.1515/tlr-2019-2029.
- Rizou, Vasiliki. 2021. Verbal aspect in heritage Greek across majority languages: A preliminary study. In Arnstein Hjelde & Åshild Søfteland (eds.), *Selected pro-*

- ceedings of the 10th Workshop on Immigrant Languages in the Americas (WILA 10), 39–48. Sommerville, MA: Cascadilla Proceedings Project.
- Rizou, Vasiliki, Onur Özsoy, Maria Martynova, Luka Szucsich, Artemis Alexiadou & Natalia Gagarina. 2024. Grammatical aspect in heritage and monolingual Greek, Russian and Turkish. *Register Aspects of Language in Situation* 3(3). DOI: 10.18452/29345.
- Rizou, Vasiliki, Despina Papadopoulou & Artemis Alexiadou. 2025. *Novel forms for the expression of aspect in heritage Greek*.
- Romanova, Natalia. 2008. Mechanisms of verbal morphology processing in heritage speakers of Russian. *Heritage Language Journal* 6(1). 105–127. DOI: 10.46538/hlj.6.1.5.
- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2). 155–163. DOI: 10.1177/1367006909339814.
- Rothman, Jason, Fatih Bayram, Vincent DeLuca, Grazia Di Pisa, Jon Andoni Duñabeitia, Khadij Gharibi, Jiuzhou Hao, Nadine Kolb, Maki Kubota, Tanja Kupisch, Tim Laméris, Alicia Luque, Brechje van Osch, Sergio Miguel Pereira Soares, Yanina Prystauka, Deniz Tat, Aleksandra Tomić, Toms Voits & Stefanie Wulff. 2023. Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics* 44(3). 316–329. DOI: 10.1017/S0142716422000315.
- Sauer, Simon & Anke Lüdeling. 2016. Flexible multi-layer spoken dialogue corpora. *International Journal of Corpus Linguistics* 21(3). 419–438. DOI: 10.1075/ijcl.21.3.06sau.
- Schmid, Helmut. 2008. Tokenizing and part-of-speech tagging. In Anke Lüdeling & Merja Kyölä (eds.), *Corpus linguistics: An international handbook*, vol. 1, 527–551. Berlin: Walter de Gruyter.
- Schmidt, Thomas & Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *Handbook on corpus phonology*, 402–419. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199571932.013.030.
- Scontras, Gregory, Zuzanna Fuchs & Maria Polinsky. 2015. Heritage language and linguistic theory. *Frontiers in Psychology* 6. 1545. DOI: 10.3389/fpsyg.2015.01545.
- Seaman, P. David. 1972. *Modern Greek and American English in contact*. The Hague/Paris: Mouton.
- Slabakova, Roumyana. 2014. The bottleneck of second language acquisition. *Foreign Language Teaching and Research* 46(4). 543–559.

- Spyropoulos, Vassilios, Anthi Revithiadou & Giorgos Markopoulos. 2017. *Root strength as (under)specification: Evidence from root allomorphy in Greek*. Talk presented at Strength in Grammar workshop, November 10–11, 2017. Leipzig University.
- Stephany, Ursula. 1997. The acquisition of Greek. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition*, 183–333. Hillsdale, NJ: Erlbaum.
- Swain, Merrill & Sharon Lapkin. 1995. Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics* 16(3). 371–391.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Valdés, Guadelupe. 2005. Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? *The Modern Language Journal* 89(3). 410–426. DOI: 10.1111/j.1540-4781.2005.00314.x.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo & Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). DOI: 10.21105/joss.01686.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Anke Lüdeling, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Tracy Rosemarie, Christoph Schroeder, Luka Szucsich, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2019. *RUEG Corpus*. Version 0.2.0. Zenodo. DOI: 10.5281/zenodo.3236069.



# Chapter 8

## Clause combining in narrative discourse: A contrastive analysis across heritage and majority languages

Christoph Schroeder<sup>a</sup>, Kateryna Iefremenko<sup>a,b</sup>, Kalliopi Katsika<sup>c</sup>, Annika Labrenz<sup>d</sup> & Shanley E. M. Allen<sup>c</sup>

<sup>a</sup>University of Potsdam <sup>b</sup>Leibniz-Centre General Linguistics <sup>c</sup>University of Kaiserslautern-Landau <sup>d</sup>Humboldt-Universität zu Berlin

In this chapter, we analyze how monolingual and bilingual speakers combine clauses in narratives. We investigate both languages of bilingual speakers, that is, heritage and majority languages of speakers in Germany and the US with Russian, Greek or Turkish as heritage languages and German or English as majority languages. We ask to what extent differences and similarities between the varieties can be related i) to the typological structures of the languages, ii) to communicative situations within which the narratives were produced, and iii) to language contact. Our findings indicate that the typological differences between the languages in question regarding preferences for subordination strategies are preserved also in the language contact settings. Furthermore, we do not find much evidence of language contact effects in either the majority or heritage languages. However, the findings suggest that communicative situations play a crucial role. In instances where differences between heritage and majority varieties of the investigated languages are observed, they arise mainly due to register levelling in the heritage languages of heritage speakers.

### 1 Introduction

In this chapter, we ask how monolingual and bilingual speakers, with different linguistic repertoires, combine clauses in narrations. We analyze both languages



of bilingual speakers, that is, heritage and majority languages of speakers in Germany and the US with Russian, Greek or Turkish as heritage languages and German or English as majority languages. Generally speaking, distributions of clause combining strategies in a given language are co-determined by several parameters. On the one hand, typological properties of the language provide a predisposition in a certain direction; for example, a language prefers either finite or non-finite subordination based on its typological structure (see Section 1.1 for typological discussion and Section 1.2 for the strategies of the languages in question here). At the same time, preferences for clause combining strategies emerge in different varieties of the language, and this includes, quite centrally for us, varieties arising from situational parameters such as (in)formality and mode (spoken/written) (see Section 1.3). We furthermore assume that the choice of text type (argumentation, narration, description, report, ...) and possibly also speech acts within text types (descriptive, evaluative, text organizing, ...) affect clause combining strategies. Therefore, we concentrate on one particular text type: narrations (see Section 1.4). And last but not least, in our data, the bilingualism of speakers has to be considered as a specific constellation which may have an impact on the strategies used, through language contact or otherwise (see Section 1.5).

The data we analyze were collected using the same method across languages and speaker groups (see Section 2), and were controlled for (in)formality and mode. This allows us to investigate the factors listed above. Hence, in the chapter we explore whether there are similar dynamic patterns in clause combining strategies among the speakers of the five languages under research. Our research question is the following: To what extent are potential differences and similarities between the varieties related to i) the typological structures of the languages, ii) communicative situations (comm-sits) of the narratives, and iii) language contact. Further elaboration on each of these points will be provided below.

Our study is exploratory in nature. That is, we aim to discuss tendencies that we observe based on the analysis presented in figures. We believe that observations will offer a more holistic understanding of the issue under investigation and identified patterns can later be tested more rigorously with statistical methods.

## **1.1 Structural and typological issues of clause combining**

In this chapter, we adopt a broad notion of clause in the sense of the use of the term by Lehmann (1988: 182); in other words, any syntagma containing a predication is a clause. The minimal requirement for this is a verbal form (except

for nominal clauses). Our terminology with respect to the structural aspects of clause combining is as follows:

1. *Independent clauses* are clauses that are syntactically independent.
2. *Matrix clauses* are clauses which embed a subordinate clause.
3. *Subordinate clauses* are clauses that are dependent on a matrix clause in the sense that they are embedded in it, that is, they form a constituent within a matrix clause. Subordinate clauses can be finite or non-finite, and they can be reduced, meaning that arguments of the verb, further optional complements, and/or aspecto-temporal or other clausal properties are suppressed.
4. A *finite clause* is one with a finite verb (i.e., a verb that has verbal agreement morphology). A non-finite clause is a clause formed with a non-finite verb (i.e., a verb that has no verbal agreement morphology).<sup>1</sup>
5. *Extended clauses* consist of a matrix clause and one or more subordinate clauses; simple clauses consist of only an independent clause.
6. Clause combining strategies can be either *hypotactic* (subordinated) or *paratactic* (coordinated).<sup>2</sup>
7. Paratactic clause combining can be *syndetic* (i.e., by means of an overt connector) or *asyndetic* (i.e., without an overt connector).

Languages typically make use of a spectrum of clause combining strategies (see, for example, the typological studies by Haspelmath (2004)). In principle, we understand the distribution of clause combining strategies as ordered on continua, based on Lehmann (1988). Lehmann develops three continua of clause linkage (or, as we mostly call it here, clause combining). These are as follows: i) the continuum of integration of one clause into the other, ii) the continuum of the expansion vs. reduction of the clauses, and iii) the continuum of the mutual isolation vs. linkage of the two clauses. We concentrate here on the first and the second continuum. One pole is that of parataxis, where two clauses are coordinated and are syntactically independent from each other. The other is that of

---

<sup>1</sup>We are aware of the fact that definitions of (non-)finiteness are always disputable. This definition is meant to apply to the languages that are at issue here.

<sup>2</sup>We take as a starting point Haspelmath's (2007: 1) definition of coordination: "The term coordination refers to syntactic constructions in which two or more units of the same type are combined into a larger unit and still have the same semantic relations with other surrounding elements."

hypotaxis, where two clauses are in a dependency relation to each other in the sense that one of the clauses either lies within the temporal, aspectual, modal or negative scope of its matrix clause or has a syntactic function, in that it is embedded in the matrix clause.<sup>3</sup> Languages may prefer finite subordination on the hypotactic pole, as is the case with English, German, Russian and Greek, or they may prefer non-finite subordination, as is the case with Turkish (see below). Whatever means the individual languages prefer, the relationship between finite and non-finite appears to be directed in the sense of the second continuum Lehmann (1988) develops, namely, that of reduction. In the opposition between finite and non-finite subordination in a given language (at least of the five languages which are at issue here), non-finite clauses are typically more reduced than finite subordinated clauses, or at least allow more reduction, and the non-finite subordination type may move up further towards the pole of reduction by getting “stripped” of further characteristics of verbality (e.g., binding arguments, a subject other than that of the matrix clause, marking for tense or aspect). Distributions of clause combining strategies in linguistic practice can be ordered along these continua.

## **1.2 Clause combining strategies of each of the languages involved**

In German, English, Greek and Russian, subordination is realized mainly by finite right-branching clauses with a preposed subjunctor. Subjunctors in relative clauses have pronominal character, though non-pronominal subjunctors are also possible, as are combined subjunctors. Finite adverbial subordinated clauses are introduced by subjunctors that define the nature of the temporal, spatial or logical relation to the matrix clause. Finite complement clauses are headed by various complementizers; specific matrix clause predicates (e.g. *verba dicendi*) allow complement clauses without an (overt) complementizer.

All four languages also have non-finite means of subordination, albeit to different extents:

English has reduced infinitival complement clauses, where the infinitive combines with the preposition *to*. It also has non-finite relative clauses which are

---

<sup>3</sup>Some works distinguish “chained clauses” from “subordinate clauses”, where “chained clauses” are dependent only in the sense that they lie within the TMA or negation scope of the matrix clause but have no syntactic position in it, see Sarvary & Choi (2020). In the languages at issue here, this distinction might apply to certain clause types in Turkish (Ögel-Balaban & Aksu-Koç 2020), which in our approach are subordinate adverbial (converb) clauses, see also Iefremenko et al. (2021).

postposed to their head noun: *the person to ask*. Furthermore, English has a non-finite verb-based form in *-ing*. This can be an attributive participle and an adverbial gerund (Nedjalkov 1995, Kortmann 1995). Clauses formed with this non-finite verb allow the use of connectives (e.g. *while going to school*, ...), they also allow the binding of direct objects and adverbial phrases, and can, in the form of “absolute” small clauses, also have subjects that are different from the matrix clause (e.g. *Dodo joined him, two laden bellboys following*, see Kortmann (1995: 217) and the discussion therein). *-ing*-forms have aspectual (perfective and continuous) differentiations.

German also has reduced infinitival clauses. These may be complement clauses, where the infinitival verb combines with the preposition *zu*, but may, with *verba dicendi* in the matrix clause, also stand without the complementizer. In addition, adverbial infinitive clauses may be formed with *um zu*. Furthermore, German has two participle forms, one in the present tense (*Partizip I*), one perfective (*Partizip II*). Both forms can form adverbial and relative clauses. When forming relative clauses, the participles are preposed to their head noun and inflect like adjectives. The German participles are not only much less frequent than their English counterparts, but the clauses they form are also more restricted in use; they can only be same-subject and usually consist only of the non-finite verbal form (cf. König & Gast 2018, 74f., 241f.).

Russian also has infinitives that form complement clauses (Shagal et al. 2022) and can form adverbial non-finite clauses (Madariaga 2011). In the latter function they are usually used together with such connectors as (*dlya togo*) *čtoby* ‘(in order) to’, *daby* ‘to’. Russian furthermore has adverbial participles (Kupriianova 2002), sometimes referred to as “gerunds” or “verbal adverbs”. *Deepričastije* is “a bare, non-finite verb phrase lacking a subject NP and headed by a particular type of verb form with adverbial features supplied by the adverbial participle suffix” (Babby & Franks 1998: 245). Non-finite relative clauses in Russian, also referred to as participles (in Russian linguistics usually called *pričastije*), inflect for gender and number and also express aspect (Kupriianova 2002).

Greek has no infinitival complements (Joseph 2020). It does, however, have a complement type where a finite form of the verb in subjunctive forms a clause introduced by the particle *na* (‘to’) and is inflected for person, number, tense and aspect (Joseph 2020). Greek also has non-finite verbal forms, which are formed by attaching the affix *-ondas* to the stem of the verb (e.g. *perpat-ondas* ‘walk-ing’). Gerunds have no tense and agreement features and they cannot bind arguments, thus gerundial/converbal clauses in Greek are highly reduced (see Roussou 2000: 87). They do have a temporal interpretation however, and can either be anterior or simultaneous to the main clause event, but never posterior (Tsimpli 2000: 137).

Contrary to English gerunds, Greek gerunds may not be introduced by any type of connective.

As for Turkish, we already hinted at the fact that subordination is realized mainly by means of non-finite forms. Complement clauses are clausal nominalizations, relative clauses are formed with participles, and adverbial clauses are formed with conversbs or other non-finite verbs which are combined with postpositions. Non-finite subordination is preposed to its governing structure, so Turkish hypotaxis is left-branching (Johanson 1992: 267). Non-finite subordinate clauses in Turkish can be fully expanded; there are certain adverbial clause forms, certain nominalizations and certain relative clauses which do allow independent subjects, and others which do not (see Iefremenko et al. 2021 for details). Turkish also has a few means of finite subordination with non-pronominal subjunctions; these form adverbial clauses and complement clauses. Finite complement clauses without an (overt) complementizer are also possible (Kerslake 2007).

Summarizing this brief overview, we see that four of the languages involved use finite subordination as the dominant strategy for hypotactic linking; one language, Turkish, prefers non-finite subordination. At the same time, there seem to be differences in the breadth of means for non-finite subordination among the languages that prefer finite subordination. Greek appears to be most restricted, in that non-finite clauses are highly reduced in form, in function, and with regard to the expansion they allow. German is a little richer as it has also non-finite complementation, but it also is fairly strict with regard to the expansion of non-finite clauses. Russian and English are broader in their employment of non-finite forms, both in terms of the variety of forms and functions, as well as with regard to lesser restrictions imposed on their expansion. And Turkish clearly has the broadest portfolio of non-finite clauses, in all aspects, but is on the other hand more restricted with regard to finite subordinate clauses. Overall, degrees of intensity and breadth of means for non-finite subordination in the languages at issue can be ordered the following way:

Turkish > Russian / English > German > Greek.

### **1.3 Situational parameters and clause combining**

The typological overview and the descriptions of the individual languages shows that clause combining strategies in individual languages are variable, in that speakers can use a variety of strategies in a given inter-propositional relation. We assume that the distribution of strategies used in each language is not random but

that, within each language, different varieties may prefer different clause combining strategies. This includes also varieties arising from different comm-sits (“registers”). In this context, Chafe (1976)’s early work on “information packaging”, as well as investigations of discourse structure (e.g., Miller & Weinert, 1998) and register-oriented approaches (Biber et al. 2011, Biber & Conrad 2019) show that in a given comm-sit, the choice between a strategy which tends more towards the paratactic and/or less reduced pole or one which tends more towards the hypotactic and/or more reduced pole depends on situational parameters. In ad-hoc comm-sits of higher context-dependence (e.g., informal situations), the frequency of paratactic linking structures increases, and in comm-sits of higher context-independence (e.g., formal situations), the frequency of hypotactic linking structures increases. This has also been shown, or at least discussed, for most of the languages we look at; for instance, see Kožina (2011) for Russian, Schroeder (2002, 2016) for Turkish, also under the heading of “academic language” for English in Schleppegrell (2004) and under “*Bildungssprache*” for German (Gogolin 2006, 2013, Haberzettl 2016). Related to this is the claim that embedded clauses indicate syntactic complexity (Sanchez Abchi & de Mier 2017, Housen et al. 2019), and they are proposed as diagnostic criteria to determine competence in first or second language acquisition in language testing,<sup>4</sup> or to determine complexity in written texts (Biber et al. 2011). On the other hand, Martynova et al. (2024) conclude that embedding rate in Russian is not an appropriate measure of complexity.

#### 1.4 Narration

As highlighted in the introduction, in addition to typological and situational factors, the selection of text type and the specific speech acts within that text type also impact the strategies employed in clause combining. Therefore, this chapter delves into the examination of structures within narrations. Narrations have three levels. The first and broadest level is that of text type. Narrations are a particular text type: A speaker conceptualizes a sequence of events that they experienced (e.g., participated in, saw, read, imagined) as meaningfully coherent and worth telling, and articulates this linguistically in the form of a coherent text (Labov & Waletzky 1997). The narrations we examine in the chapter have a particular macro structure, consisting of an opening, a narrative body, and a

---

<sup>4</sup>For example the German pre-school language test “Lise-DaZ”, see Schulz & Tracy (2011).

closing.<sup>5</sup> The second level of narration is the narrative body, where the narration of a plotline (i.e., a meaningfully knitted sequence of events) is dominant. It must be made clear here that the narrative text body does not consist only of iconically sequentialized descriptions of events. Woven into the narrative are also evaluative propositions, propositions that organize the text itself, dialogic interchanges with an interlocutor, and direct or indirect speech, all of which create a particular stance reflecting the personal experience of the speaker (Lehmann et al. 2023). The specific composition of the linguistic forms within the narrative body create the narrative structure. In the outline of their crosslinguistic developmental study “Relating events in the narrative”, Berman & Slobin (2013: 19) distinguish here between five categories of functions that these forms may express: Temporality, Event Conflation, Perspective, Narrative Type and Connectivity. Connectivity, which is in the focus of this chapter, is described as “knitting the fabric” of narrative discourse by means of syntactic conjunction and subordination” (*ibid.*). We investigate connectivity in the narrative body of the texts. In doing so, we assume that event descriptions must far outweigh other speech acts in narrations in completed language acquisition of adolescents and adults, in order to create a “well-formed” narration (Berman & Slobin 2013: 40). But we nevertheless assume that the degree to which other speech acts are interwoven bears a relation to mode (spoken, written), formality (informal, formal), the culture of story-telling in a particular language or society, and personal style (Lambrou 2005, Andrews et al. 2013).

## 1.5 Language contact and bilingualism

The question we are pursuing in this chapter requires taking into account yet another variable, namely that of the bilingualism of the speakers. This brings us, on the one hand, to the issue of change through language contact. Here the notion of “convergence” is important to us. We understand it as the structural approximation of languages to each other in language contact, which can express itself for example in the increased frequency of structures existent in both languages (Matras 2009, 2020, Grant 2020).

However, convergence must not necessarily be understood as mutual approximation between the languages involved. The bilingualism in focus here is not balanced (if such a thing exists at all). Rather, one of the languages (the “heritage language”, i.e. Russian, Greek, Turkish) is acquired as a first language, but

---

<sup>5</sup>We do not claim this macro structure to be a universal “story grammar”; see the respective critical discussion of Labov & Waletzky (1997) in Lambrou (2005), Andrews et al. (2013), amongst others.

is used only in specific comm-sits – predominantly in the family and with peers, less often in informal public settings – and there is little access to literacy or to language used in more formal comm-sits. The other language (i.e. German, English) is acquired as a second first language or early second language (age of onset before age 4), and it is the majority language of the social context and the dominant language of the formal public sphere including school, work life, social institutions, etc. (Maas 2008). On the basis of this sociolinguistic imbalance between the languages involved, we might expect convergence to express itself in structures of the heritage language(s) rather than in both languages involved.

The specific sociolinguistic distribution of languages brings into play yet other possible language contact dynamics beyond convergence, and these are related to language acquisition. For a long time, research on multilingualism was dominated by a claim of “bilingual disadvantage”, meaning that bilingualism leads to higher processing costs in both languages (see Higby et al. 2013 for a neurolinguistic overview). Also, on the side of the heritage language, studies spoke of “incomplete acquisition”, “attrition” or “erosion” of structures (see Aalberse et al. 2019: ch. 7 for a critical overview). Across RUEG projects, however, we assume that different varieties in the speech of monolingual vs. bilingual heritage speakers rather emerge from limited access of heritage speakers to the structures of the written standard of the language and to the registers of formality. This may lead to a levelling between the linguistic practices used in different comm-sits (Wiese et al. 2022). Importantly, however, there may be differences in the relationship between majority and heritage language(s) in different societies and there are reasons to assume that this also has an impact on the development of the heritage language (Iefremenko et al. 2021). The size of the speaker group may play a role here, and the vitality of the particular heritage language in the country (group size and vitality may or may not correlate).

Generally speaking, clause combining is a grammatical area which is clearly sensitive to language contact and convergence. For example, the fact that Russian uses more non-finite subordination than any other Slavic language has been attributed to centuries of language contact of Russian with Uralic and Altaic languages, which rely more heavily on non-finite subordination (Grenoble 2013: 772). For Turkish, numerous works show shifts from non-finite to finite subordination in languages of the Turkic language family, due to contact with Indo-European languages that prefer finite subordination (Johanson et al. 2020: 566 for an overview, Keskin 2023, 2025 [this volume] for the Balkans). Also, expansions of syndetic paratactic clause combining strategies have been shown to be accelerated by language contact (Mithun 1988), and coordinating conjunctions seem

to be a part of speech that is frequently borrowed between languages (Winford 2003).

As for clause combining and bilingualism, studies on German as an early second language speak of delayed acquisition of subordination structures in bilingual children starting school (age of onset of L2 between 2 and 4 years; Grimm & Schulz 2014). Other studies investigating monolingual and bilingual German-speaking pupils do not find evidence of a “bilingual disadvantage”; see Goschler (2017) for primary school pupils and (Haberzettl 2016) for 13-year-olds. For clause combining in heritage languages on the other side, we find a claim that heritage speakers use less subordination than majority speakers of the same language; see Polinsky (2018: 47) for a generalization across heritage languages. Polinsky (2008) also states this to be valid for Russian heritage speakers in the United States. For heritage Turkish, Treffers-Daller et al. (2006), Bayram (2013) and Yilmaz-Wörfel (2022) argue in a similar direction. This development has in the past sometimes been attributed to “incomplete acquisition” (Polinsky 2008, Bayram 2013), sometimes also to “attrition” (Yagmur 1997) or “erosion” (Dussias 2004). However, Sanchez Abchi & de Mier (2017) do not find any evidence for a “decline” in subordination in the written narrative texts of Spanish heritage speakers in France and Germany. Several studies using the RUEG methodology also do not find such evidence for Russian heritage speakers in Germany and the US and German heritage speakers in the US (Martynova et al. 2024, Pashkova et al. 2022, Tsehayé 2023). In a similar vein, a growing body of research concentrating on Turkish as a heritage language steps away from the more deficit-oriented perspective and shows that the dynamics in heritage Turkish should rather be understood as the emergence of a new system of clause combining, which expands both paratactic connectivity and finite subordination (Herkenrath et al. 2003, Karakoç 2007, Onar Valk 2015, Schroeder 2016, Özsoy et al. 2022, Keskin et al. 2024), and is related to limited access to the formal register of the language (cf. Aalberse et al. 2019: 57).

## 2 Methodology

### 2.1 Data collection

The data used for the study come from the latest version of the RUEG corpora (RUEG-DE, EL, EN, RU, TR) (Lüdeling et al. 2024). The data were collected using the “language situations” method described in Wiese (2020). This method elicits naturalistic and ecologically valid narrative data across different formalities and modes (see Wiese et al. 2025 [this volume]). Participants were shown a short

video with a staged minor car accident and asked to imagine that they witnessed it. Then participants were asked to retell this accident in four different contexts: to a friend via a WhatsApp voice message (informal-spoken), to a friend via a WhatsApp text message (informal-written), to the police via a voicemail (formal-spoken), and to the police in the form of a written witness report (formal-written). Bilingual participants completed the task twice, with a break of at least three days between sessions: once in their heritage language and once in the majority language. Monolingual speakers' data were elicited only in their native language. All participants filled out an extensive online language background questionnaire, including questions about their metalinguistic profiles.

## 2.2 Participants

For the current study, we analyze data from five languages, namely Greek, Russian, Turkish, German and English. In the first three languages, we investigate heritage varieties in Germany and in the US and monolingual varieties in Greece, Russia and Turkey, respectively (see Table 1 for the number of speakers in each group). German and English are analyzed as majority languages of bi- and monolingual speakers in Germany and the US respectively (see Table 2 for the number of speakers in each group). All participants are teenagers (15–17 years old) who at the time of data collection were attending school. As mentioned in Wiese et al. (2025 [this volume]), the data were controlled for (in)formality (formal vs. informal) and mode (spoken vs. written). As a result, each speaker produced four productions in each of their languages. In total, we analyzed 1588 texts produced by a total of 277 participants.

## 2.3 Annotation

The main body of each narrative was annotated for connectivity. The main body comprises the text that excludes openings and closings, that is, it is the part of the text that contains the story line (see Katsika et al. 2025 [this volume]). Using EXMARaLDA (Schmidt & Wörner 2014), we annotated the data on three tiers for the following features:

- I. Speech act
- II. Clause type
- III. Connectivity elements

Table 1: Number of speakers and texts for Greek, Russian, Turkish speakers

| Language | Status   | Country | Number of speakers | Number of texts |
|----------|----------|---------|--------------------|-----------------|
| Greek    | Heritage | Germany | 20                 | 80              |
|          | Heritage | USA     | 24                 | 96              |
|          | Majority | Greece  | 24                 | 96              |
| Russian  | Heritage | Germany | 28                 | 112             |
|          | Heritage | USA     | 34                 | 136             |
|          | Majority | Greece  | 35                 | 140             |
| Turkish  | Heritage | Germany | 24                 | 96              |
|          | Heritage | USA     | 24                 | 96              |
|          | Majority | Greece  | 24                 | 96              |

Table 2: Number of speakers and texts for German and English speakers

| Language | Status   | Country | Group           | Number of speakers | Number of texts |
|----------|----------|---------|-----------------|--------------------|-----------------|
| German   | Majority | Germany | German-Greek    |                    |                 |
|          |          |         | bilinguals      | 16                 | 64              |
|          |          |         | German-Russian  |                    |                 |
|          |          |         | bilinguals      | 16                 | 64              |
|          |          |         | German-Turkish  |                    |                 |
|          |          |         | bilinguals      | 16                 | 64              |
| English  | Majority | USA     | German          |                    |                 |
|          |          |         | monolinguals    | 16                 | 64              |
|          |          |         | English-Greek   |                    |                 |
|          |          |         | bilinguals      | 24                 | 96              |
|          |          |         | English-Russian |                    |                 |
|          |          |         | bilinguals      | 24                 | 96              |
|          |          |         | English-Turkish |                    |                 |
|          |          |         | bilinguals      | 24                 | 96              |
|          |          |         | English         |                    |                 |
|          |          |         | monolinguals    | 24                 | 96              |
|          |          |         |                 |                    |                 |
|          |          |         |                 |                    |                 |

I. The *Speech act tier* comprises the following categories, which only apply to clauses.

- *Situational*: Includes all clauses that directly relate to the description of the events displayed in the stimulus video. This tier also includes instances of negation and modality. An example of a situational speech act is given in (1).

- (1) und der Mann hat-te ein Ball in der Hand genau.  
 and the man have-PST.3SG a ball in the hand yeah  
 ‘And the man had a ball in his hands, yeah.’ (DEbi61MG\_fsD)<sup>6</sup>

- *Subjective*: Includes all predication which explicitly express the speaker’s personal stance, such as evaluations, explanations, personal statements, e.g. example (2).

- (2) It was a mess.  
 (-)  
 (DEbi61MG\_fsD)

- *Intersubjective*: Includes all dialogic clausal instances in which the speaker addresses the imaginary interlocutor (see example 3).

- (3) aber jo (-) ah weiß-t du was ich krass  
 but well ah know-PRS.3SG you what I extreme  
 fand?  
 find.PST.1SG  
 ‘But do you know what I found extreme?’ (DEmo53FD\_isD)

---

<sup>6</sup>The participant codes provide the following information:

- Country: DE—Germany; GR—Greece; RU—Russia; TU—Turkey; US—US
- Bi-/monolingual speaker: bi vs. mo
- Speaker number incl. age group: 01 to 49—adults; from 50 onwards—adolescents
- Gender: M vs. F (there were no speakers who identified as non-binary)
- Heritage language for bilingual speakers or only majority language for monolinguals: D for German; E for English; G for Greek; R for Russian; T for Turkish
- comm-sit: formal—f vs. informal—i; spoken—s vs. written—w
- Language of production: D for German; E for English; G for Greek; R for Russian; T for Turkish

- *Textual*: Includes predication that relate to the organization of the narrative (see example 4).

(4) bun-un-la da kal-ma-d<sub>1</sub>  
this-GEN-INST also stay-NEG-PST.3sg  
'And it didn't end here.' (TUmo57FT\_fsT)

- *Direct speech*: Includes instances of direct speech within the narration, e.g. example (5).

(5) sta-l takoj (-) ruk-i na bok takoj oj da  
stand-PST.3SG such hand-PL on side such oh yes  
izvin-i-te  
excuse-IMP-2PL  
'and he was like with his arms akimbo: "oh yeah, excuse me."  
(USbi69FR\_isR)

II. The *Clause type tier* comprises the following categories:

- a) Independent clause or matrix clause
- b) Elliptical clause
- c) Complement clause: finite or non-finite
- d) Adverbial clause: finite or non-finite
- e) Relative clause: finite or non-finite

Not all types of clauses are present in each language. Thus, for example, there are no instances of finite relative clauses in Turkish (as expected).

III. The *Connectivity tier* comprises the categories listed in Table 3.

Asyndetic linking was interpreted as the absence of overt linking devices between independent extended or simple clauses.

### 3 Results

This subchapter displays the descriptive quantitative results of our investigation. We present figures and explain them briefly, focusing on four topics that were introduced in Sections 1 and 2. First, we look at the degree to which the texts,

Table 3: Connectivity tier

| Connectivity  | Example (from English, for convenience)                |
|---|--|
| Neutral (NCON)  | <i>And</i>   |
| Event-related and logical (ELCON)   | <i>but, moreover, fortunately, luckily, either, or</i> |
| Spatial (SAD)   | <i>behind (him), across (from him), (from) behind</i>  |
| Temporal (TAD)  | <i>then, afterwards, at that time, at that moment</i>  |
| Temporal adverbial clause, finite or non-finite (which explicitly refers to the time of a previously narrated event) (TACA) | <i>when they arrived</i>                               |
| Subordinating connector (SCON)  | <i>while, because, who (relative)</i>                  |
| Linking discourse marker (LDM)  | <i>yeah, like, well</i>                                |
| Filled pause (FP)<br>(tagged only when there are no other connectivity devices in the clause)                               | <i>um, e, ee</i>                                       |

stripped of their openings and closings, are narrations on the second level, as discussed in Section 1.4. That is, we investigate whether a description of events dominates, as expected, and to what extent other speech acts are involved (Section 3.1). Second, we focus on subordination and investigate in more detail the frequency and distribution of subordination strategies (finite vs. non-finite, different clause types) (Section 3.2). Third, we look at the distribution of strategies for paratactic clause combining (Section 3.3). Fourth, throughout the Subsections 3.1 to 3.3, we inquire into the linguistic practice in the different comm-sits distinguished here.

### 3.1 Speech acts

Figures 1 and 2 present the frequency distribution of speech acts in heritage and majority languages of the speakers. Both figures clearly show that the situational speech act by far dominates. Across all speaker groups, it makes up between 89 and 98% of all clauses. Thus, the text sections that we are investigating indeed

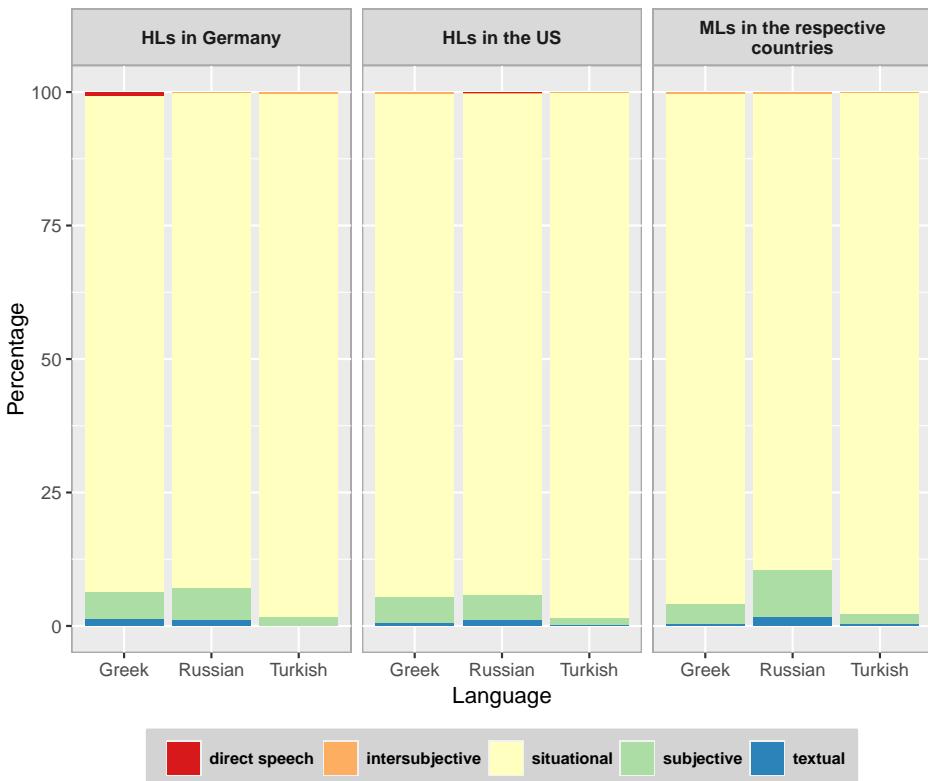


Figure 1: Percentage of different types of speech acts in the bilingual speakers' heritage language and monolingual speakers' majority language

qualify as narratives. Next to the descriptive speech act, only the speech act of evaluation (subjective) is worth mentioning, which has an overall percentage of 1 to almost 9%. The percentage of evaluative speech acts is the highest in the Russian texts; in the monolingual Russian speakers from Russia they reach 8.8%, while in the Russian heritage speakers in Germany evaluations are at 6%. The lowest percentage of evaluations is found in the Turkish texts; the three Turkish speaker groups (monolinguals, heritage speakers in the US and in Germany) have between 1.4 and 1.9% of evaluative comments in their texts. The next most frequent, at 3.7%, is found in Greek-English bilinguals in their English productions. There is no evidence of cross-linguistic influence from the heritage languages to the majority languages in percentage of evaluative speech acts. In German, there are no notable differences between the speaker groups, and in English, the

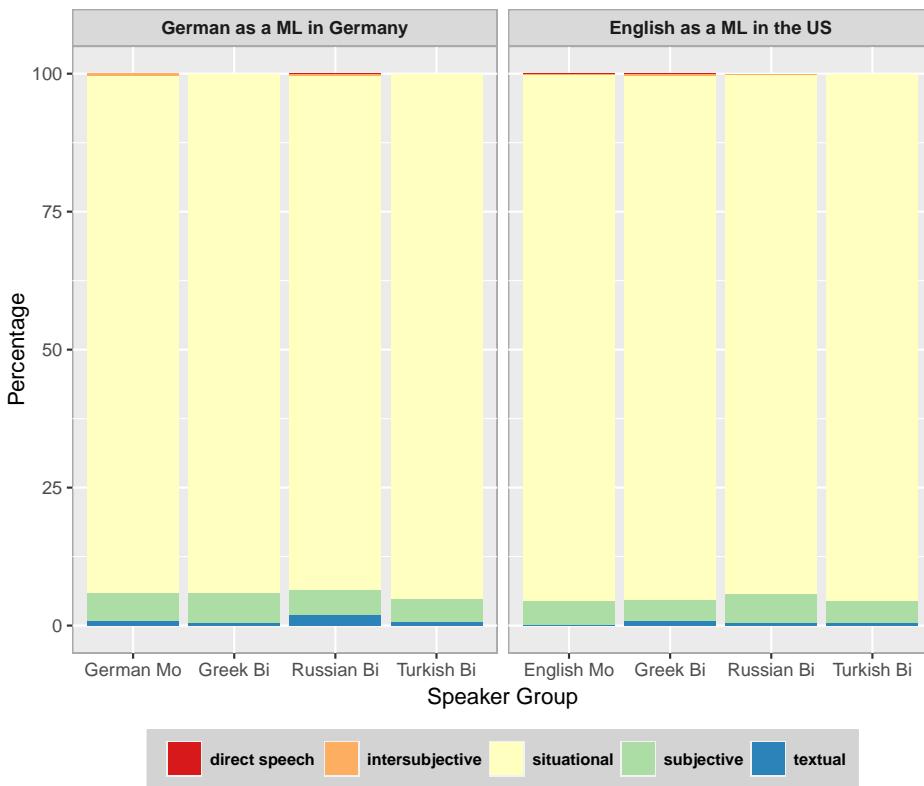


Figure 2: Percentage of different types of speech acts in the bi- and monolingual speakers' majority languages

Russian-English bilinguals may have the highest ratio (5.5%), but the monolingual speakers score on the same level (5.2%), while the Greek-English bilinguals and the Turkish-English bilinguals have a considerably lower ratio (3.7% and 3.7%, respectively).

Figures 3 and 4 present frequency distribution of speech acts in heritage and majority languages of the speakers across the four comm-sits: informal-spoken, informal-written, formal-spoken and formal-written. From the figures we see that the situational speech act dominates in the four situations. At the same time, the figures also show that subjectivity is very much a characteristic of informal situations, while in the formal situations, particularly the written mode, subjective comments are scarce. Such differences between the comm-sits are most evident in the monolingual majority varieties of the investigated languages, less so in the bilingual majority varieties, and are least pronounced in the heritage varieties.

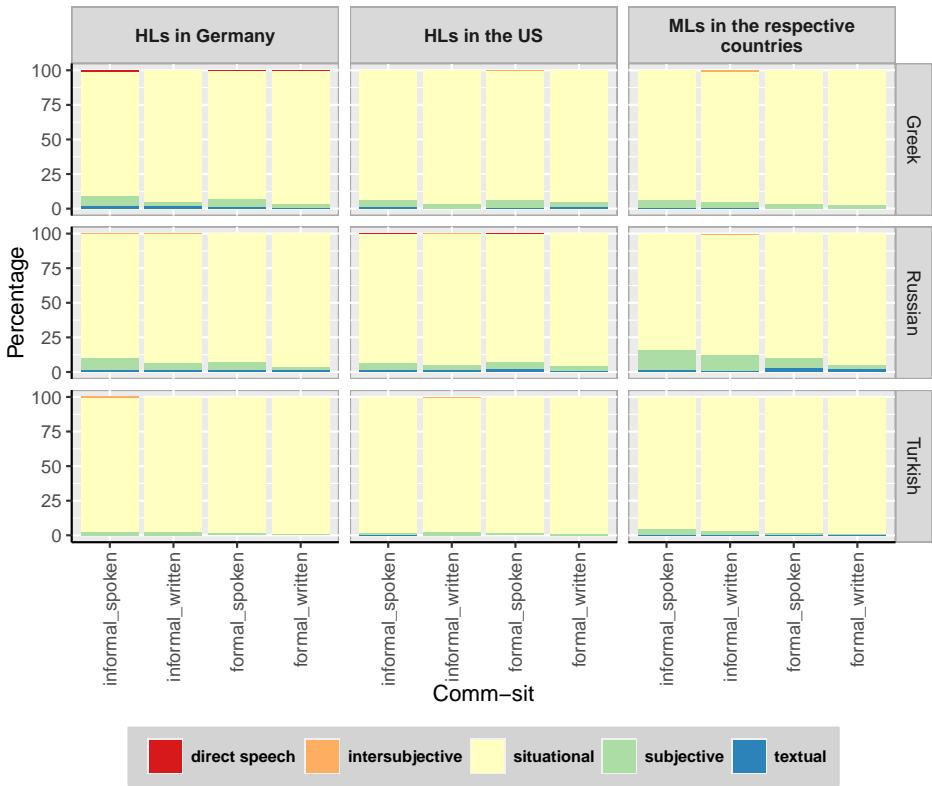


Figure 3: Percentage of different types of speech acts in the bilingual speakers' heritage language and monolingual speakers' majority language across comm-sits.

### 3.2 Subordination and subordination strategies

We start with a general overview on the quantities of subordinate clauses and then proceed to details with regard to subordination strategies.

Figure 5 and 6 show frequency distribution of main and subordinate clauses in heritage and majority languages of the speakers across four comm-sits. When we look at the ratio of subordinate clauses within all clauses in Russian, Greek and Turkish (Figure 5), we find that in Turkish and Greek, the intensity of the use of subordinate clauses drops considerably in the heritage varieties, while in Russian, it does not.

In the majority languages in Germany and the US, the percentage of subordinate clauses does not differ much between the speaker groups. However,

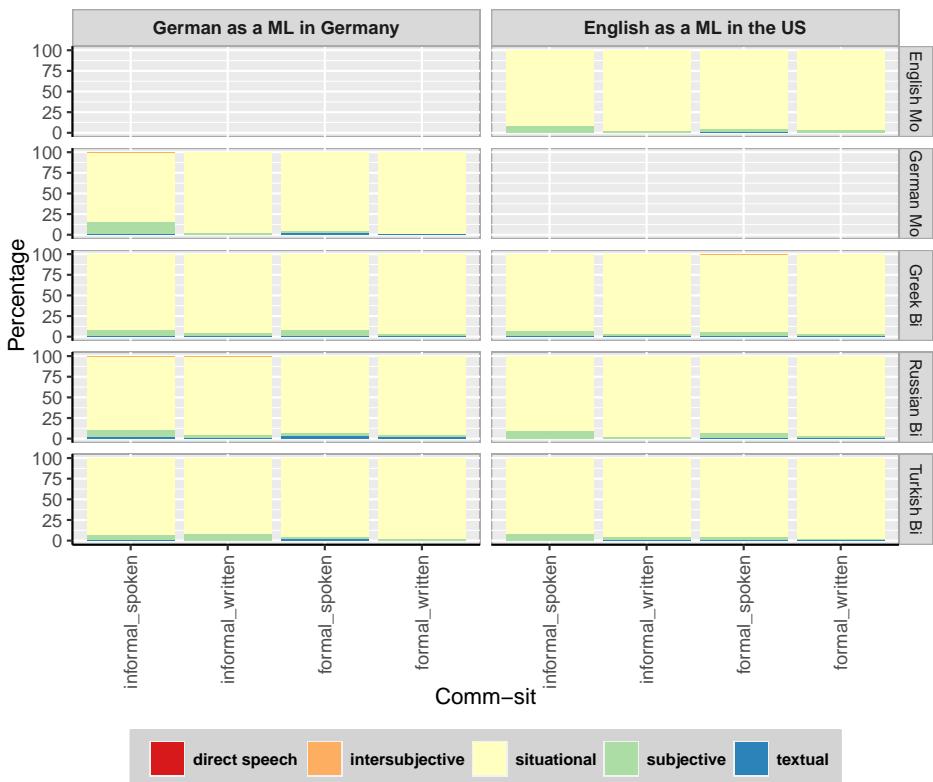


Figure 4: Percentage of different types of speech acts in the bi- and monolingual speakers' majority languages across comm-sits.

German-Greek bilingual speakers tend to employ more subordinate clauses than German-Russian and German-Turkish bilinguals.

Furthermore, the figures show that formality leads to more subordination: the number of subordinate clauses in the informal-spoken and informal-written comm-sits is lower than in the formal comm-sits. On the other hand, mode (spoken vs. written) cannot be generalized to make a difference: In some speaker groups, formal-written texts have the highest percentage of subordinate clauses, while in other speaker groups the formal-spoken texts have the highest percentage, and vice versa for informal-written versus informal-spoken and (lower) percentages of subordinate clauses.

In the heritage varieties, the difference in intensity of use of subordinate structures between formal and informal is smaller, compared to the corresponding monolingual language practice – lesser so for heritage speakers in the US than

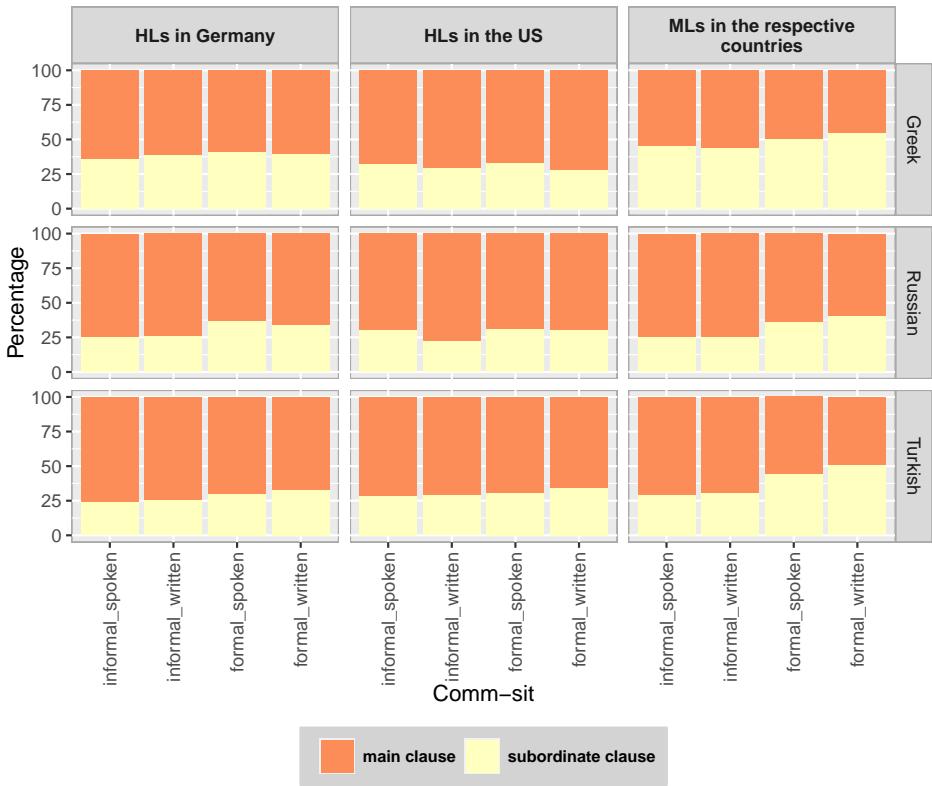


Figure 5: Percentage of main clauses vs. subordinate clauses in the bilingual speakers' heritage language and monolingual speakers' majority language across comm-sits.

in Germany. Note, however, that the frequency of the use of subordinate structures in the informal-spoken comm-sit is similar between the monolingual and the heritage varieties (for Turkish also in the informal-written and for Russian also in informal-written and formal-spoken comm-sit). As for the monolingual and bilingual speakers of the majority languages English and German, we find differences between the groups, but no opposition between monolinguals on the one hand and bilinguals on the other.

Next, we zoom in on subordinate clauses and discuss frequency distribution of finite vs. non-finite structures in the languages under research here.

Figures 7 and 8 demonstrate the frequency distribution of finite and non-finite subordinate clauses in heritage and majority languages of the speakers across four comm-sits. In general, figures show that the typological characteristics hold.

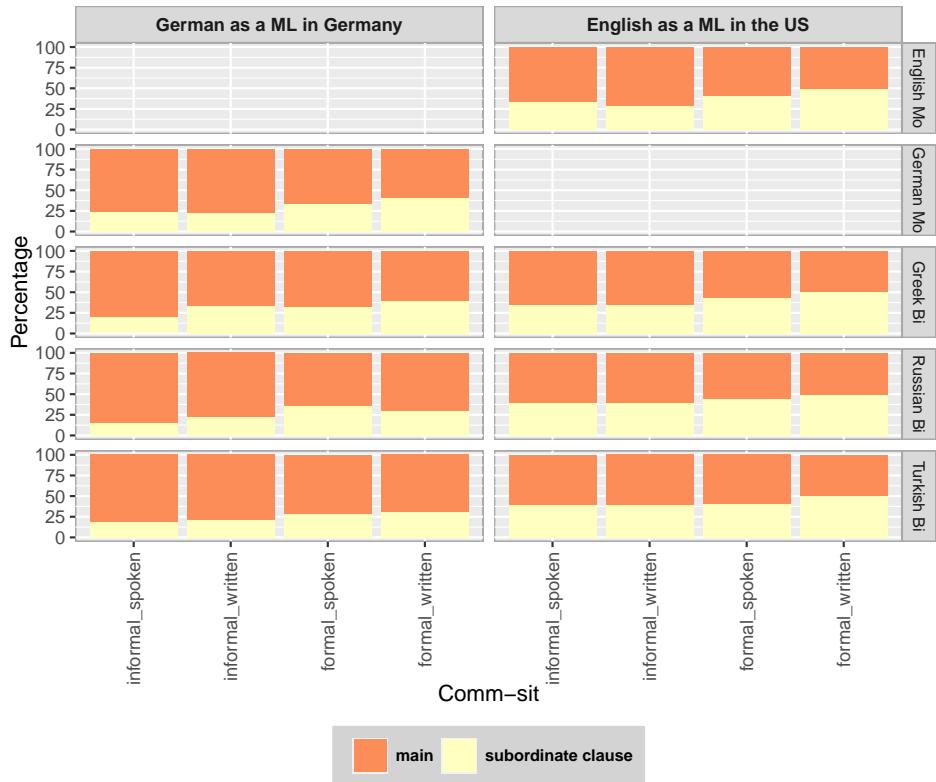


Figure 6: Percentage of main clauses vs. subordinate clauses in the bi- and monolingual speakers' majority languages across comm-sits.

That is, all Turkish speaker groups overwhelmingly use non-finite subordination strategies; the Greek speakers use the lowest percentage of non-finite clauses as compared to finite clauses, the Russian and the English speakers again use more non-finite structures than the German speakers. Thus, the intensity of the use of non-finite structures goes as shown in Section 1.2:

$$\text{Turkish} > \text{Russian, English} > \text{German} > \text{Greek}$$

Furthermore, Figures 7 and 8 show different distributions of finite vs. non-finite subordinate clauses across heritage and majority speaker groups in different comm-sits. In the monolingual varieties of Greek, Russian, and Turkish, formality leads to the use of more non-finite subordinate clauses. Heritage speakers of the three languages produce more finite subordinate clauses in contrast to non-finite subordinate clauses, compared to monolinguals (although the difference in

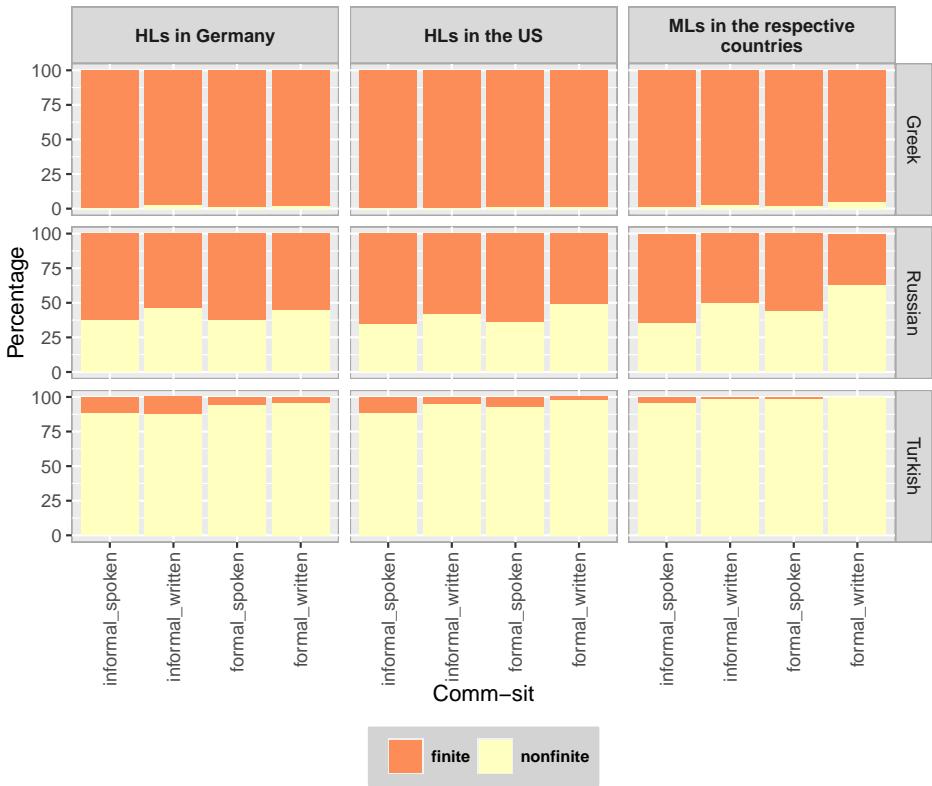


Figure 7: Percentage of finite vs. non-finite subordinate clauses in the bilingual speakers' heritage language and monolingual speakers' majority language across comm-sit.

percentage varies considerably across the languages). However, for Russian and Greek heritage speakers this is only observed in the formal comm-sits, while Turkish, it is observed across all four comm-sits.

As for the majority languages, Figure 8 shows that German monolinguals prefer finite subordinate clauses, particularly in the spoken mode. While German-Russian bilinguals closely resemble German monolinguals, German-Turkish and German-Greek bilinguals employ more finite subordinate clauses than German monolinguals, particularly in the written mode. As for English, all groups behave similarly in terms of frequency distribution of finite vs. non-finite subordinate clauses.

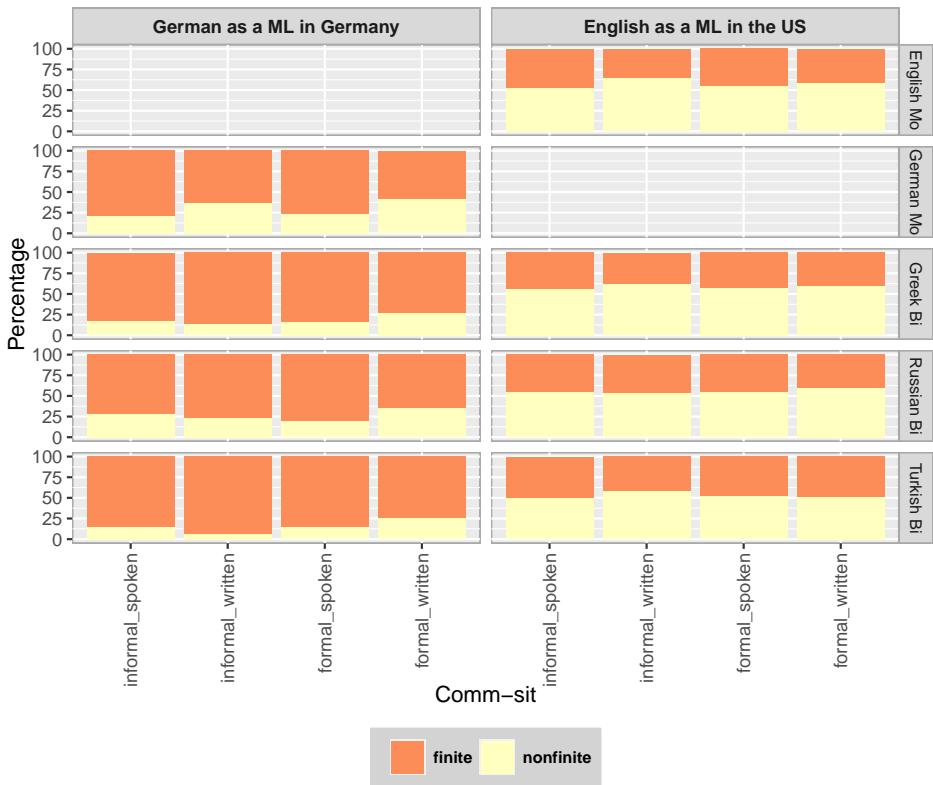


Figure 8: Percentage of finite vs. non-finite subordinate clauses in the bi- and monolingual speakers' majority languages across comm-sits.

### 3.3 Paratactic clause combining strategies

A further observation concerns paratactic clause combining devices. Here we distinguish between different semantic kinds of connectors (see Section 2) and present the percentage of asyndetically connected clauses.

Figures 9 and 10 present frequency distribution of different kinds of connectors across comm-sits in the five languages. First of all, we can go back to Figures 5 and 6 and generalize that the reverse of what was said for subordinate clauses counts for paratactic linking; in other words, the lesser the percentage of subordination, the more the use of paratactic combining of clauses. That means that the heritage varieties of Greek and Turkish use overall more parataxis than the monolingual varieties, while for Russian, this generalization does not apply.

Turning now to the details of paratactic clause combining, we do not find notable differences between the different groups of the majority languages English

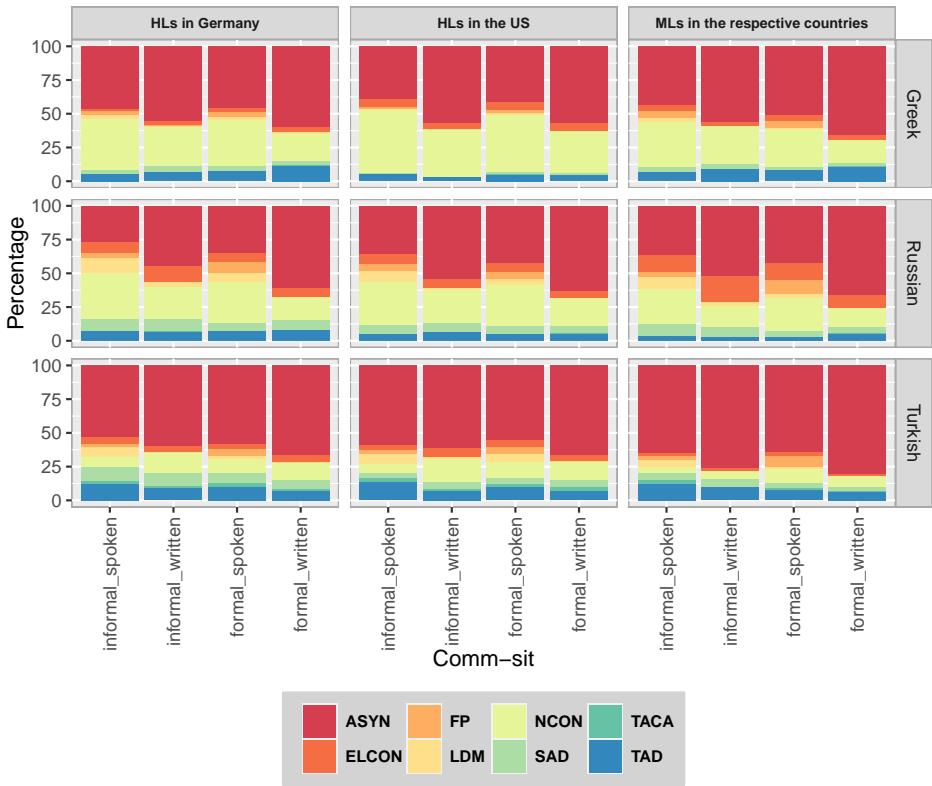


Figure 9: Percentage of different kinds of connecting devices in the bilingual speakers' heritage language and monolingual speakers' majority language across comm-sits.

and German. In Greek and Russian, on the other hand, the heritage speakers use the neutral connector more frequently than the corresponding monolingual speakers. However, what is interesting is that this does not seem to raise the frequency or variety of other connecting devices used by the monolingual speakers. Rather, the monolinguals use a higher percentage of asyndetic connections than the bilinguals, that is, no overt connecting device at all.

We also find that in all speaker groups with the exception of Turkish, there is a considerable drop in (relative) frequency of neutral connectors in the formal-written texts, as opposed to the other comm-sits, and, vice versa, a rise of the use of semantically more specific paratactic connectors (event-related or temporal connectors). In Greek and Russian, this drop is less pronounced in the heritage varieties than in the monolingual varieties, that is, the heritage speakers use the

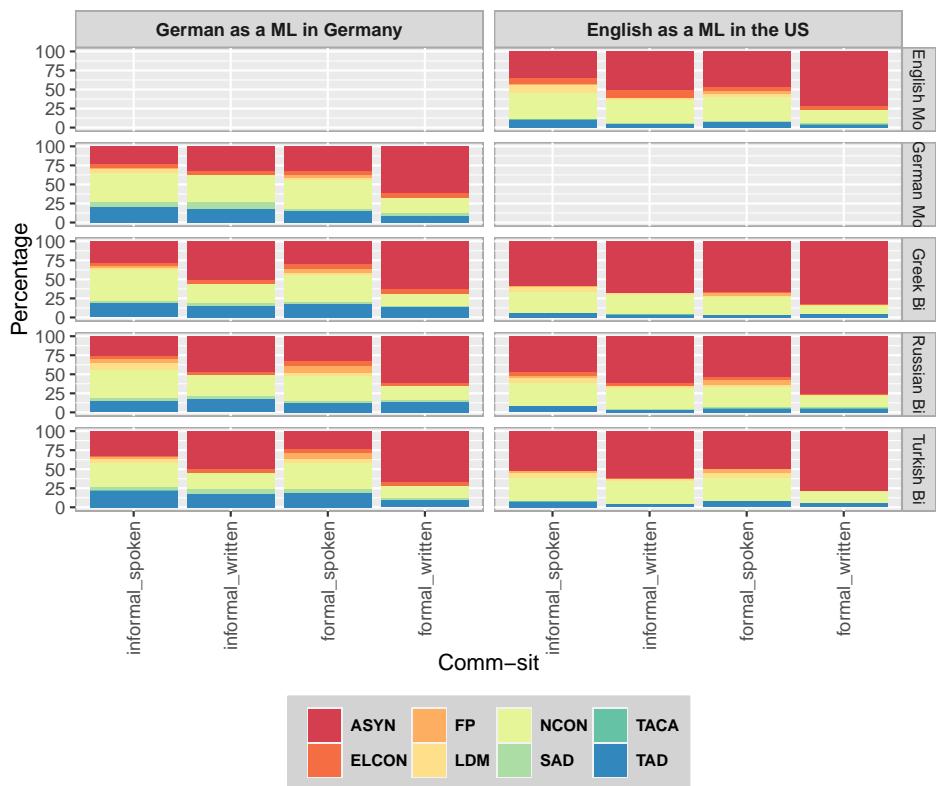


Figure 10: Percentage of different kinds of connecting devices in the bi- and monolingual speakers' majority languages across comm-sits.

neutral connector relatively more than the monolingual speakers in the formal comm-sits.

For Turkish, the picture looks quite different. First of all, compared to all other languages, the absence of overt paratactic connecting devices is most prominent here – over 50 per cent of all paratactically linked clauses have no connector (compared to 30–40% in the other languages), and in monolingual speakers in Turkey this absence is more prevalent (60–80% depending on the comm-sit) than in heritage speakers (55–65% depending on the comm-sit). This shows a stronger tendency among heritage speakers towards the use of overt paratactic linking devices, and they also use a broader variety of overt linkers than the monolingual speakers, including new connectors based on discourse markers (Labrenz et al. 2025 [this volume]).

Furthermore, we also observe a high frequency of event-related and logical connectors in the Russian speakers, with the highest occurrence found in monolingual Russian speakers. Interestingly, in Russian-German bilinguals this pattern is exclusive to the Russian language and is not found in their German productions. Such an increase in the frequency of event-related and logical connectors may be the result of the Russian speakers using a high number of evaluative comments (as discussed in Section 3.1).

## 4 Discussion

In this chapter, we investigated clause combining in typologically different languages: four Indo-European languages, namely English, German, Greek and Russian, and an Altaic language, Turkish. We explored the majority as well as heritage varieties of these languages, thus investigating both languages of the bilinguals. The goal was to explore whether there are similar dynamic patterns in clause combining strategies among the speakers of the five languages under research, and to find out to what extent the potential differences and similarities between the varieties are related to the typological structures of the languages, to communicative situations, and to language contact.

First of all, we can safely say that the typological differences between the languages at issue concerning preferences of subordination strategies are preserved also in the language contact settings addressed here. For instance, in terms of the finiteness of subordinate clauses, heritage varieties pattern with the corresponding majority varieties. Thus, Turkish predominantly relies on non-finite subordination; Greek speakers overwhelmingly use finite structures, while Russian, English and German speakers employ both strategies, although to different degrees. Bilingualism does not bring radical changes, neither to the majority languages nor to the heritage varieties we investigate here, which cling to the general typological characteristics of the corresponding monolingual varieties. This adherence exists in spite of intensive language contact which is, however, relatively recent.

Nevertheless, some phenomena raise further questions and some changes seem to be arising in the heritage and majority varieties.

To begin with the majority languages, we certainly cannot generalize a distinction between monolinguals on the one side and bilinguals on the other, as the claim of a “bilingual disadvantage” (see Section 1.5) would have it. What we do find are different preferences across the speaker groups, for example differing frequencies of the uses of subordination in general, and differing frequencies

in the employment of non-finite structures. In very selected domains, these differing preferences may allow us to discuss issues of convergence. For example, we see that the Greek speakers in Germany have the highest percentage of subordination among the heritage speaker groups, and they are also the bilingual group which has the highest percentage of subordination in German. All in all, language contact does not provide a good argument for the different distributions of clause combining strategies between the speaker groups in the majority languages. At minimum, language contact cannot be the only factor. Thus, we have to leave this question to further research and discussion.

We did not find much evidence for effects of language contact in the majority languages, but do we find evidence in the structures of the heritage languages? The clearest phenomenon we find here is the “levelling of registers”. The linguistic practices (here: clause combining strategies) in the different comm-sits show fewer cross-linguistic differences in the heritage varieties than in the monolingual varieties. We already hinted at a possible reason here: Heritage speakers have less access to the variety of registers of the standard language, at least to formal registers. Consequently, structures that would be classified as belonging to informal registers in the monolingual varieties are generalized also to more formal comm-sits in the heritage varieties. Interestingly, this levelling seems more pronounced in the US than in Germany. Is this related to a higher vitality of Russian, Greek and Turkish in Germany as compared to the US, in the sense that speakers of the heritage languages can use or are exposed to a larger variety of comm-sits in these languages in Germany than in the US? For Turkish and Russian, this seems to be the case, as Iefremenko et al. (2021) and Schroeder et al. (2024) argue for Turkish and Martynova et al. (2024) for Russian, but it remains to be investigated whether the same argument holds also for Greek.

Furthermore, our research here is not consistent with a characterization of heritage languages in general as “incomplete acquisition” or “attrition” or “erosion”. All subordinating structures are used by all speaker groups, and at least in the informal-spoken comm-sit, there are no differences between monolingual and heritage speakers in terms of the frequency of uses of subordinate structures. Thus, the overall higher frequency of subordinate structures in the monolingual varieties is rather related to the speakers’ differing practices in more formal comm-sits, where they use more subordinating and less paratactic structuring, while the heritage speakers adhere somewhat more to the linguistic practices that they also use in the informal comm-sits. Thus, it is not that the heritage speakers do not “have” the respective subordinating structures, it is just that they do not employ them in similar frequencies. The same holds for the use of the neutral connector in Russian and Greek. And it also accounts for the very

small percentage of non-finite structures in the language use of the Greek heritage speakers. These structures seem to belong to the formal register of standard Greek, to which the Greek speakers in Germany and the US have lesser access.

A further interesting point is the observation that asyndetic linking seems to be more frequent in the monolingual varieties of Greek, Russian and Turkish, as compared to the heritage varieties. Does this point to a higher “explicitness” and “transparency” in the language use of heritage speakers, as pointed out in Polinsky (2018) – in this case a more intensive use of overt paratactic linking devices? For Turkish, the tendency to use overt paratactic linking devices goes hand in hand with an expansion of the system of these devices and a stronger use of finite subordination strategies, as opposed to non-finite subordination. Together with Özsoy et al. (2022) we argue that here, register levelling and convergence combine together. The speakers expand a particular subordination strategy that is already there in the language, and they expand it according to the example of their majority languages.

All in all, then, just as we found for the majority languages, language contact does not seem to be the decisive factor initializing and/or driving change in the heritage languages. Rather, the strongest factor seems to be the particular sociolinguistic situation of heritage languages in the majority societies, narrowing the spectrum of language practice in different comm-sits, and leading to an expansion of the language practice from informal comm-sits.

The exploratory study presented in this chapter lays strong foundations for further work. Although we have examined several aspects of subordination structures in this chapter, other aspects remain to be examined that will undoubtedly shed further light on tendencies to change in the system. These include more detailed analyses of degrees of expansion or reduction of subordinate clauses along the continua proposed in Lehmann (1988), as well as case hierarchies in relative clause formation (Polinsky 2008). Both points relate less to the frequency and occurrence of subordinate structures than to the breadth of what the structures are used for. Furthermore, we believe that a closer look at individual variation would strengthen our argument in the direction of a sociolinguistic explanation (Özsoy & Blum 2023).

Summarizing the main results from the chapter, we can say that the study contributes to a better understanding of the causes for changes in both languages of heritage speakers. Importantly, it is not the bilingualism of the speakers but rather their limited access to formal registers in their heritage language that is the reason for the observed differences. Hence, our study corroborates the findings of the several recent studies from our research unit that demonstrated that

the cause for the differences in the linguistic structures between heritage and majority varieties can be attributed to register levelling in heritage speakers (Alexiadou & Rizou 2023, Wiese et al. 2022, Tsehayé 2023, Schroeder et al. 2024, among others). Furthermore, our study supports the claim that vitality of a heritage language plays an important role in the development of the language, as has already been discussed for Turkish and Russian in recent works from our research group (Iefremenko et al. 2021, Schroeder et al. 2024, Martynova et al. 2024).

## Acknowledgements

We are indebted to our student assistants, who carried out the immense annotation tasks: Iliuza Akhmetzianova, Franziska Groth, Ksenia Idrisova, Oğuzan Kuyrukçu, Simge Türe, and Barbara Zeyer. Also, we want to extend special thanks to our student assistant, Özce Özceçelik, for handling the formatting-related tasks in L<sup>A</sup>T<sub>E</sub>X, as well as to Lea Coy for help with pre-publication formatting. We also thank to the anonymous reviewers for their valuable and highly constructive comments. The research was supported through funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit *Emerging Grammars in Language Contact Situations*, project P9 (grant number: 313607803).

## Abbreviations

|     |            |      |              |      |           |
|-----|------------|------|--------------|------|-----------|
| 2   | 2nd person | INST | instrumental | PRET | preterite |
| 3   | 3rd person | NEG  | negative     | PST  | past      |
| GEN | genitive   | PL   | plural       | SG   | singular  |
| IMP | imperative | PRES | present      |      |           |

## References

- Aalberse, Suzanne P., Ad Backus & Pieter Muysken (eds.). 2019. *Heritage languages: A language contact approach* (Studies in Bilingualism 58). Amsterdam: John Benjamins. DOI: 10.1075/sibil.58.
- Alexiadou, Artemis & Vasiliki Rizou. 2023. The use of periphrasis for the expression of aspect by Greek heritage speakers: A case study of register variation narrowing. *Register Studies* 5(1). 82–110. DOI: 10.1075/rs.20022.ale.
- Andrews, Molly, Corinne Squire & Maria Tamboukou. 2013. *Doing narrative research*. Thousand Oaks, CA: Sage.

- Babby, Leonard & Steven Franks. 1998. The syntax of adverbial participles in Russian revisited. *The Slavic and East European Journal* 42(3). 483–516.
- Bayram, Fatih. 2013. *Acquisition of Turkish by heritage speakers: A processability approach*. Newcastle: Newcastle University. (Doctoral dissertation).
- Berman, Ruth A. & Dan I. Slobin. 2013. Relating events in narrative: A crosslinguistic developmental study. In 17–35. New York: Psychology Press.
- Biber, Douglas & Susan Conrad. 2019. *Register, genre, and style*. 2nd edn. Cambridge University Press. DOI: 10.1017/9781108686136.
- Biber, Douglas, Bethany Gray & Kornwipa Poonpon. 2011. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly* 45(1). 5–35.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.
- Dussias, Paola E. 2004. Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English bilinguals. *International Journal of Bilingualism* 8(3). 355–371.
- Gogolin, Ingrid. 2006. Bilingualität und die Bildungssprache der Schule. In *Die Macht der Sprachen*, 79–85. Münster: Waxmann.
- Gogolin, Ingrid. 2013. Mehrsprachigkeit und bildungssprachliche Fähigkeiten. In Ingrid Gogolin, Imke Lange, Ute Michel & Hans H. Reich (eds.), *Herausforderung Bildungssprache – und wie man sie meistert*, 7–18. Münster: Waxmann.
- Goschler, Juliana. 2017. Syntaxerwerb bei ein- und mehrsprachigen Grundschüler/innen: Eine quantitative Untersuchung [The acquisition of syntactic structures of children in primary school: A quantitative study]. *Zeitschrift für Angewandte Linguistik* 66. 59–77. DOI: 10.1515/zfal-2017-0002.
- Grant, Anthony P. 2020. Contact and language convergence. In Raymond Hickey (ed.), *The handbook of language contact*, 2nd edn., 113–128. Hoboken, NJ: Wiley.
- Grenoble, Lenore A. 2013. The syntax and pragmatics of Tungusic revisited. In *Language typology and historical contingency: In honor of Johanna Nichols*, 357–382. John Benjamins.
- Grimm, Angela & Petra Schulz. 2014. Sprachfähigkeiten von Kindern mit DaZ bei Schuleintritt. In B. Lütke & I Petersen (eds.), *Deutsch als Zweitsprache: erwerben, lernen und lehren. Beiträge zum 9. Workshop Kinder mit Migrationshintergrund*, 35–50. Stuttgart: Klett Fillibach.
- Haberzettl, Stefanie. 2016. Bildungssprache im Kontext von Mehrsprachigkeit: Eine Untersuchung von Berichtstexten ein- und mehrsprachiger Schüler. *Diskurs Kindheits- und Jugendforschung/Discourse: Journal of Childhood and Adolescence Research* 11(1). 61–79.

- Haspelmath, Martin. 2004. *Coordinating constructions* (Typological studies in language 58). Amsterdam: John Benjamins.
- Haspelmath, Martin. 2007. Coordination. In Timothy Shopen (ed.), *Language typology and syntactic description*, 2nd edn., vol. 2: Complex constructions, 1–51. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511619434.001.
- Herkenrath, Annette, Birsel Karakoç & Jochen Rehbein. 2003. Interrogative elements as subordinators in Turkish. In Natascha Müller (ed.), *(In)vulnerable domains in multilingualism*, 221–269. Amsterdam: John Benjamins. DOI: 10.1075/hsm.1.09her.
- Higby, Eve, Jungna Kim & Loraine K. Obler. 2013. Multilingualism and the brain. *Annual Review of Applied Linguistics* 33. 68–101.
- Housen, Alex, Bastien De Clercq, Folkert Kuiken & Ineke Vedder. 2019. Multiple approaches to complexity in second language research. *Second Language Research* 35(1). 3–21.
- Iefremenko, Kateryna, Christoph Schroeder & Jaklin Kornfilt. 2021. Converbs in heritage Turkish: A contrastive approach. *Nordic Journal of Linguistics* 44(2). 130–154. DOI: 10.1017/S0332586521000160.
- Johanson, Lars. 1992. *Strukturelle Faktoren in türkischen Sprachkontakten* (Sitzungsberichte der wissenschaftlichen Gesellschaft an der Johann Wolfgang Goethe-Universität Frankfurt am Main). Stuttgart: Steiner.
- Johanson, Lars, Eva Csato & Birsel Karakoç. 2020. Turkic language contacts. In Raymond Hickey (ed.), *The handbook of language contact*, 2nd edn., 551–570. Hoboken, NJ: Wiley.
- Joseph, Brian D. 2020. Greek infinitive-retreat versus grammaticalization: An assessment. In Dariya Rafiyenko & Ilja A. Seržant (eds.), *Postclassical Greek: Contemporary approaches to philology and linguistics*, 145–162. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110677522-007.
- Karakoç, Birsel. 2007. Connectivity by means of finite elements in monolingual and bilingual Turkish discourse. In Jochen Rehbein, Christiane Hohenstein & Lukas Pietsch (eds.), *Connectivity in grammar and discourse* (Hamburg Studies on Multilingualism 5), 199–227. Amsterdam: John Benjamins.
- Katsika, Kalliopi, Annika Labrenz, Kateryna Iefremenko & Shanley E.M. Allen. 2025. Discourse openings and closings across languages in contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 527–576. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775189.
- Kerslake, Celia. 2007. Alternative subordination strategies in Turkish. In Jochen Rehbein, Christiane Hohenstein & Lukas Pietsch (eds.), *Connectivity in gram-*

- mar and discourse* (Hamburg Studies on Multilingualism 5), 231–258. Amsterdam: John Benjamins.
- Keskin, Cem. 2023. Transient subordinate clauses in Balkan Turkic in its shift to Standard Average European subordination: Dialectal and historical evidence. *Folia Linguistica* (57). 155–197.
- Keskin, Cem. 2025. Balkan Turkic as a model for understanding contact-induced change in Turkish. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 287–323. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775175.
- Keskin, Cem, Kateryna Iefremenko, Jaklin Kornfilt & Christoph Schroeder. 2024. Hybrid clause combining strategies in Turkish language contacts. *Studia Linguistica* (Early View). DOI: 10.1111/stul.12256.
- König, Ekkehard & Volker Gast. 2018. *Understanding English-German contrasts*. 4th edn. Berlin: Erich Schmidt Verlag.
- Kortmann, Bernd. 1995. Adverbial participial clauses in English. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms – Adverbial participles, gerunds* (Empirical Approaches to Language Typology 13), 189–238. Berlin: Mouton De Gruyter.
- Kožina, Margarita N. 2011. *Stilistika russkogo jazyka*. FLINTA.
- Kupriianova, T. F. 2002. *Знакомьтесь: Причастие: Учебное пособие для изучающих русский язык (продвинутый этап)* (*znakom'tes': Pričastie: Učebnoe posobie dlja izučajuščih russkij jazyk (prodvinutyyj ètap)*). Saint Petersburg: Златоуст (Zlatoust).
- Labov, William & Joshua Waletzky. 1997. Narrative analysis: Oral versions of personal experience. *Journal of Narrative & Life History* 7(1–4). 3–38.
- Labrenz, Annika, Kateryna Iefremenko, Kalliopi Katsika, Shanley E.M. Allen, Christoph Schroeder & Heike Wiese. 2025. Dynamics of discourse markers in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 493–526. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775187.
- Lambrou, Marina. 2005. *Story patterns in oral narratives: A variationist critique of labov and waletzky's model of narrative schemas*. Middlesex University. (Doctoral dissertation).
- Lehmann, Christian. 1988. Towards a typology of clause linkage. In John Haiman & Sandra A. Thompson (eds.), *Clause combining in grammar and discourse*

- (Typological Studies in Language 18), 181–225. Amsterdam: John Benjamins.  
DOI: 10.1075/tsl.18.09leh.
- Lehmann, Nico, Dina Serova, Julia Lukassek, Sophia Döring, Frank Goymann, Anke Lüdeling & Roodabeh Akbari. 2023. Guidelines for the annotation of parameters of narration. *Register Aspects of Language in Situation* 2(5). 1–42.  
DOI: 10.18452/26437.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Anniqa Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehaye, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Maas, Utz. 2008. *Sprache und Sprachen in der Migrationsgesellschaft: Die schriftkulturelle Dimension*. (Schriften des Instituts für Migrationsforschung und Interkulturelle Studien (IMIS) der Universität Osnabrück 15). Göttingen: V&R unipress.
- Madariaga, Nerea. 2011. Infinitive clauses and dative subjects in russian. *Russian Linguistics* 35(3). 301–329.
- Martynova, Maria, Yulia Zuban, Natalia Gagarina & Luka Szucsich. 2024. Use of embedded clauses in heritage and monolingual russian. *Languages* 9(5). 157.
- Matras, Yaron. 2009. *Language contact*. 1st edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Matras, Yaron. 2020. *Language contact*. 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Miller, James Edward & Regina Weinert. 1998. *Spontaneous spoken language: Syntax and discourse*. Oxford: Oxford University Press.
- Mithun, Marianne. 1988. The grammaticalization of coordination. In John Haiman & Sandra A. Thompson (eds.), *Clause combining in grammar and discourse* (Typological Studies in Language), 332–359. Amsterdam: John Benjamins.
- Nedjalkov, Vladimir P. 1995. Some typological parameters of converbs. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms – Adverbial participles, gerunds* (Empirical Approaches to Language Typology 13), 97–136. Berlin: Mouton De Gruyter.

- Ögel-Balaban, Hale & Ayhan Aksu-Koç. 2020. Clause chaining and discourse continuity in Turkish children's narratives. *Frontiers in Psychology* 11. 115. DOI: 10.3389/fpsyg.2020.00115.
- Onar Valk, Pelin. 2015. *Transformation in Dutch Turkish subordination? Converging evidence of change regarding finiteness and word order in complex clauses*. Utrecht: LOT Publications.
- Özsoy, Onur & Frederic Blum. 2023. Exploring individual variation in Turkish heritage speakers' complex linguistic productions: Evidence from discourse markers. *Applied Psycholinguistics* 44(4). 534–564. DOI: 10.1017/S0142716423000267.
- Özsoy, Onur, Kateryna Iefremenko & Christoph Schroeder. 2022. Shifting and expanding clause combining strategies in heritage Turkish varieties. *Languages* 7(3). 242. DOI: 10.3390/languages7030242.
- Pashkova, Tatiana, Wintai Tsehayé, Shanley E. M. Allen & Rosemarie Tracy. 2022. Syntactic optionality in heritage language use: Clause type preferences of German heritage speakers in a majority English context. *Heritage Language Journal* 19(1). 1–41. DOI: 10.1163/15507076-12340022.
- Polinsky, Maria. 2008. Heritage language narratives. In Donna M. Brinton, Olga Kagan & Susan Bauckus (eds.), *Heritage language education: A new field emerging*, 149–164. New York: Routledge.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Roussou, Anna. 2000. On the left periphery: Modal particles and complementizers. *Journal of Greek Linguistics* 1(1). 65–94.
- Sanchez Abchi, Veronica Soledad & Mariela Vanesa de Mier. 2017. Syntactic complexity in narratives written by Spanish heritage speakers. *Vial: Vigo International Journal of Applied Linguistics* 14. 125–148.
- Sarvary, Hannah S. & Soonja Choi. 2020. Acquisition of clause chaining. *Frontiers in Psychology* 11. 583431. DOI: 10.3389/fpsyg.2020.583431.
- Schleppegrell, Mary J. 2004. *The language of schooling: A functional linguistics perspective*. Ney York: Routledge.
- Schmidt, Thomas & Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *Handbook on corpus phonology*, 402–419. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199571932.013.030.
- Schroeder, Christoph. 2002. On the structure of spoken Turkish. *Essener Linguistische Skripte – elektronisch* 2(1). 73–90. [https://www.unidue.de/imperia/md/content/elise/ausgabe\\_1\\_2002\\_Schroeder.pdf](https://www.unidue.de/imperia/md/content/elise/ausgabe_1_2002_Schroeder.pdf).

- Schroeder, Christoph. 2016. Clause combining in Turkish as a minority language in Germany. In Mine Güven, Didar Akar, Balkiz Öztürk & Meltem Kelepir (eds.), *Exploring the Turkish linguistic landscape: Essays in honor of Eser Erguvanlı-Taylan* (Studies in Language Companion Series), 81–102. Amsterdam: John Benjamins. DOI: 10.1075/slcs.175.06sch.
- Schroeder, Christoph, Kateryna Iefremenko & Mehmet Öncü. 2024. The postverbal position in heritage Turkish: A comparative perspective with a focus on non-clausal elements. In Zeynep Kalkavan-Aydin & Yazgül Şimşek (eds.), *Deutsch-Türkische Zweisprachigkeit mit besonderem Fokus auf Jugendliche [Turkish-German bilingualism with a special focus on adolescents]* (Mehrsprachigkeit [Multilingualism] 58), 109–132. Münster: Waxmann.
- Schulz, Petra & Rosemarie Tracy. 2011. *Linguistische Sprachstandserhebung: Deutsch als Zweitsprache (LiSe-DaZ): Language test for children with German as a second language*. Göttingen: Hogrefe.
- Shagal, Ksenia, Pavel Rudnev & Anna Volkova. 2022. Multifunctionality and syncretism in non-finite forms: An introduction. *Folia Linguistica* 56(3). 529–557.
- Treffers-Daller, Jeanine, A. Sumru Özsoy & Roeland van Hout. 2006. Oral language proficiency of Turkish–German bilinguals in Germany and Turkey: An analysis of complex embeddings in Turkish picture descriptions. In Hendrik E. Boeschoten, Lars Johanson, Sevgi Ağcagül & Vildan Milani (eds.), *Turkic languages in contact*, 203–219. Wiesbaden: Harrassowitz.
- Tsehayre, Wintai. 2023. Light-weights placed right: Post-field constituents in heritage German. *Frontiers in Psychology* 14. DOI: 10.3389/fpsyg.2023.1122129.
- Tsimpli, Ianthi-Maria. 2000. Gerunds in Greek. *Journal of Greek Linguistics* 1(1). 133–169.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehayre, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Lin-*

- guistic dynamics in heritage speakers: Insights from the RUEG group*, 33–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775159.
- Winford, Donald. 2003. *An introduction to contact linguistics* (Language in Society 33). Malden, MA: Blackwell.
- Yagmur, Kutlay. 1997. *First language attrition among Turkish speakers in Sydney* (Studies in Multilingualism 7). Tilburg: Tilburg University Press.
- Yilmaz-Wörfel, Seda. 2022. *Adverbial relations in Turkish-German bilingualism* (Mehrsprachigkeit/Multilingualism 53). Münster: Waxmann.

# Chapter 9

## Balkan Turkic as a model for understanding contact-induced change in Turkish

Cem Keskin

University of Potsdam, Leibniz-Centre General Linguistics (ZAS)

How good a model is the Standard Average European–Turkic contact in the Balkans for understanding contact-induced syntactic changes in Turkish in the West? Patterns of head directionality change and innovations in clause combining replicated across several contact situations suggest that the answer to this question is a positive superlative. However, there is no single homogeneous response that Turkic varieties give under language contact but a complex pattern. Also the causal factors behind this pattern are fairly general in nature. So, a more reasonable conclusion that suggests itself is that Balkan Turkic constitutes a good model for understanding contact-induced syntactic change in Turkish in the West insofar as it can help reveal general contact dynamics, not as a Turkic contact variety per se. And on that front the in-depth investigation of Balkan Turkic promises to deliver useful observations.

### 1 Introduction

As is well-known, a significant new era of intensive Standard Average European<sup>1</sup> (SAE)–Turkic contacts can be said to have begun with the arrival of the so-called guest workers in northwest Europe in the late 1950s and early 1960s (Küçükcan & Güngör 2009: 83–84). This new contact constellation naturally stimulated a large

<sup>1</sup>The term *Standard Average European* (originally from Whorf 1944) refers in recent studies to a proposed sprachbund that spans Romance, Germanic, Balto-Slavic, the Balkan languages, etc. (see e.g., Haspelmath 1998, 2001, van der Auwera 2011).



body of linguistic research on the changes that were beginning to be observed in Turkish in the region.

One lacuna in this literature has been the lack of references to other earlier SAE–Turkic contact situations as potential models of contact-induced change in Turkic that could help us understand emerging changes in Turkish in the new contact constellation in northwest Europe. This would have been a reasonable approach, given that Turkic varieties have been in contact with SAE languages for several centuries, mostly in the Balkan sprachbund, and have undergone extensive well-documented changes (see Johanson 2021: Ch. 10 for an overview and Johanson 1992, 2013 for a detailed account).

This is the gap that I attempt to fill in this chapter. More precisely, I try to answer the question of how good a model other SAE–Turkic contact situations (more specifically SAE–Turkic contact in the Balkans) can be when trying to understand contact-induced syntactic changes in Turkish in the West. Let me reveal the answer from the beginning: Balkan Turkic (BT) constitutes a good model for understanding contact-induced syntactic change in Turkish in the West insofar as it can help reveal general contact dynamics, not necessarily by virtue of being a Turkic contact variety.<sup>2</sup> In the process of detailing this answer, I will be presenting three sets of findings: (i) those on contact-induced changes in BT,<sup>3</sup> (ii) those on contact-induced changes in heritage Turkish as spoken in Germany and the US and the Turkish of Kurmanji–Turkish bilinguals, and (iii) monolingual Turkey’s Turkish.<sup>4</sup>

The chapter is made up of two main sections, which will follow the next section on data sources, method, and statistics: Section 3 is on contact-induced head directionality changes, and Section 4 details contact-induced changes in clause combining with a focus on subordination. Both sections begin with data from BT, later turning to data from the other contact varieties and showing the connections between the former and the latter.

---

<sup>2</sup>I prefer to use the more general term *Balkan Turkic*, rather than the *Balkan Turkish* of most works in Turkological literature (e.g. Johanson 2021). The reason for this is the following: Even though the Southwest (or Oghuz) Turkic varieties spoken in the Balkans, which the present paper focuses on, are very closely related to Turkish varieties in the east, it is not entirely clear to what degree they have diverged and so whether at least some should be considered separate languages or can still be seen as dialects of Turkish.

<sup>3</sup>Note, however, that I will not be concerned with the so-called Balkanisms here (see e.g. Joseph 2020: 540–541), but only with the more general SAE features seen in BT.

<sup>4</sup>It should be emphasized that the latter two sets of findings are the outcome of extensive collaboration with Kateryna Iefremenko, Christoph Schroeder, and to a lesser extent Jaklin Kornfilt.

## 2 Data sources, method, and statistics

The data used in the analysis of BT come from the Balkan Turkic Corpus (Keskin et al. in preparation). The texts in this corpus, totaling around 80,000 words, were culled from the following sources:

1. Dialect texts
  - a) West Rumelian Turkish
    - i. Kosovar Turkish: personal accounts in Sulçevsi (2019)
    - ii. Macedonian Turkish: folk tales in Destanov (2016) and Kakuk (1972), folk tales and personal accounts in Katona (1969)
    - iii. Western Bulgarian Turkish: folk tales and accounts of traditions in Kakuk (1961a,b)
  - b) North Rumelian Turkic
    - i. Gagauz: Folk tales selected from numerous sources and published by Özkan (2007)
    - ii. Dobruja Turkish: miscellaneous texts in Haliloglu (2017)
  - c) East Rumelian Turkish: folk tales and accounts of traditions in Hazai (1960) and Kakuk (1958)<sup>5</sup>
2. Historical texts that show Early Balkan Turkic features (the so-called “transcription texts”)
  - a) 14th century: Schiltberger’s Our Father published by Helmholdt (1966)
  - b) 15th century: Yusof and Jakob Papas’ letters published by Brendemoen (1980), Pietro Bruto and Adriano Fino’s Bible verses in Weil (1953)

---

<sup>5</sup>These two sources provide only two small samples from central and southern Bulgaria, selected as representatives of East Rumelian as a whole. Data from the Turkish of Western Thrace (see e.g. Petrou 2021) and elsewhere were not included in the East Rumelian part of the corpus for two reasons. First, the East Rumelian group does not show the syntactic changes that were of interest to the research on which this chapter is based, so a small control sample was judged to be sufficient. Second, some of these texts (particularly the more recent ones in Petrou 2021) were not available during the text collection phase of the corpus. More East Rumelian data are planned to be added in a later version of the corpus.

- c) 16th century: Filippo Argenti's phrases in Adamović (2001), Bartholomaeus Georgievits' dialogue, Our Father, the Apostles' Creed, etc. in Heffening (1942), Marco Antonio Begliarmati's dialogue in Teza (1892), the anonymous phrases in Adamović (1976), Guillaume Postel's phrases in Drimba (1966), Reinhold Lubenau's phrases in Adamović (1977)
- d) 17th century: Pietro Ferraguto's dialogue in Bombaci (1940), Giovan Battista Montalbano's sayings in Gallotta (1986), sample text in Du Ryer (1630), the anonymous dialogue in Blau (1868), Miklós Illésházy's dialogue in Németh (1970), dialogue and Our Father in Herbinius (1675)

These texts were first coded sentence by sentence using a sentence annotation interface for the features below, and the features were stored in a database:

1. Directionality: bare object–verb versus verb–bare object, oblique–verb versus verb–oblique, etc. (a total of 21 pairs of opposing features)
2. Clause type: main, argument, relative, adverbial
3. Finiteness: finite, non-finite
4. Metadata: century, author, genre, provenance

The texts were then analyzed based on directionality, clause type, finiteness, century, and provenance, using a query module that uses the stored grammatical properties.

In addition to Balkan Turkic data, four other sets of data were collected for the studies reported on here following the Language Situations method by Wiese (2020): (i) Turkish with German contact in Germany ( $n = 65$ ), (ii) Turkish with English contact in the US ( $n = 61$ ), (iii) with Kurmanji contact in Turkey ( $n = 30$ ), and (iv) monolingual Turkish ( $n = 66$ ). The Language Situations method brings controlled elicitation together with spontaneous speech and is suited to comparing various contact constellations. It allows for the collection of quasi-naturalistic data across different communicative situations, including formal and informal settings and written and spoken communication. In accordance with this method, participants were shown the video of a car accident and asked to describe it in four different imagined scenarios: (i) WhatsApp voice message to a friend, (ii) WhatsApp text message to a friend, (iii) voicemail to the police, and (iv) written witness report to the police. Each bilingual participant had two

sessions, one in their heritage language and one in the majority language of the society in which they lived.

A few remarks are in order about these data sources. First, the historical sources in the Balkan Turkic Corpus comprise almost all the transcription texts available to date. As the volume of these texts is fairly low, the only selection criterion applied (in addition to whether or not the text showed BT features) was the intelligibility of the author's orthography. Also, some of these sources are religious texts (e.g. translations of the Bible, etc.), and one might wonder whether these texts and their translations are inclined towards contact forms. Indeed, the religious texts in the corpus do have a stronger tendency towards contact forms than the secular sources, but the skewing effect of this can be considered negligible, as religious texts constitute only about 10% of the sources in the corpus in terms of word count. (Genre in general seems not to have a significant effect, but word count per genre is too low to reach a definitive conclusion.) Second, one might wonder whether it is legitimate to compare the material in the Balkan Turkic Corpus with the data collected using the "Language Situations" method. To clarify, the non-Gagauz material in the corpus consists of transcripts of spontaneous oral productions of informants, while the Gagauz texts seem to have been minimally edited for readability and preserve the original style of the narrators as much as possible. So, even though the two sets of data are not products of identical methodologies, this is as close as one can presently get in such a comparative study, and that seems to be an appropriate approximation.

As for statistical tests and basic mathematical operations, three separate tools were used for these: (i) Lancaster Stats Tools online that runs R code (Brezina 2018), (ii) R (version 4.3.0) (R Core Team 2021), and (iii) MS Excel (version 2202).

### 3 Contact-induced head directionality changes

#### 3.1 From head-final to head-initial

A number of syntactic changes associated with a possible shift from head-final to head-initial syntax that appear to be taking place in heritage Turkish in contact with German and English have been subject to research. Some of this research focuses on the use of the postverbal position in the canonically verb-final Turkish (see e.g. Iefremenko 2021c, Schroeder 2018, Schroeder & Iefremenko 2019b,c, Schroeder et al. 2024, see also Bunk et al. 2025 [this volume] and Iefremenko 2019, 2021a,b) and innovative clause combining strategies. The latter I will defer to Section 4 and focus here on directionality in the verbal domain, supplementing

my observations with additional data from noun phrases, adpositional phrases, and clauses.

If a shift from head-final to head-initial syntax is indeed the path that heritage Turkish is treading, it would not be the first time in the Turkic family, Macedonian Turkish (see e.g. Matras & Tufan 2007), Gagauz (see e.g. Menz 1999), Karaim (see e.g. Csató 2000), and Urum (see e.g. Böhm 2015, Skopeteas 2015) being its most well-known members which already show features of the shift, to varying degrees, as can be gleaned from the literature (see e.g. Balci 2010, Doerfer 1959: 271, Friedman 1982: 33–35, Friedman 2006: 40–41, Gülsevin 2017: 110–111, Günşen 2010: 471–475, İğci 2010: 129, Jable 2010: 148–49, Johanson 2021: 790, 936, Katona 1969: 165, Kirli 2001, Matras 2009: 251, Matras & Tufan 2007, Menz 1999: 40, Menz 2014: 61–62, Özkan 1996: 209–210; see also Petrou 2019: 226 and Petrou 2021: 342 for some initial observations on the shift to head-initial syntax in the Turkish of Western Thrace in contact with Greek).

The current state of heritage Turkish, then, appears to reveal the initial stages of a syntactic shift in the same direction as the head-initial varieties of Turkic, and is probably comparable to (i) Turkic contact varieties in their earlier periods and (ii) Turkic contact varieties that are more on the head-final pole of a range that extends to head-initiality. Thus, a systematic comparative approach promises to help us map out possible trajectories of syntactic change in Turkic, provide insights into the changes taking place in heritage Turkish at present, and perhaps even predict the changes that might occur in the future.

Following this rationale, I will be exploiting, in what follows, two avenues for comparison between BT and heritage Turkish: (i) a dialectological angle that investigates the available range of modern BT varieties with varying preferences for head-final versus head-initial orders, and (ii) a diachronic angle with an eye on older forms of BT on the way to head-initiality.

However, as an anonymous reviewer rightly points out, BT and heritage Turkish in Germany or the US are and have been in very different sociolinguistic settings, meaning that it would be unreasonable to expect wholesale convergences between them. I will mention one sociolinguistic consideration. The contact varieties in the West maintain contacts with Turkey's Turkish, and their speakers still regard the latter as the “proper” variety – even if this proper variety cannot fully assert its influence. Even with these reservations in mind (e.g. the linguistic orientation of later generations, the degree of access to the written standard), this situation is very different than that of BT, for which there has never been a strong Turkic standard towards which the speakers could orient themselves, even during the Ottoman era (Johanson 1989). These differences will no doubt

have their consequences with respect to the patterns that the newer contact varieties will show in the longer term, as well as the timelines of these patterns. Yet, I believe there is still some merit in comparing these contact varieties simply by the fact of their all being Turkic varieties in contact with SAE, and, as we will see during the discussion of the findings, the results seem to justify this approach.

After this introductory discussion, I now move on to detailed descriptions of contact-induced head directionality changes in the relevant contact varieties where I will present detailed data on the constituent orders of 13 head and dependent pairs. As will soon become clear, the data show substantial amounts of variation across syntactic structures and varieties, and reveal spectrums of head directionality (Sections 3.2.1 and 3.2.2), partly brought about by an interplay of progressive and conservative syntactic domains (Section 3.2.3).<sup>6</sup> In addition to this variation, another aspect of these syntactic patterns is what we could call *variability*, which can be briefly described as follows: Progressive and conservative domains are not equally so. As dialects and syntactic pairs shift from head-final to head-initial order, they diverge from each other in the degree to which they do that, presenting a heterogeneous pattern of changes (Section 3.2.4). This multi-faceted pattern can be attributed to the interaction of several factors, such as degree of contact and the differential responses of syntactic structures (Section 3.3).

## 3.2 Change, variation, and variability

### 3.2.1 Two spectrums of directionality in Balkan Turkic

Then, let us begin by addressing the following question: What kind of word order changes does BT show due to SAE contact? As pointed out in the introductory discussion, the answer to this question will set a standard of comparison for the Turkish contact varieties in the West. But before getting into the details of the answer, let me provide in (1) an illustrative example of the kinds of changes that I will be discussing:

- (1) a. Ben ak ekmek al-dī-m.  
1SG white bread buy-PST-1SG  
'I bought some white bread.' (Hazai 1960: 187)
- b. Çocuk al-mış tuz.  
child take-EVID.3SG salt  
'The child took some salt.' (Özkan 2007: 137)

---

<sup>6</sup>The terms *progressive*, *advanced*, *conservative*, etc. contain no value judgments and are intended only to describe the degree of and the responses to the shift to head-initial order.

Example (1a) is from the dialect group referred to as *East Rumelian Turkish*, which is spoken in the southeastern Balkans and known for its conservative syntactic features (see e.g. Johanson 2021: 49). The example instantiates canonical Turkic word order in the verbal domain with the bare direct object *ak ekmek* ‘white bread’ preceding the verb *al* ‘buy’. It exemplifies, in other words, the dependent-head constituent order typical of Turkic. Example (1b) is from Gagauz, spoken mostly in Moldova and known for its pervasive contact-induced syntactic features (see e.g. Johanson 2021: 50–51). It contains a head-initial VP, in which the direct object *tuz* follows the verb *al* ‘take’. That is to say, it exemplifies the head-dependent constituent order commonly seen in Turkic varieties going through contact-induced word order shifts.

Comparable shifts (as in Gagauz) or lack thereof (as in East Rumelian) form an overall pattern summarized in Table 1, which shows the distribution of thirteen head-dependent orders in main clauses across BT (see Keskin 2023a for another, similar account of word order changes in BT).

The head-dependent combinations are indicated in the top row.<sup>7</sup> They are sorted on the total percentages of head-initial orders across 25 BT dialect locales (given in the bottommost row) in increasing order from left to right. Each cell indicates the percentage of head-initial order per given head+dependent combination in a given dialect locale. Dialect locales are indicated in the leftmost column, sorted on the total percentages of all head-initial orders observed in those locales (given in the rightmost column) in increasing order from top to bottom.

Table 1 immediately makes the intricacies involved in answering the question above clear. Two spectrums of directionality can be discerned in the pattern it displays, contingent upon (i) type of head and dependent, and (ii) dialect locale. The interaction of these two spectrums, thus, brings about a rather complex pattern of change.<sup>8</sup>

Let us begin with the spectrum of head-dependent pairs. First, we see that there is a gradual transition from very low to very high incidences of head-dependent orders as we move rightward – a transition from head-final to head-initial order, in other words. Clausal dependents (i.e. relative clauses, finite complement and adverbial clauses, and non-finite complement and adverbial clauses)

<sup>7</sup>NAdj=noun–adjective, NGen=noun–genitive, PN=adposition–noun, VAdv=verb–adverb, VSubj=verb–subject, VPP=verb–adpositional phrase, VBar=verb–bare object, VNFin=verb–non-finite dependent clause, VAcc=verb–accusative-marked object, VObl=verb–oblique object, VFin=verb–finite dependent clause, CCl=connector–dependent clause, NRC=noun–relative clause, HD=head–dependent order

<sup>8</sup>Color coding was used in the table for ease of interpretation. Gray cells show cases for which I only have a few examples.

Table 1: Head-dependent orders in Balkan Turkic

|            | HEAD-FINAL |      |     |      |       |     |      |       |      |          | ⇒    | HEAD-INITIAL |      |     |  | HD |
|------------|------------|------|-----|------|-------|-----|------|-------|------|----------|------|--------------|------|-----|--|----|
|            | NAdj       | NGen | PN  | VAdv | VSubj | VPP | VBar | VNFin | VAcc | VOblique | VFin | CCI          | NRC  |     |  |    |
| Mamushë    | 0%         | 3%   | 7%  | 0%   | 17%   | 27% | 9%   | 80%   | 38%  | 24%      | 43%  | 56%          | 83%  | 16% |  |    |
| Kazanlăk   | 0%         | 1%   | 0%  | 3%   | 13%   | 25% | 16%  | 17%   | 28%  | 34%      | 41%  | 32%          | 0%   | 18% |  |    |
| Gjilani    | 3%         | 0%   | 0%  | 3%   | 3%    | 11% | 26%  | 0%    | 63%  | 18%      | 56%  | 94%          | 100% | 18% |  |    |
| Kărdžali   | 0%         | 2%   | 0%  | 12%  | 23%   | 25% | 15%  | 9%    | 48%  | 44%      | 15%  | 32%          | 0%   | 20% |  |    |
| Resen      | 0%         | 4%   | 0%  | 13%  | 5%    | 18% | 8%   | 50%   | 22%  | 29%      | 76%  | 69%          | 67%  | 23% |  |    |
| Razgrad    | 0%         | 6%   | 0%  | 11%  | 16%   | 24% | 14%  | 25%   | 44%  | 45%      | 50%  | 58%          | 33%  | 24% |  |    |
| Prizreni   | 3%         | 8%   | 6%  | 6%   | 19%   | 44% | 17%  | 0%    | 38%  | 47%      | 50%  | 94%          | 100% | 26% |  |    |
| Ohrid      | 0%         | 0%   | 0%  | 11%  | 9%    | 13% | 31%  | 40%   | 18%  | 25%      | 68%  | 85%          | 50%  | 26% |  |    |
| Jeni Maale | 0%         | 9%   | 0%  | 26%  | 24%   | 36% | 39%  | 38%   | 62%  | 46%      | 41%  | 63%          | 0%   | 28% |  |    |
| Ali Koč    | 0%         | 0%   | 0%  | 11%  | 21%   | 67% | 14%  | 36%   | 46%  | 63%      | 29%  | 65%          | 0%   | 29% |  |    |
| Janevë     | 3%         | 0%   | 0%  | 6%   | 41%   | 67% | 41%  | 100%  | 50%  | 68%      | 55%  | 75%          | 100% | 31% |  |    |
| Prishtinë  | 3%         | 3%   | 5%  | 10%  | 11%   | 38% | 58%  | 0%    | 59%  | 57%      | 71%  | 75%          | 100% | 32% |  |    |
| Konče      | 0%         | 11%  | 0%  | 19%  | 24%   | 22% | 37%  | 33%   | 68%  | 53%      | 56%  | 69%          | 50%  | 34% |  |    |
| Dobrčane   | 0%         | 13%  | 9%  | 24%  | 15%   | 35% | 54%  | 0%    | 65%  | 55%      | 77%  | 62%          | 100% | 36% |  |    |
| Pejë       | 3%         | 4%   | 25% | 29%  | 2%    | 50% | 84%  | 100%  | 56%  | 65%      | 27%  | 81%          | 50%  | 36% |  |    |
| Chişinău   | 0%         | 3%   | 6%  | 12%  | 17%   | 38% | 38%  | 59%   | 63%  | 57%      | 74%  | 88%          | 100% | 36% |  |    |
| Vuştrri    | 2%         | 14%  | 8%  | 9%   | 33%   | 50% | 67%  | 50%   | 45%  | 72%      | 73%  | 91%          | 100% | 38% |  |    |
| Montana    | 1%         | 4%   | 0%  | 23%  | 32%   | 14% | 35%  | 17%   | 69%  | 70%      | 77%  | 82%          | 100% | 39% |  |    |
| Kyustendil | 2%         | 14%  | 0%  | 5%   | 22%   | 57% | 41%  | 43%   | 70%  | 76%      | 73%  | 78%          | 25%  | 40% |  |    |
| Mitrovica  | 13%        | 7%   | 20% | 18%  | 18%   | 70% | 55%  | 0%    | 64%  | 58%      | 72%  | 75%          | 100% | 41% |  |    |
| Bučim      | 2%         | 30%  | 0%  | 39%  | 25%   | 20% | 34%  | 60%   | 65%  | 63%      | 68%  | 84%          | 33%  | 43% |  |    |
| Odesa      | 0%         | 0%   | 0%  | 18%  | 24%   | 33% | 50%  | 50%   | 79%  | 81%      | 77%  | 90%          | 75%  | 45% |  |    |
| Bessarabia | 0%         | 5%   | 0%  | 25%  | 19%   | 67% | 74%  | 53%   | 73%  | 74%      | 85%  | 73%          | 33%  | 45% |  |    |
| Ossetia    | 0%         | 6%   | 33% | 27%  | 31%   | 33% | 57%  | 69%   | 74%  | 88%      | 55%  | 76%          | 100% | 47% |  |    |
| Tomai      | 0%         | 22%  | 30% | 20%  | 42%   | 33% | 67%  | 63%   | 79%  | 70%      | 80%  | 100%         | 100% | 51% |  |    |
| HD         | 1%         | 5%   | 5%  | 14%  | 18%   | 35% | 35%  | 46%   | 53%  | 54%      | 59%  | 74%          | 74%  |     |  |    |

appear to be the most progressive constituents as a category. (Non-finite clauses, however, are markedly less progressive.) By contrast, adpositional phrases and noun phrases with non-clausal dependents are the most conservative group. Finally, verbs with non-clausal dependents (i.e. obliques, bare objects, etc.) can be said to be distributed over the middle range.

As for the spectrum of dialects, we see that some dialects, such as Kărdžali and Kazanlăk (members of the dialect group East Rumelian Turkish), are conservative. Some others, such as Tomai and Chişinău (varieties of Gagauz), on the other hand, are progressive.

### 3.2.2 Spectrums of directionality in other contact varieties

Let us now turn to the word order changes in the contact varieties in the West. To repeat, these are Turkish with German contact in Germany (abbreviated as *DEbi*) and with English contact in the US (*USbi*) (which are considered heritage languages). I will also include Turkish with Kurmanji contact in Turkey<sup>9,10</sup> (*TUbi*, spoken by Kurmanji–Turkish bilinguals, not a heritage language) and Early Balkan Turkic (EBT, 14th–17th c.)<sup>11</sup> in the analysis for reference, and will refer to *DEbi*, *USbi*, and *TUbi* as the *new contact varieties* (NCV). The findings are summarized in Table 2 (see also Iefremenko et al. 2023 for another similar analysis).

To begin, it is interesting to note that, broadly speaking, the same pattern of word order changes replicates itself across four different contact situations (Table 2a). As in BT, noun and adposition phrases are the most conservative domains

<sup>9</sup>Some grammatical features of Kurmanji that are relevant for this study are as follows (extracted from the relevant sections in Haig & Öpengin 2018 and McCarus 2009). In the nominal domain, possessors, adjectives, and relative clauses follow, while demonstratives and numerals precede the head noun. In the verbal domain, obliques and complement clauses follow the verb, while direct objects precede it. Adverbs can precede or follow the verb, and adverbial clauses can precede the main clause or be postverbal. Conjunctions that introduce clauses are clause-initial. The language makes use of prepositions and circumpositions, as well as locational nouns that precede their dependents.

<sup>10</sup>Turkish and Kurmanji have been in contact since at least the 11th century, but the extent of present-day contact is unprecedented, determined by the language policies of the Turkish state in the post-1923 period in conjunction with the rise of mass media (see e.g. Bulut 2006: 96, 116, Dorleijn 2006: 74, Haig 2017: 398–399, Haig & Öpengin 2018: 222–225, Johanson 2006: 8, Varol 2014: 59, 65). To cite specific but limited figures which point in this direction, the percentage of Kurmanji–Turkish bilinguals in Turkey seems to have increased from around 26% to around 40% just in the 1935–1965 period, the only period from which we have relevant data, collected during the censuses (census data available in Dündar 2000: 171–216). For this reason, I assume that the phenomena that we observe in the Turkish contact variety spoken by Kurmanji–Turkish bilinguals are by and large relatively new developments.

<sup>11</sup>An anonymous reviewer questions the power of transcription texts to represent EBT. The question of whether transcription texts can be considered reliable sources for the historical linguistics of Turkish is a valid one. In the field of Turkology during the 1960s and early 1970s, there was a debate on the extent to which the Turkish language observed in transcription texts (particularly Georgievits 1544 published by Heffening 1942 and Illésházy 1668 published by Németh 1970) could accurately represent a specific Turkish variety (see e.g. Hazai 1990: 64–67, Stein 2016: 161–162). Németh (1968, 1970) argued that these texts were representative of the Balkan dialects of Turkish. On the other hand, Kissling (1968) claimed that the texts contained linguistic mixtures unique to themselves and reflected an imperfectly learned Turkish influenced by the Balkan region. Ultimately, the debate was resolved in favor of the former position, and as a result, a body of scholarly work emerged that utilizes transcription texts as reliable sources for conducting historical linguistic studies of Turkish (see e.g. Csató et al. 2016).

Table 2: Head-dependent orders in NCV and EBT

(a) New contact varieties

|      | ⇒    |      |    |     |      |       |      |       |      |      |     |     | HEAD-INITIAL |     |
|------|------|------|----|-----|------|-------|------|-------|------|------|-----|-----|--------------|-----|
|      |      |      |    |     |      |       |      |       |      |      |     |     | HD           |     |
|      | NGen | NAdj | PN | NRC | VAcc | VSubj | VBar | VNFin | VObl | VAdv | VPP | CCI | VFin         | HD  |
| TUbi | 0%   | 0%   | 0% | 0%  | 4%   | 5%    | 4%   | 5%    | 8%   | 12%  | 21% | 17% | 45%          | 8%  |
| USbi | 0%   | 0%   | 0% | 0%  | 3%   | 2%    | 7%   | 9%    | 18%  | 11%  | 19% | 23% | 68%          | 9%  |
| DEbi | 0%   | 0%   | 0% | 3%  | 4%   | 10%   | 8%   | 12%   | 13%  | 18%  | 19% | 30% | 42%          | 12% |
| HD   | 0%   | 0%   | 0% | 1%  | 4%   | 6%    | 6%   | 8%    | 13%  | 14%  | 20% | 21% | 48%          |     |

(b) Early Balkan Turkic

| EBT | PN | NAdj | VSubj | NGen | VAdv | VBar | VPP | VAcc | VObl | VNFin | NRC | VFin | CCI | HD  |
|-----|----|------|-------|------|------|------|-----|------|------|-------|-----|------|-----|-----|
|     | 4% | 5%   | 5%    | 6%   | 12%  | 14%  | 22% | 32%  | 38%  | 46%   | 50% | 70%  | 86% | 30% |

in NCV. Relative clauses, however, present a major difference between BT and NCV. BT tends to use SAE-type finite postpositive relative clauses, the high incidence of noun-relative clause (NRC) order in Table 1 being a consequence of that. As these have not yet clearly emerged in NCV, almost all relative clauses in use are non-finite prepositive Turkish relatives. Next, again as with BT, finite clausal dependents are progressive, with the difference between finite and non-finite clauses being replicated here. Finally, non-clausal dependents of verbs can again be said to be distributed over the middle range.

The pattern just described obtains not only across four contact situations but across different historical periods as well. Consider the distribution of head-dependent orders in EBT given in Table 2b. As before, noun phrases with non-clausal dependents and adpositional phrases are maximally conservative. Note that relative clauses in EBT behave more like those in modern BT, as SAE-type relative clauses had already emerged in BT by that period. (The relatively less progressive position they occupy in the EBT spectrum may be symptomatic of the earlier stage of this development.) Also, clausal dependents are the most progressive elements and non-clausal dependents of the verb occupy the middle of the spectrum. Note, finally, that the asymmetry between finite and non-finite clauses can also be observed in EBT.

Turning now to the dialect spectrum: There are some differences between the three new contact varieties, but they do not differ much from one another with respect to their overall preferences for head-initial order, and they are all more conservative than the most conservative modern BT variety. Finally, in terms of its overall preference for head-initial order, EBT looks like a mid-range modern BT variety, such as the Ali Koč and Janevë dialects.

To sum up, two spectrums of directionality can be discerned in the new contact varieties as well, with much less marked differences along both dimensions when compared to BT (likely due to differences in the duration of contact).

After these general observations, I now move on to a more refined comparison, based on a cluster analysis of these findings.

### 3.2.3 Progressive versus conservative domains

Figure 1 shows three tree plots detailing how the thirteen head-dependent pairs in the three groups of contact varieties (i.e. NCV, EBT, and BT) may be analyzed into six cluster groups in terms of how progressive or conservative they are *within their respective dialect groups* (hierarchical agglomerative cluster analysis, Manhattan distance, average linkage).

To make it easier to compare them, I will use Table 3, which assigns a numerical value to each of these six clusters, cluster one being the most conservative and cluster six the most progressive. In addition, the table gives the means and mean absolute deviations (MAD) of these cluster values per head-dependent pair.

Table 3: Head-dependent clusters

|              | <b>BT</b> | <b>EBT</b> | <b>NCV</b> | <b>Mean</b> | <b>MAD</b> |
|--------------|-----------|------------|------------|-------------|------------|
| <b>CCI</b>   | 6         | 6          | 5          | 5.67        | 0.44       |
| <b>VFin</b>  | 5         | 5          | 6          | 5.33        | 0.44       |
| <b>NRc</b>   | 6         | 4          | 1          | 3.67        | 1.78       |
| <b>VObl</b>  | 5         | 2          | 3          | 3.33        | 1.11       |
| <b>VNFin</b> | 4         | 3          | 2          | 3.00        | 0.67       |
| <b>VAcc</b>  | 5         | 2          | 2          | 3.00        | 1.33       |
| <b>VPP</b>   | 2         | 1          | 4          | 2.33        | 1.11       |
| <b>VAdv</b>  | 1         | 1          | 3          | 1.67        | 0.89       |
| <b>VBar</b>  | 2         | 1          | 2          | 1.67        | 0.44       |
| <b>VSubj</b> | 1         | 1          | 2          | 1.33        | 0.44       |
| <b>NGen</b>  | 1         | 1          | 1          | 1.00        | 0.00       |
| <b>PN</b>    | 1         | 1          | 1          | 1.00        | 0.00       |
| <b>NAdj</b>  | 1         | 1          | 1          | 1.00        | 0.00       |

Echoing the previous observations, noun phrases with non-clausal dependents and adpositional phrases are consistently the most conservative group (i.e. are strongly head-final). Verb-subject, verb-bare object, and verb-adverb pairs also tend to be conservatively head-final, but less consistently so than the previous

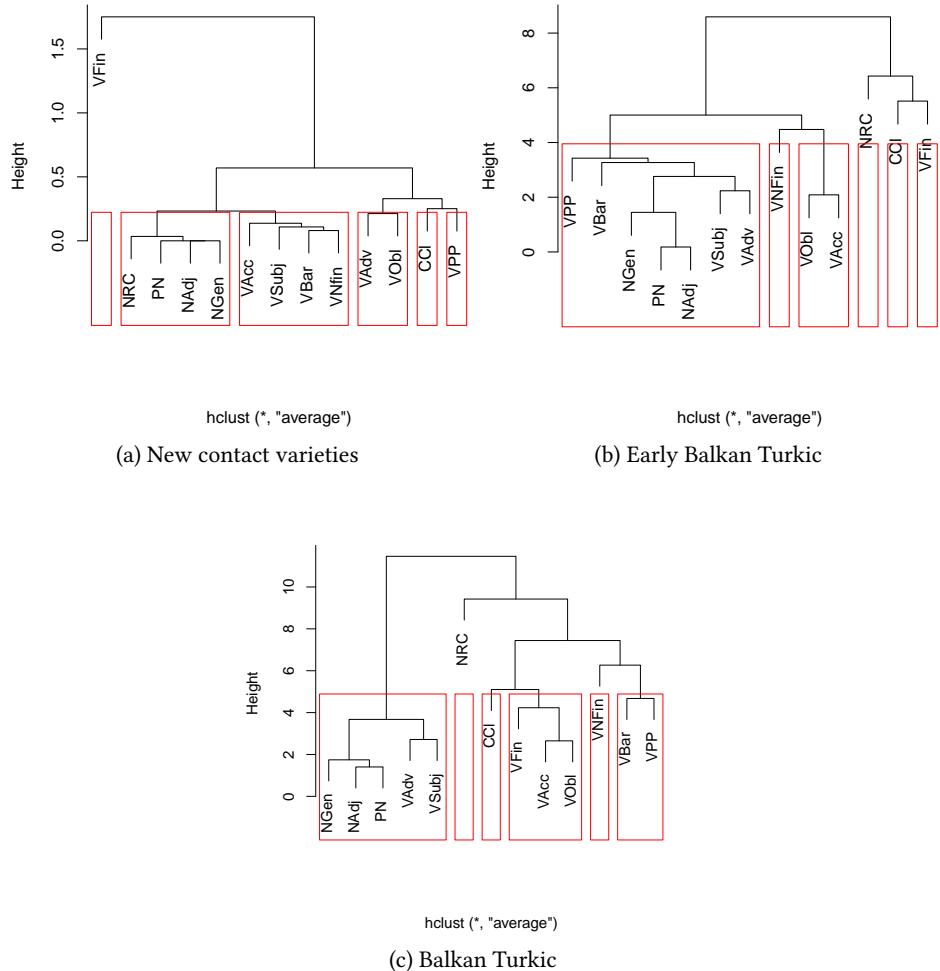


Figure 1: Cluster analysis of Turkic contact varieties

group, as shown by the higher MAD values. Skipping, then, to the top of the table, we see that finite dependent clauses and the clause-initial connectors associated with them are fairly consistently the most progressive group (see the MAD values). The difference between finite and non-finite clauses is clear here as well. In this portion of the table, the lack of SAE-type finite postpositive relative clauses in NCV, mentioned earlier, creates a noteworthy but explainable inconsistency. Finally, verbs with non-clausal dependents are distributed over the middle range (cf. mean  $\approx 3$ ), with the most variation observed in this zone.

### 3.2.4 Change and variability

The discussion that presently follows attempts to answer two questions raised by the data in Table 1:

1. How much do the syntactic domains of a given variety diverge from one another as that variety shifts to head-initial order?
2. How much does a given domain diverge across the varieties in a group (e.g. modern BT) as that domain shifts to head-initial order?

In other words, the following two sections dive into the variation we see along (i) the horizontal dimension (variability within varieties), and (ii) the vertical dimension (variability across varieties) of Table 1. The main generalization that captures the patterns in the data is that there is a strong positive correlation between change and variability along both dimensions.

#### 3.2.4.1 Variability within varieties

The answer to the first question can be visually summarized as in Figure 2.

The scatter plot in Figure 2a shows the link between the incidence of head-initial order in each BT variety as a percentage value (horizontal axis) and MAD across the thirteen syntactic pairs in that variety (vertical axis) (cf. Table 1). The best fitting regression model (i.e. logarithmic regression, adjusted  $R^2 = 0.251$ ), has also been plotted on the graph.

To spell out the generalization suggested by these data: The more a variety shifts to head-initial order, the more the syntactic domains of that variety diverge from one another in head directionality. (I consider some potential internal and external causal factors in Section 3.3.) This finding is supported by a Pearson's correlation test, which showed that there is a strong positive correlation between the incidence of head-initial order and MAD ( $r = 0.517$ ,  $p = 0.008$ , 95% CI: [0.153, 0.757]).

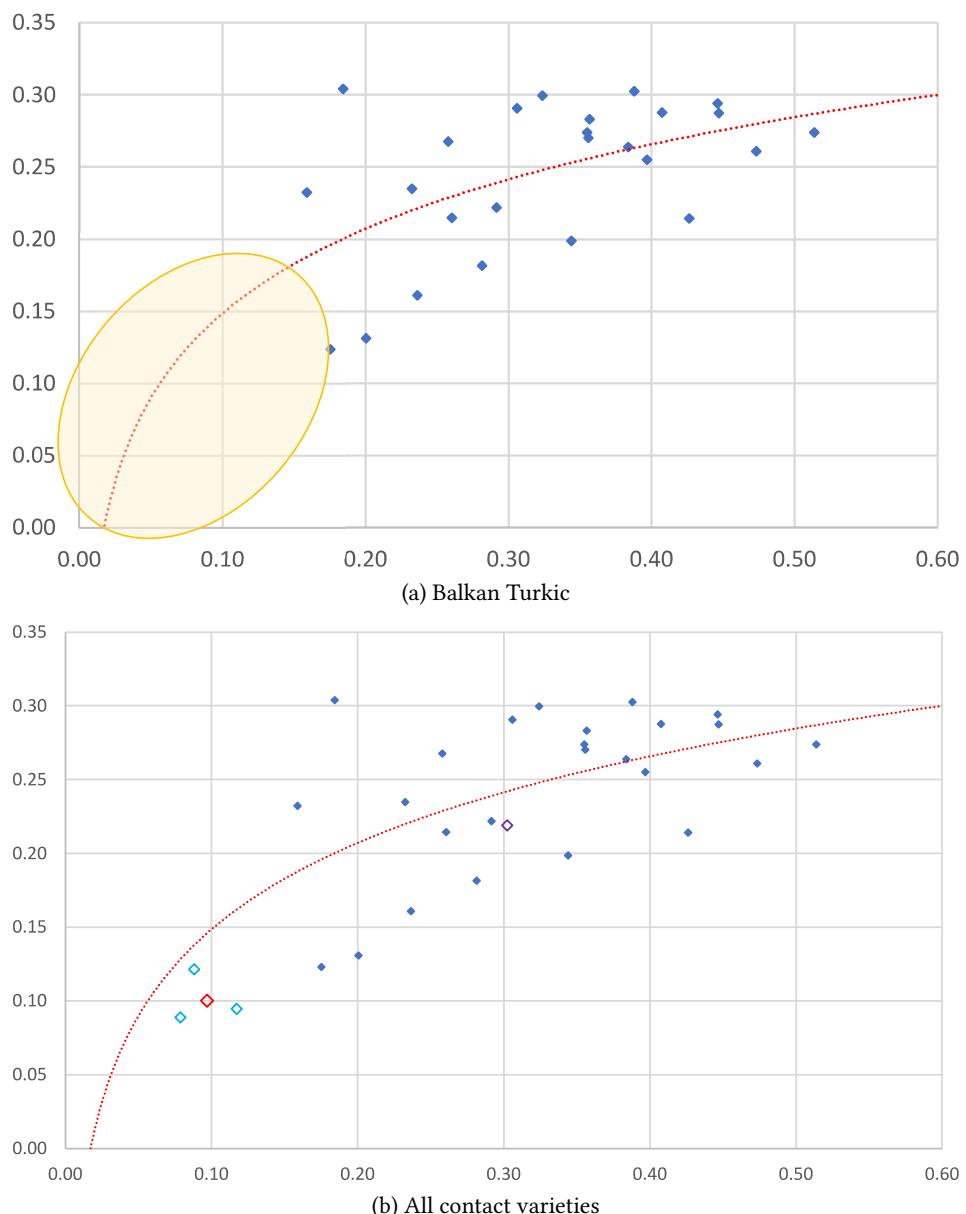


Figure 2: Head-initial order and mean absolute deviation per variety

Based on this generalization, the logarithmic regression model predicts that varieties that are less advanced than modern BT should occur somewhere in the yellow zone in the plot, i.e. that they should be less divergent internally. Indeed, as shown in Figure 2b the three new contact varieties (cyan-colored markers) and monolingual Turkish (red-colored marker) are located close to the trend line. In other words, they show only a limited amount of head-initial word order and are internally less heterogeneous in this regard. Note that the pattern in EBT (purple-colored marker at (0.3, 0.22)), which can be compared to a moderately advanced present-day BT variety, is also consistent with this model.

### 3.2.4.2 Variability across varieties

Let us now turn to the second question above: How much does a given head-dependent pair (e.g. noun–adjective) diverge across varieties as that pair shifts to head-initial order? The answer to this question is given graphically in Figure 3. Similar to Figure 2, this scatter plot indicates the incidence of head-initial order on the horizontal axis and MAD on the vertical. The relationship represented, however, is between the amount of change observed in a given syntactic pair and the MAD of the amount of change in that syntactic pair across all contact varieties.

The plot makes it clear that the more a head-dependent pair shifts to head-initial order, the more it will diverge across varieties in head directionality. As before, Pearson’s correlation shows that this is a very strong positive correlation ( $r = 0.829$ ,  $p < 0.001$ , 95% CI: [0.511, 0.947]).

## 3.3 An explanatory account of change, variation, and variability

Let me now address two questions that arise in connection with the observations in the preceding section:

1. What do these findings mean for the future of NCV?
2. What are the driving factors behind the patterns in Figures 1–3 and Table 3?

I will start with the first question: These findings predict that the more DEbi, USbi, and TUbi adopt patterns of head-initial word order in contact with German, English and Kurmanji, respectively, the more they will diverge to some extent (i) from one another (cf. Figure 3) as syntactic domains undergoing change diverge across varieties, and (ii) internally (cf. Figure 2).

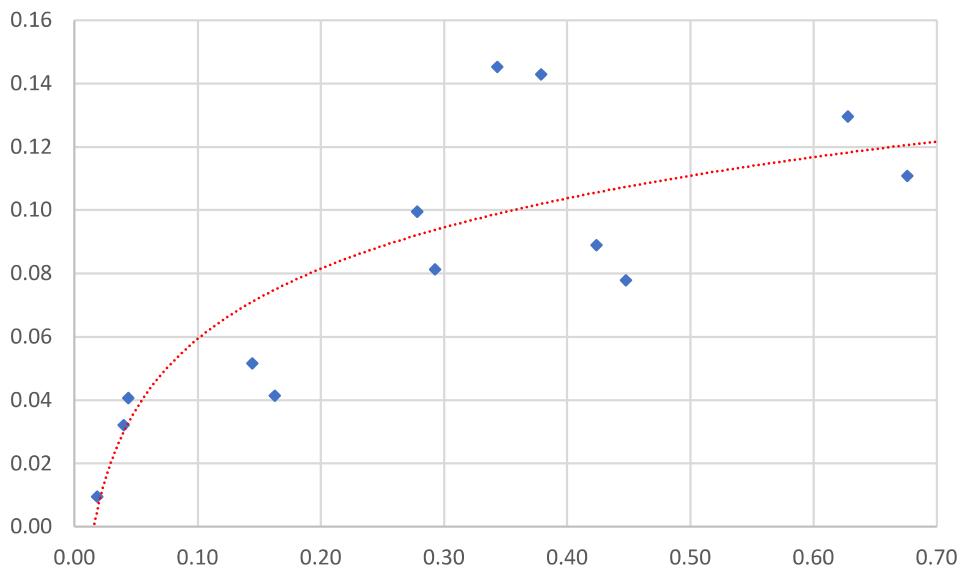


Figure 3: Head-initial order and mean absolute deviation per domain

In an attempt to answer the second question, I will offer some initial thoughts on four explanatory factors common to divergence both across and within varieties, touched upon in the preceding discussion.

### 3.3.1 Degree of exposure to contact

This factor has two facets. The first is time depth or the diachronic angle. BT varieties have been in a multitude of contact situations across the Balkans for a far longer period of time than NCV, and even the most conservative BT variety is more advanced in the shift to head-initial order than the most progressive NCV, as shown in Figure 2b. As the duration of contact between NCV and the majority languages increases, this can be expected to produce stronger effects. As an illustration, consider again the six clusters in Table 3. In that discussion, it was noted that the uppermost and the lowermost ranges of head-dependent pairs are homogeneous across dialect groups, while the middle range (i.e. from NRC down to VPP) contains the highest amount of variation. Despite this heterogeneity, the middle range in BT could be expected to be more advanced than that of NCV due to duration of contact. The means of cluster values in the middle range seem to corroborate this: 4.4 for BT versus 2.4 for NCV. Also, it would not be surprising if NCV and EBT were similar to each other in this regard (as both have had less exposure to contact than modern BT), which they indeed are: Mean of cluster

values in the middle range of both varieties = 2.4. It should be noted, however, that the less consistently conservative VAdv, VBar, and VSubj are surprising from this perspective (and VPP can perhaps also be included among these), as they are more advanced in NCV despite the shorter contact.

The second facet of the degree of exposure involves a synchronic perspective. Keskin (2023a) shows that the incidence of head-initial order in BT strongly and positively correlates with distance from the Turkish border (Pearson's correlation:  $r = 0.438, p = 0.029, 95\% \text{ CI: } [0.052, 0.71]$ ). Keskin proposes that this is likely due to the decreasing size of Turkish-speaking communities in the Balkans as one travels away from Turkey. This idea is supported by a strong negative correlation between the percentage of Turkish speakers in a municipality and the frequency of head-initial order in that municipality (Pearson's correlation:  $r = -0.53, p = 0.04, 95\% \text{ CI: } [-0.82, -0.03]$ ). In other words, speakers tend to use more head-initial order as speech communities shrink.<sup>12</sup> This is probably due to increased exposure to the majority language, as it becomes increasingly unlikely for speakers not to be exposed to the majority language as the group that speaks Turkish becomes smaller. This seems to predict that USbi should shift to head-initial order more than DEbi, as it is spoken by a loosely connected community made up of small, scattered micro-communities, while DEbi is spoken by a tightly connected community made up of larger micro-communities (see e.g. Iefremenko, Schroeder, et al. 2021, Özsoy et al. 2022).

A possible counterbalancing factor from both a diachronic and a synchronic perspective is contact with standard Turkish. As pointed out in Section 3.1, there is a strong connection between NCV and standard Turkish thanks to the internet, media, and regular travels, while such contact with standard Turkish would have been unlikely for most of the period that BT varieties were in contact with SAE. An anonymous review suggests that contact with standard Turkish (however weak it may have been) may also have had an influence on the patterns of change in BT, with communities further west possibly having less contact with standard Turkish.

### 3.3.2 Differential responses of different domains

When the syntactic system undergoes change, different elements or components within the system respond in different ways. This is clearly seen in the differences between head-dependent pairs within one variety.

---

<sup>12</sup>The concept of speech community size should be understood in this context as the size of what we could call a *micro-community*, i.e. for instance, the number of speakers in regular contact with one another in a village and not as the whole number of speakers in a given country.

Regardless of the duration of contact, noun and adposition phrases have remained by and large unchanged across the board, and will likely continue to do so, while finite clausal dependents are inclined towards head-initial order and will possibly become increasingly more head-initial. In this connection, the intriguing divergence between non-finite and finite clauses is another case of differential response replicated across periods/groups of varieties.

Further, obliques and accusative objects seem to be slower to react (and change) than finite clauses, as can be judged from their lower cluster values in EBT and NCV in contrast to their higher values in BT. In other words, it may be that finite clauses are likely the first constituents to change (and rather quickly at that), followed by the categories of the middle range.

Finally, two dependent types with surprising behavior are PP dependents of verbs and adverbs: They react differently to change and fall into different clusters, and are far more postverbal in NCV than in (E)BT.

These cross-categorial differences, particularly the differences between the most progressive and the most conservative categories, will lead to an increase in heterogeneity within the system as their responses diverge further, pulling the syntactic system apart, so to speak.

### 3.3.3 Emergence of new elements

As previously mentioned, an important difference between (E)BT and NCV is the use of SAE-type relative clauses in the former. Prolonged contact may bring about the emergence of these relative clauses in NCV, and they will have different characteristics or behaviors compared to Turkic-type RCs with which they will coexist for a period of time, thereby increasing heterogeneity within one variety.

### 3.3.4 Universal constraints

Despite all the factors that contribute to heterogeneity, there is a limit to this tendency. Due to the logarithmic trends that we have seen, heterogeneity will increase less sharply after a certain point, perhaps eventually leveling off. This can perhaps be attributed to universal constraints of grammar (in a theory-neutral sense of the term). What this implies for NCV is that as they evolve, their syntactic domains are likely to converge towards these universal constraints, after some period of divergence. Over time, the available syntactic options may become more constrained, leading to a slower increase in heterogeneity. For instance, we saw in Section 3.2.3 that subordinators (CCl) and finite clauses (VFin) have similar behaviors in that both favor head-initial orders. This can potentially

be due to a “head-parameter” which pushes these two elements towards convergence (i.e. head-initial order).

## 4 Contact-induced changes in clause combining

As pointed out in Section 3.1, a second set of contact-induced syntactic changes that appear to be taking place in heritage Turkish in the West, connected to a possible shift from head-final to head-initial syntax, involves innovative clause combining strategies, which is what I presently turn to. I will be focusing on a phenomenon which is revealing itself as we take into account what may at first glance appear as noise, namely a class of hybrid, blended, or mixed clauses that seem to emerge as part of contact-induced innovations in the subordination system.

I begin the discussion by describing the subordination models used in Turkic languages (Section 4.1) and the changes seen in clause combining in Turkish in the West (Section 4.2). This will set the stage for the treatment of the phenomenon of interest, which begins with a presentation of hybrid subordinate clauses in Balkan Turkic (Section 4.3). Subsequently, I show that similar hybrid subordinate clauses are also being generated in Western contact varieties (Section 4.4).<sup>13</sup>

### 4.1 Subordination models in Turkic

Turkic languages in contact with languages of Indo-European (IE) stock incorporate IE-type subordination strategies into their repertoires and often shift from the Turkic subordination template to the IE pattern (see e.g. Johanson 2021: 55.2.6, 55.3.8, 903–904, 913–916, 923–924). Balkan Turkic and Karaim with SAE contact, and Khalaj, Uzbek, and Azeri with Persian contact can be cited as examples.

The Turkic and the IE subordination models that these Turkic languages make use of side by side are diametrically opposed to each other. A typical subordinate clause (SC) conforming to the Turkic pattern has four main features (see e.g. Csató & Johanson 1998: 223–224, 229–233, Johanson 1998: 48, 57–66, 2021: 854–931): (i) Its predicate is a non-finite form; (ii) It is positioned before the head noun or the matrix verb, i.e. it is prepositive; (iii) It is subordinated by means of a subordinative element suffixed to its predicate; (iv) if it does involve a free subordinative element, that element is clause-final. Two illustrative examples are given in (2):

---

<sup>13</sup>Note that the material that follows is a summary of various parts of Iefremenko et al. (2022), Keskin (2023b), and Keskin et al. (2024).

- (2) a. [Gel-diğ-in-i] bil-iyor-um.  
           come-NIND-2SG.POSS-ACC know-PROG-1SG  
           'I know that you have come.'  
       b. sen gel-dik-ten sonra  
           2SG come-NIND-ABL after  
           'after you have come'

The bracketed argument clause in (2a) contains a non-finite predicate, signaled by the nominal indicative suffix which also acts as a subordinative element. The clause is positioned before the verb *biliyorum* 'I know' and is marked in accusative case, functioning as the direct object of the verb *bil* 'know'. The adverbial clause in (2b) is again a nominal indicative, but this time including the postposition *sonra* 'after' that functions as an additional subordinative element.

The IE template, by contrast, makes SCs with an opposing set of characteristics possible (see e.g. Göksel & Kerslake 2005: 409–411, 457–460, 463–465, Johanson 1998: 65–66, 2021: 867–868, 894–899, Kerslake 2007, Kornfilt 1997: 3, 46, 60, 321–323, 439–440, 443): (i) Their predicates are finite verb forms; (ii) They are positioned after the head noun or the matrix verb, i.e. they are postpositive; (iii) They are linked to the superordinate clause by means of free subordinative elements; (iv) These subordinative elements are clause-initial. This strategy is exemplified in (3).

- (3) Bil-iyor-um [ki gel-di-n].  
       know-PROG-1SG CONN come-PST-2SG  
       'I know that you have come.'

The argument clause here can be contrasted with the one in (2a). It has a finite predicate marked in past tense and is positioned after the matrix verb *biliyorum*. It is introduced by the free clause-initial connector *ki* of Persian origin.

#### 4.2 Clause combining in Turkish in the West

Turkish in Western Europe and the US may have started to undergo changes in subordination that are comparable to the changes seen in other Turkic varieties with IE contact, as suggested by the observations of studies on heritage Turkish in the West (see e.g. Bayram 2013, Iefremenko, Özsoy, et al. 2021, Iefremenko, Schroeder, et al. 2021, Karakoç 2007, Onar Valk 2015, Özsoy et al. 2022, Schroeder 2016, Schroeder & Iefremenko 2019a, Treffers-Daller et al. 2006, Turan et al. 2020, see also Schroeder et al. 2025 [this volume]). It has begun to use

non-finite clauses less frequently when compared to Turkey's Turkish. It prefers finite clauses instead, with the consequence that it now makes use of paratactic clause combining in place of subordination. The latter strategy was exemplified in (2a), and an illustrative example of the former is given in (4):

- (4) Gel-di-n        bil-iyor-um.  
          come-PST-2SG know-PROG-1SG  
          'I know you have come.'

Example (4), in contrast to (2a), involves two syntactically independent finite clauses that have simply been strung together sequentially without any differences in their morphological form or syntactic function.

IE-type finite subordination can probably not be said to have developed yet in the Western contact varieties of Turkish. However, concomitantly with the move from non-finite to finite clauses, one observes changes in the functions and positions of paratactic connectors, the emergence of new connectors, and a higher frequency of clause combining connectors overall. These developments suggest the beginnings of finite subordination, and they might culminate in Turkic subordination becoming marginal and IE-type subordination dominating in the Turkish varieties in the West, as with BT.

What is critical at this juncture is an observation on BT subordination: The shift to IE-type dependent clauses in BT seems to have happened through the creation of several kinds of hybrid dependent clauses that persist in present-day BT varieties as marginal types of dependent clauses. These will be detailed in the next two sections, before I return to Turkish in the West.

### 4.3 Hybrid subordinate clauses in Balkan Turkic

To elaborate on the preceding remark, the shift from Turkic to IE-type subordination in the case of BT, as I have argued in Keskin (2023b), involves a process that combines, in an apparently random fashion, the properties of Turkic and IE-type subordination outlined in Section 4.1, creating a range of SCs that show mixed properties (*X-clauses*). X-clauses seem restricted to West and North Rumelian, most probably because it is these two groups that make extensive use of IE-type subordination. I have not been able to identify any hybrid SCs in the conservative ERT varieties.

Consider now the examples in (5–10) that illustrate the six subtypes of X-clauses which occur at above-average frequency within this class of SCs and account for a vast majority of their occurrences:

- (5) [Ani sirala-di-m] to urba-lar-i giy-ē-sin.  
     CONN tell-PST-1SG DIST clothes-PL-ACC wear-AOR-2SG  
     ‘You’ll put on the clothes that I told you about.’ (Murtaza 2016: 81)
- (6) Çuval doqū-du-lā onlā-dan [ani zāre-ler-i quy-mā].  
     sack weave.AOR-PST-3PL 3PL-ABL CONN grain-PL-ACC put-NSUB.DAT  
     ‘They weaved sacks from them to put the grains in.’ (Haliloglu 2017: 214)
- (7) Kimsä de-yär-mış [ani ki o nicä gün duuması].  
     some say-PROG-EVID.3SG CONN CONN 3SG like sun rise  
     ‘Some were saying that she is like the sunrise.’ (Çimpoeş 1988: 5 via  
     Özkan 2007: 157)
- (8) Sevin-ēr-im [ani mizin ol-du deye].  
     rejoice-AOR-1SG CONN muezzin become-PST.3SG CONN  
     ‘I am happy that he became a muezzin.’ (Murtaza 2016: 112)
- (9) O sene [bu cade ne yap-il-di] çalış-i-dr-k biz  
     DIST year PROX road CONN make-PASS-PST.3SG work-PROG-PST-1PL 1PL  
     or-da.  
     there  
     ‘The year that this road was built, we were working there.’ (Sulçevsi 2019:  
     255)
- (10) Sū-müş o [ani bizim yaşa-dī-mız] yer-i.  
     plough-EVID.3SG 3SG CONN 1PL.GEN live-NSREL-1PL.POSS place-ACC  
     ‘He ploughed the place where we lived.’ (Karaşinik 2011: 230)

The SC in (5) is almost like an IE-type SC with its finite predicate and clause-initial connector (viz. *ani*), but it is prepositive like a Turkic SC. The SC in (6) is partly like an IE-type SC, since it is postpositive and introduced by an initial subordinator, however it is non-finite like a Turkic SC, as indicated by the nominalized subjunctive marker. The SCs in (7), (8), and (9) are finite and postpositive, again partly conforming to the IE template, but the subordinators that introduce them are atypical. The first has two clause-initial subordinators (viz. *ani ki*); the second has two subordinators, one clause-initial and the other clause-final (viz. *ani...deye*), which I refer to as a *circumclausal subordinator*; and the last has a clause-internal subordinator (viz. *ne*). Finally, the SC in example (10) is non-finite and prepositive following the Turkic template, but it is introduced by a clause-initial subordinator like an IE-type SC.

#### 4.4 Hybrid subordinate clauses in Turkish in the West

So to repeat, the shift from Turkic to IE-type subordination in BT seems to involve a process that combines the properties of Turkic and IE-type subordination, creating a range of X-clauses. If this is true, X-clauses could also be identifiable – at the very least in part – in other Turkic varieties with SAE contact, most relevantly in DEbi and USbi, given the present context.<sup>14</sup> That is the possibility that we investigate in Keskin et al. (2024), and I now move on to the discussion of the examples of X-clauses identified in these Turkish contact varieties as part of that study. Using the description of the subordination models in Turkic in Section 4.1 as a point of reference, I present and discuss the data in terms of (i) clause position and (ii) subordinator type and position.

##### 4.4.1 Clause position

Innovations in clause combining in Turkish as spoken in Germany and US are perhaps most clearly observed in clause position. These are seen significantly more in bilingual data than in monolingual data and involve canonically preverbal SCs occurring in the postverbal position. An example is given in (11):<sup>15</sup>

- (11) İnan-mi-ycak-sın [demin n=ol-düğ-un-a].  
believe-NEG-FUT-2SG just what=happen-NIND-3SG-DAT  
'You won't believe what just happened.' [DEbi33MT\_isT]

Apart from its position relative to the matrix verb, the SC in this example is entirely ordinary. Now, Turkic SCs can and do occur postverbally under the right conditions, given the flexible constituent order of Turkic languages. However, this has a typical frequency signature: Only about a quarter of SCs are postverbal in standard spoken Turkish (22% in our data). In bilingual data, by contrast, we observe a significantly higher frequency of postpositive SCs (40%, with DEbi = 42%

<sup>14</sup>As Keskin (2023b) observes, components of the X-clause phenomenon are actually attested in several unrelated contact situations, such as Indo-Aryan languages with Dravidian contact (Bayer 2001, Dhongde & Wali 2009: 214, Hock 2021: 541–542, Pandharipande 1997: 2, 6, 70) and Laz with Turkish contact (Demirok & Öztürk 2022). Preliminary research suggests that the phenomenon is also seen in Romeyka with Turkish contact (see e.g. Keskin et al. 2023, Schreiber 2019, 2022).

<sup>15</sup>The codes next to the examples in Section 4.4 provide the following information: Country: DE=Germany, TU=Turkey, US=US; Bi-/monolingual speaker: bi vs. mo; Speaker number and age group: 1–50=adults; from 51 onward=adolescents; Gender: M vs. F (there were no speakers who identified as non-binary); Family language: T=Turkish; Communicative situation: f=formal vs. i=informal; s=spoken vs. w=written; Language of production: T=Turkish.

and USbi = 38%), as shown by a generalized linear mixed model. The model was built with the individual speakers as random effects and bilingualism, clause type, and age group as fixed effects. It showed a significant effect of bilingualism and clause type ( $p = 0.017$  and  $p < 0.001$ , respectively): Bilinguals produce significantly more postpositive SCs than monolinguals, and finite SCs are significantly more postpositive than non-finite SCs.

These observations point towards a blending of the IE feature of postpositiveness with the Turkic features of non-finiteness, etc.

#### 4.4.2 Subordinator type and position

I begin the discussion of subordinators and their features with the issue of the possible emergence of free clause-initial connectors that introduce finite SCs (cf. the IE model of subordination). I then present potential clause-initial connectors that introduce non-finite clauses – X-clauses per excellence. I finally connect these two sets of data to irregularities in subordinator position, which present further examples of X-clauses.

The potentially emerging connectors that occupy the central stage in these findings are *hani*, *yani*, and *iste*, all derived from homophonous discourse markers. An important feature of most of the examples which contain these elements is an ambiguity between subordinate and coordinate readings, which can be taken as a sign of the beginning stages of the development of finite subordination in DEbi and USbi.<sup>16</sup>

##### 4.4.2.1 Subordinator type 1: Clause-initial connectors introducing finite subordinate clauses

Consider, first, the example in (12):

- (12) Araba [hani bun-u gör-dü] dur-du.  
 car CONN this-ACC see-PST.3SG stop-PST.3SG  
 'The car that saw this stopped.' [DEbi06FT\_isT]

---

<sup>16</sup>I should point out an important difference between this section and the preceding ones. Here, I will not be presenting any statistical analyses and I confine myself to a presentation of several examples which seem to involve clause combining connectors. The reason for this is that the uses of the above-mentioned elements in the data are open to interpretation and hence their analysis is marred by a high degree of subjectivity. A more systematic and objective investigation of these data is left to future research.

In this example we have, in square brackets, what appears to be a typical IE-type relative clause introduced by *hani*: finite, postpositive, with a free clause-initial subordinator.

The next example in (13) exemplifies the use of *yani* as what again seems to be a subordinator:

- (13) Anla-dı-n [yani bi çift var-di top-lu].  
understand-PST-2SG CONN a couple exist-PST.3SG ball-ATTR  
'You understood that/well there was a couple with a ball.'  
[DEbi18MT\_isT]

Here, the bracketed clause is a possible complement clause but the interpretation of the example is ambiguous: The bracketed segment can also be interpreted as the member of a coordinate structure together with the preceding segment.

Finally, consider the example in (14), demonstrating the use of *işte* as connector:

- (14) Karşı taraf-ta da bi kadın var [işte alışveriş yap-mış san-ır-im].  
opposite side-LOC FOC a woman exist CONN shopping do-PRF.3SG think-AOR-1SG  
'There is a woman on the opposite side who/and she had been shopping I think.' [DEbi08FT\_isT]

As with the preceding two examples, the bracketed clause here can be interpreted as an SC, an extraposed relative detached from its head noun. An alternative interpretation is as a second conjunct.

#### 4.4.2.2 Subordinator type 2: Clause-initial connectors introducing non-finite subordinate clauses

In addition to the cases in (12–14) which strongly resemble IE-type SCs, another pattern that we observe in DEbi and USbi data is the frequent occurrence of *yani* and *işte* immediately preceding various non-finite SCs, as exemplified in (15–17):

- (15) Top=la [işte gid-er-ken] oynuyo-du.  
ball=with CONN go-AOR-CONV play.PROG-PST.3SG  
'S/he was playing with the ball as he went.' [USbi74FT\_fsT]
- (16) [Yani top-u düş-ür-en] adam kadın-a yardım et-ti.  
CONN ball-ACC fall-CAUS-SREL man woman-DAT help-PST.3SG  
'The man who dropped the ball helped the woman.' [DEbi60FT\_isT]

- (17) Bi adam [yani dokuz yüz on bir-i ara-ma-ya] çalış-tı.  
 a man CONN 911-ACC call-NSUB-DAT try-PST.3SG  
 'A man tried to call 911.' [USbi52FT\_isT]

The bracketed segment in (15) contains a preverbal converb (viz. *giderken*). Immediately to the left of the converb is *iste* which could be interpreted as introducing the converbial clause. Next, in example (16), we have a comparable structure which involves *yani* on its left periphery and a prepositive participial relative clause typical for Turkic languages. Finally, (17) contains an infinitival clause preceded by *yani*, producing the same structure as before.

These observations suggest that the IE feature of free clause-initial subordinators is being blended with the Turkic features of non-finiteness and prepositiveness in DEbi and USbi.

#### 4.4.2.3 Subordinator position

I now move on to non-canonical uses of subordinator position that give rise to further examples of X-clauses. Recall from Section 4.3 that BT developed, in addition to canonical free clause-final subordinators (e.g. *diye*) in its repertoire, free subordinators that occupy non-clause-final positions, using them as part of its X-clause inventory, such as double clause-initial (7), circumclausal (8), clause-initial (5–6 and 10), and clause-internal (9) subordinators.

When we turn to Turkish in Germany and the US in the light of this possibly inexhaustive list, we readily observe the occurrence of circumclausal subordinators. This phenomenon involves both well-established clause-initial and final connectors, such as *çünkü* 'because', *için* 'for', *ki* 'which, that', *diye* 'that', but also the potential emergent connectors presented above, i.e. *hani*, *yani*, and *iste*. This appears to be a largely bilingual phenomenon: We have identified three relevant examples in monolingual data (all with the emergent connectors) as opposed to 30 in bilinguals (with both new and old connectors).

Consider the example in (18) with well-established connectors:

- (18) Gerek-iyor-muş [çünkü bunlar-a şahit ol-duğ-um için].  
 be.necessary-PROG-EVID because these-DAT witness-NIND-1SG for  
 'It was necessary because I witnessed these.' [DEbi05FT\_isT]

The adverbial clause in this example has a non-finite predicate, as attested by the nominalized indicative suffix on it – a hallmark of Turkic SCs. The clause contains not only the free clause-final subordinator *için* (Turkic-type) but also the

clause-initial subordinator *çünkü* (IE-type), forming a circumclausal free subordinator. Finally, the clause is positioned after the main verb like an IE-type clause. This blending of Turkic and IE features make this adverbial clause an X-clause.

Next is an example in (19) with an emergent connector:

- (19) Öndeki araba dur-du [hani köpeğ-e çarp-miyim diye].  
in.front car stop-PST.3SG CONN dog-DAT hit-NEG.OPT.1SG CONN  
'The car in front stopped so as not to hit the dog.' [USbi52FT\_isT]

The SC in this example follows a pattern close to that in example (18): postpositive with a circumclausal free subordinator. It is, however, finite, which means that it is a twin of the X-clause in (8).

## 5 Conclusion

Let me begin the concluding section by remembering the question with which I set out: How good a model is the SAE-Turkic contact in the Balkans for understanding contact-induced syntactic changes in Turkish in the West?

As an answer to this question, I presented patterns of change (i.e. the two spectrums of directionality, resistant versus progressive domains, and variability within and across varieties) replicated across contact situations past and present. In the light of these observations, one may be inclined to answer this question with a positive superlative.

Still, two things need to be underlined. First, what can be observed across periods and contact situations is not a single homogeneous response to language contact but a complex pattern with a broad range into which all the contact varieties investigated fit rather well. Second, this pattern may be brought about by explanatory factors which are hardly specific to any given language or family of languages: degree of exposure to contact, differential responses of different syntactic domains, emergence of new elements, and universal constraints.

The same comments apply, mutatis mutandis, to the contact effects observed in the clause combining system as well. The emergence of X-clauses in both Balkan Turkic and the Western contact varieties is a striking convergence. However, the reader will recall that they are attested in other contact situations too (e.g. Romeyka-Turkish contact).

So, the longer and more reasonable answer to the research question above seems to be that Balkan Turkic constitutes a good model for understanding contact-induced syntactic change in Turkish in the West insofar as it can help reveal general contact dynamics, not as a Turkic contact variety per se. And on that front the investigation of Balkan Turkic can be said to have been fruitful.

## Abbreviations

|      |                |       |                         |
|------|----------------|-------|-------------------------|
| 1    | first person   | LOC   | locative                |
| 2    | second person  | NEG   | negation                |
| 3    | third person   | NIND  | nominalized indicative  |
| ABL  | ablative       | NSREL | nonsubject relative     |
| ACC  | accusative     | NSUB  | nominalized subjunctive |
| AOR  | aorist         | OPT   | optative                |
| ATTR | attributive    | PASS  | passive                 |
| CAUS | causative      | PL    | plural                  |
| CONN | connector      | POSS  | possessive              |
| DAT  | dative         | PRF   | perfective              |
| DIST | distal         | PROG  | progressive             |
| EVID | evidential     | PROX  | proximate               |
| FOC  | focus particle | PST   | past                    |
| FUT  | future         | SG    | singular                |
| GEN  | genitive       | SREL  | subject relative        |

## Acknowledgments

The academic research that constitutes the foundation of this article was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant, as part of the projects *Head directionality change in Turkic in contact situations: A diachronic comparison between heritage Turkish and Balkan Turkic* (grant number: 394841858) and *Clause combining in Balkan Turkic: Pathways and stages of contact-induced grammaticalization* (grant number: 313607803) within the Research Unit *Emerging Grammars in Language Contact Situations: A Comparative Approach* (FOR 2537).

First and foremost, I thank Christoph Schroeder for his unwavering encouragement and support, as well as his comments, suggestions, exhortations, etc. I am grateful to Heike Wiese for allowing me to join the RUEG network with my supplemental projects and to Shanley E. M. Allen to link my research with the project that she co-directed with Heike and Christoph. I also thank the two anonymous external reviewers, the internal reviewers Onur Özsoy and Judith Purkarthofer, and the following colleagues (in alphabetical order): Eystein Dahl, Kateryna Iefremenko, Jaklin Kornfilt, Nils Picksak, Sergey Say, Ilja Seržant, Anna Shadrova, and Nadine Zürn. Finally, I thank Lea Coy for help with pre-publication formatting. The usual disclaimers apply.

## References

- Adamović, Milan. 1976. *Vocabulario Nuovo mit seinem Türkischen Teil*. *Rocznik Orientalistyczny* 38. 45–50.
- Adamović, Milan. 1977. *Das osmanisch-türkische Sprachgut bei R. Lubenau* (Beiträge zur Kenntnis Südosteuropas und des Nahen Orients 25). München: R. Trofenik.
- Adamović, Milan. 2001. *Das Türkische des 16. Jahrhunderts: Nach den Aufzeichnungen des Florentiners Filippo Argenti* (1533). Göttingen: Pontus Verlag.
- Balci, Mustafa. 2010. Kosova ve Makedonya Türk ağızlarında devrik cümle meselesi. *Avrasya Etüdleri* 38(2). 43–60.
- Bayer, Josef. 2001. Two grammars in one: Sentential complements and complementizers in Bengali and other South Asian languages. In Rajendra Singh (ed.), *The yearbook of South Asian languages and linguistics*, 11–36. New Delhi: Sage.
- Bayram, Fatih. 2013. *Acquisition of Turkish by heritage speakers: A processability approach*. Newcastle: Newcastle University. (Doctoral dissertation).
- Blau, Otto. 1868. *Bosnisch-türkische Sprachdenkmäler*. Leipzig: Deutsche Morgenländische Gesellschaft.
- Böhm, Stefanie. 2015. Differential object marking in standard Turkish and Caucasian Urum. *Language Typology and Universals* 68(4). 421–438.
- Bombaci, Alessio. 1940. Padre Pietro Ferraguto e la sua Grammatica turca (1611). *Annali del R. Istituto Superiore Orientale di Napoli*. 205–236.
- Brendemoen, Bernt. 1980. Labiyal ünlü uyumunun gelişmesi üzerine bazı notlar. *Türkiyat Mecmuası* 19. 223–240.
- Brezina, Vaclav. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press. <http://corpora.lancs.ac.uk/stats/index.php>.
- Bulut, Christiane. 2006. Turkish elements in spoken Kurdish. In Hendrik Boeschoten & Lars Johanson (eds.), *Turkic languages in contact*, 95–121. Wiesbaden: Harrassowitz Verlag.
- Bunk, Oliver, Shanley E. M. Allen, Sabine Zerbian, Tatiana Pashkova, Yulia Zuban & Erica Conti. 2025. Information packaging and word order dynamics in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 379–412. Berlin: Language Science Press. DOI: 10.5281/zenodo.1577518.
- Çimpoeş, Lübov S. 1988. *Literatura okumakları 9–10 klasslar*. Chişinău.
- Csató, Éva Á. 2000. Syntactic code-copying in Karaim. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *The circum-baltic languages: Their typology and contacts*, 265–277. Amsterdam: John Benjamins.

- Csató, Éva Á. & Lars Johanson. 1998. Turkish. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, 203–235. London: Routledge.
- Csató, Éva Á., Astrid Menz & Fikret Turan (eds.). 2016. *Spoken Ottoman in mediator texts*. Wiesbaden: Harrassowitz Verlag.
- Demirok, Ömer & Balkız Öztürk. 2022. The emergence of clausal nominalizations in Laz. Talk delivered at Formal Approaches to Contact in/with Turkish, UiT, Tromsø.
- Destanov, Ekrem. 2016. *Doğu Makedonya sözlü geleneğinde masallar üzerine bir inceleme*. Istanbul: Yıldız Technical University. (Dissertation).
- Dhongde, Ramesh V. & Kashi Wali. 2009. *Marathi*. 3rd edn. (London Oriental and African Language Library 13). Amsterdam: John Benjamins.
- Doerfer, Gerhard. 1959. Das Gagausische. In Jean Deny, Kaare Grønbech & Scheel, Helmuth, Togan, Zeki Velidi (eds.), *Philologiae turcicae fundamenta I*, 260–271. Wiesbaden: Steiner.
- Dorleijn, Margreet. 2006. Turkish–Kurdish language contact. In Hendrik Boeschoten & Lars Johanson (eds.), *Turkic languages in contact*, 75–94. Wiesbaden: Harrassowitz Verlag.
- Drimba, Vladimir. 1966. « L'instruction des mots de la langue turquesque » de Guillaume Postel (1575). *Türk Dili Araştırmaları Yıllığı-Belleten*. 77–101.
- Du Ryer, André. 1630. *Rudimenta grammatices linguae turcicae*. Paris: Antoniu Vitray.
- Dündar, Fuat. 2000. *Türkiye nüfus sayımlarında azınlıklar*. 2nd edn. İstanbul: Çiviyaziları.
- Friedman, Victor A. 1982. Balkanology and Turkology: West Rumelian Turkish in Yugoslavia as reflected in prescriptive grammar. In Anthony A. Barentsen, Ronald Sprenger & Marinus G. M. Tielemans (eds.), *South Slavic and Balkan linguistics*, vol. 2 (Studies in Slavic and General Linguistics), 1–77. Leiden: Brill.
- Friedman, Victor A. 2006. West Rumelian Turkish in Macedonia and adjacent areas. In Hendrik Boeschoten & Lars Johanson (eds.), *Turkic languages in contact*, 27–45. Wiesbaden: Harrassowitz Verlag.
- Gallotta, Aldo. 1986. Latin harfleri ile yazılmış bir kaç Osmanlı atasözü. *İstanbul Üniversitesi Edebiyat Fakültesi Türk Dili ve Edebiyatı Dergisi* 24. 235–249.
- Georgievits, Bartholomaeus. 1544. De turcarum ritu et caeremoniis.
- Göksel, Aslı & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London: Routledge.
- Gülsevin, Gürer. 2017. *Yaşayan Türkiye Türkçesi ağızlarının tarihî dönemlerini belirleyebilmek için bir yöntem denemesi örneği: XVII. yüzyıl Batı Rumeli Türkçesi ağızları*. Ankara: Türk Dil Kurumu Yayınları.

- Günşen, Ahmet. 2010. Rumeli ağızlarının söz dizimi üzerine I: Makedonya ve Kosova Türk ağızları örneği. *Turkish Studies* 5(1). 462–494.
- Haig, Geoffrey. 2017. Western Asia: East Anatolia as a transition zone. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 396–423. Cambridge: Cambridge University Press.
- Haig, Geoffrey & Ergin Öpengin. 2018. Kurmanji Kurdish in Turkey: Structure, varieties, and status. In Christiane Bulut (ed.), *Linguistic minorities in Turkey and Turkic-speaking minorities of the peripheries*, 157–230. Wiesbaden: Harrassowitz Verlag.
- Haliloglu, Neşe. 2017. *Zavet köyü Türk ağızı*. Kırşehir: Ahi Evran University. (Master's thesis).
- Haspelmath, Martin. 1998. How young is Standard Average European? *Language Sciences* 20. 271–287.
- Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*, vol. 2, 1492–1510. Berlin: De Gruyter.
- Hazai, György. 1960. Textes turcs du Rhodope. *Acta Orientalia Academiae Scientiarum Hungaricae* 10(2). 185–229.
- Hazai, György. 1990. Die Denkmäler des Osmanisch-Türkeitürkischen in nicht-arabischen Schriften. In György Hazai (ed.), *Handbuch des türkischen Sprachwissenschaft*, 63–73. Budapest: Akadémiai Kiadó.
- Heffening, Willi. 1942. *Die türkischen Transkriptionstexte des Bartholomaeus Georgievits aus den Jahren 1544–1548*. Leipzig: Brockhaus.
- Helmholdt, Wolfgang. 1966. Das türkische Vaterunser in Hans Schiltbergers Reisebuch. *Central Asiatic Journal* 11(2). 141–143.
- Herbinius, Johannes. 1675. *Horae Turcico-Catecheticae: Sive institutio brevis catechetica, cuiusdam Turcae circumcisii*. Danzig: David-Fridericus Rhetius.
- Hock, Hans H. 2021. *Principles of historical linguistics*. 3rd edn. (Trends in Linguistics 34). Berlin: De Gruyter.
- Iefremenko, Kateryna. 2019. *Placement of the locational goal elements in Kurmanji Kurdish in contact with Turkish*. Talk delivered at the 4th International Conference on Kurdish Linguistics, University of Rouen, France.
- Iefremenko, Kateryna. 2021a. *Placement of the locational goal elements in Kurmanji Kurdish in contact with Turkish and German*. Talk delivered at the 9th International Conference on Iranian Linguistics (ICIL 9), Österreichische Akademie der Wissenschaften, Vienna, Austria.

- Iefremenko, Kateryna. 2021b. *Post-predicate position in Kurmanji in contact with Turkish and German*. Talk delivered at the Workshop on Turkic and Languages in Contact with Turkic (TU+), University of Toronto, Canada.
- Iefremenko, Kateryna. 2021c. *Word order in Turkish in contact with German and Kurmanji*. Talk delivered at the RUEG Conference - Dynamics of Language Contact, Humboldt Universität, Berlin, Germany [online].
- Iefremenko, Kateryna, Cem Keskin, Jaklin Kornfilt & Christoph Schroeder. 2022. *Individual variation in heritage Turkish clause combining*. Talk delivered at Formal Approaches to Contact in/with Turkish, Tromsø.
- Iefremenko, Kateryna, Cem Keskin & Christoph Schroeder. 2023. The post-predicate position in Turkish in language contact situations: How resistant to change? Talk delivered at DGfS 2023, Universität zu Köln.
- Iefremenko, Kateryna, Onur Özsoy & Christoph Schroeder. 2021. *Paratactic clause combining in heritage Turkish in Germany and the US*. Talk delivered at the International Conference on Turkish Linguistics, Anadolu University, Eskisehir, Turkey [online].
- Iefremenko, Kateryna, Christoph Schroeder & Jaklin Kornfilt. 2021. Converbs in heritage Turkish: A contrastive approach. *Nordic Journal of Linguistics* 44(2). 130–154. DOI: 10.1017/S0332586521000160.
- İğci, Alpay. 2010. *Viçitirin-Kosova Türk ağzı*. Izmir: Ege University. (Master's thesis).
- Illésházy, Miklós. 1668. Dictionarium turcico-latinum.
- Jable, Ergin. 2010. *Kosova Türk ağızları*. Sakarya: Sakarya University. (Dissertation).
- Johanson, Lars. 1989. Substandard und Sprachwandel im Türkischen. In Günter Holtus & Edgar Radtke (eds.), *Sprachlicher Substandard II* (Konzepte der Sprach- und Literaturwissenschaft), 83–112. Tübingen: Niemeyer.
- Johanson, Lars. 1992. *Strukturelle Faktoren in türkischen Sprachkontakten* (Sitzungsberichte der wissenschaftlichen Gesellschaft an der Johann Wolfgang Goethe-Universität Frankfurt am Main). Stuttgart: Steiner.
- Johanson, Lars. 1998. The structure of Turkic. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, 31–66. London: Routledge.
- Johanson, Lars. 2006. Turkic language contacts in a typology of code interaction. In Hendrik Boeschoten & Lars Johanson (eds.), *Turkic languages in contact*, 4–26. Wiesbaden: Harrassowitz Verlag.
- Johanson, Lars. 2013. *Structural factors in Turkic language contacts*. London: Routledge.
- Johanson, Lars. 2021. *Turkic* (Cambridge language surveys). Cambridge: Cambridge University Press.

- Joseph, Brian D. 2020. Language contact in the Balkans. In Raymond Hickley (ed.), *The handbook of language contact*, 2nd edn., 537–549. Chichester: John Wiley & Sons.
- Kakuk, Suzanne. 1958. Textes turcs de Kazanlyk II: Zehra Džafer. *Acta Orientalia Academiae Scientiarum Hungaricae* 8(3). 241–311.
- Kakuk, Suzanne. 1961a. Die türkischen Mundart von Küstendil und Michailograd. *Acta Linguistica Academiae Scientiarum Hungaricae* 11(3/4). 301–386.
- Kakuk, Suzanne. 1961b. Türkische Volksmärchen aus Küstendil. *Ural-altaische Jahrbücher* 33(1/2). 90–94.
- Kakuk, Suzanne. 1972. Le dialecte turc d’Ohrid en Macédoine. *Acta Orientalia Academiae Scientiarum Hungaricae* 26(2/3). 227–283.
- Karakoç, Birsel. 2007. Connectivity by means of finite elements in monolingual and bilingual Turkish discourse. In Jochen Rehbein, Christiane Hohenstein & Lukas Pietsch (eds.), *Connectivity in grammar and discourse* (Hamburg Studies on Multilingualism 5), 199–227. Amsterdam: John Benjamins.
- Karaşinik, Bahtışen. 2011. *Silistre (Silistra) ili ve yöresi ağızları*. Edirne: Trakya University. (Master’s thesis).
- Katona, Louis K. 1969. Le dialecte turc de la Macédoine de l’ouest. *Türk Dili Araştırmaları Yıllığı-Belleten* 17. 57–194.
- Kerslake, Celia. 2007. Alternative subordination strategies in Turkish. In Jochen Rehbein, Christiane Hohenstein & Lukas Pietsch (eds.), *Connectivity in grammar and discourse* (Hamburg Studies on Multilingualism 5), 231–258. Amsterdam: John Benjamins.
- Keskin, Cem. 2023a. On the directionality of the Balkan Turkic verb phrase: Variationist and theoretical perspectives. *Languages* 8(1). 1–20.
- Keskin, Cem. 2023b. Transient subordinate clauses in Balkan Turkic in its shift to Standard Average European subordination: Dialectal and historical evidence. *Folia Linguistica* (57). 155–197.
- Keskin, Cem, Kateryna Iefremenko, Jaklin Kornfilt & Christoph Schroeder. 2024. Hybrid clause combining strategies in Turkish language contacts. *Studia Linguistica* (Early View). DOI: 10.1111/stul.12256.
- Keskin, Cem, Laurentia Schreiber, Kalliopi Katsika & Metin Bağrıaçık. 2023. *Contact-induced blended subordination in Romeyka*. Talk delivered at the 3rd International Conference on Language Contact in the Balkans and Asia Minor, Aristotle University of Thessaloniki, Thessaloniki, Greece.
- Keskin, Cem, Taner Sezer, Christoph Schroeder, Lale Diklitaş, Zeynep Güler, Nil Özkul, Süleyman Yüceer, Ayşe Nur Güler & Ebru Özgenç. In preparation. *Balkan Turkic corpus*.

- Kirli, Nadejda. 2001. *N. N. Baboglu'nun 'Legandanın izi' adlı eserinin cümle bilgisi yönünden incelemesi*. Adana: Çukurova University. (Master's thesis).
- Kissling, Hans Joachim. 1968. Bemerkungen zu einigen Transkriptionstexten. *Zeitschrift für Balkanologie* 6. 119–127.
- Kornfilt, Jaklin. 1997. *Turkish*. London: Routledge.
- Küçükcan, Talip & Veyis Güngör (eds.). 2009. *Turks in Europe: Culture, identity and integration*. Amsterdam: Turkevi Research Centre.
- Matras, Yaron. 2009. *Language contact*. 1st edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Matras, Yaron & Şirin Tufan. 2007. Grammatical borrowing in Macedonian Turkish. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 215–227. Berlin: Mouton de Gruyter.
- McCarus, Ernest N. 2009. Kurdish. In Gernot Windfuhr (ed.), *The Iranian languages*, 587–633. London: Routledge.
- Menz, Astrid. 1999. *Gagauzische Syntax: Eine Studie zum kontaktinduzierten Sprachwandel* (Turcologica 41). Wiesbaden: Harrassowitz Verlag.
- Menz, Astrid. 2014. Gagauz. *Journal of Endangered Languages* 3(3). 55–69.
- Murtaza, Durgyul Metin. 2016. *Razgrad ili Kubrat ilçesi Alevi-Bektaşı köyleri ağızı (inceleme, metinler, sözlük)*. Edirne: Trakya University. (Master's thesis).
- Németh, Gyula. 1968. Die türkische sprache des Bartholomaeus Georgievits. *Acta Linguistica Academiae Scientiarum Hungaricae* 18(3-4). 263–271.
- Németh, Gyula. 1970. *Die türkische Sprache in Ungarn im siebzehnten Jahrhundert* (Bibliotheca Orientalis Hungarica 13). Budapest: Akadémiai Kiadó.
- Onar Valk, Pelin. 2015. *Transformation in Dutch Turkish subordination? Converging evidence of change regarding finiteness and word order in complex clauses*. Utrecht: LOT Publications.
- Özkan, Nevzat. 1996. *Gagavuz Türkçesi grameri* (Türk Dil Kurumu Yayınları 657). Ankara: Atatürk Kültür, Dil ve Tarih Yüksek Kurumu.
- Özkan, Nevzat. 2007. *Gagavuz destanları* (Türk Dil Kurumu Yayınları). Ankara: Atatürk Kültür, Dil ve Tarih Yüksek Kurumu.
- Özsoy, Onur, Kateryna Iefremenko & Christoph Schroeder. 2022. Shifting and expanding clause combining strategies in heritage Turkish varieties. *Languages* 7(3). 242. DOI: 10.3390/languages7030242.
- Pandharipande, Rajeshwari V. 1997. *Marathi* (Routledge descriptive grammars). London: Routledge.
- Petrou, Maria. 2019. Diglossia features and bilingualism in the Turkish-speaking community of Western Thrace. In C. Tzitzilis & G. Papanastasiou (eds.), *Γλωσσικές επαφές στα βαλκάνια και στη μ. ασία*, 2nd edn., vol. 1, 222–232. Thessaloniki: Ινστιτούτο Νεοελληνικών Σπουδών (Ιδρυμα Μ. Τριανταφυλλίδη).

- Petrou, Maria. 2021. *Der türkische Dialekt von Westthrakien* (Turcologica 125). Wiesbaden: Harrassowitz Verlag.
- R Core Team. 2021. *R: A Language and environment for statistical computing*. Version 4.3.0. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schreiber, Laurentia. 2019. Romeyka. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 892–934. Berlin: De Gruyter Mouton.
- Schreiber, Laurentia. 2022. *A (contact-)grammar of Romeyka*. Bamberg & Gent: Otto-Friedrich-Universität Bamberg & Universiteit Gent. (Dissertation).
- Schroeder, Christoph. 2016. Clause combining in Turkish as a minority language in Germany. In Mine Güven, Didar Akar, Balkız Öztürk & Meltem Kelepir (eds.), *Exploring the Turkish linguistic landscape: Essays in honor of Eser Erguvanlı-Taylan* (Studies in Language Companion Series), 81–102. Amsterdam: John Benjamins. DOI: 10.1075/slcs.175.06sch.
- Schroeder, Christoph. 2018. *The post-verbal position in heritage Turkish in Germany: Focus on register*. Talk delivered at EuroSLA, Colloquium Turkish Heritage Speakers in Germany, Münster.
- Schroeder, Christoph & Kateryna Iefremenko. 2019a. Göçmen Türkçesinde cümle birleştirme: Pilot çalışma. In Kamil İşeri (ed.), *Dilbilimde güncel tartışmalar*, 243–250. Ankara: Dilbilim Derneği Yayımları.
- Schroeder, Christoph & Kateryna Iefremenko. 2019b. *The post-verbal position in Turkish in two contact settings: Turkish–German bilinguals in Germany and Turkish–Kurdish bilinguals in Turkey*. Talk delivered at the Workshop ‘Turkish in Europe’, Universität Hildesheim.
- Schroeder, Christoph & Kateryna Iefremenko. 2019c. *The postverbal position in heritage Turkish in Germany: Focus on register*. Poster presented at the 4th workshop on Turkic and languages in contact with Turkic, New York, USA.
- Schroeder, Christoph, Kateryna Iefremenko, Kalliopi Katsika, Annika Labrenz & Shanley E. M. Allen. 2025. Clause combining in narrative discourse: A contrastive analysis across heritage and majority languages. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 251–286. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775173.
- Schroeder, Christoph, Kateryna Iefremenko & Mehmet Öncü. 2024. The postverbal position in heritage Turkish: A comparative perspective with a focus on non-clausal elements. In Zeynep Kalkavan-Aydin & Yazgül Şimşek (eds.), *Deutsch-Türkische Zweisprachigkeit mit besonderem Fokus auf Ju-*

- gendliche [Turkish-German bilingualism with a special focus on adolescents]* (Mehrsprachigkeit [Multilingualism] 58), 109–132. Münster: Waxmann.
- Skopeteas, Stavros. 2015. From OV to VO: Observables and linguistic structures. Talk delivered at the University of Athens, Philosophical Faculty, 24.04.2015.
- Stein, Heidi. 2016. The dialogue between a Turk and a Christian in the Grammatica Turchesca of Pietro Ferraguto (1611). Syntactical features. In Éva Á. Csató, Astrid Menz & Fikret Turan (eds.), *Spoken Ottoman in mediator texts*, 161–171. Wiesbaden: Harrassowitz Verlag.
- Sulçevsi, İsa. 2019. *Kosova Türkçesi ve fil yapıları*. İstanbul: İstanbul University. (Dissertation).
- Teza, Emilio. 1892. Un dialogo turco fatto in Italia nel cinquecento. *Rendiconti della R. Accademia dei Lincei* 1(6). 391–407.
- Treffers-Daller, Jeanine, A. Sumru Özsoy & Roeland van Hout. 2006. Oral language proficiency of Turkish–German bilinguals in Germany and Turkey: An analysis of complex embeddings in Turkish picture descriptions. In Hendrik E. Boeschoten, Lars Johanson, Sevgi Ağcagül & Vildan Milani (eds.), *Turkic languages in contact*, 203–219. Wiesbaden: Harrassowitz.
- Turan, Dilek, Elena Antonova-Ünlü, Çiğdem Sağın-Şimşek & Mehmet Akkuş. 2020. Looking for contact-induced language change: Converbs in heritage Turkish. *International Journal of Bilingualism* 24(5–6). 1035–1048.
- van der Auwera, Johan. 2011. Standard Average European. In Bernd Kortmann & Johan van der Auwera (eds.), *The languages and linguistics of Europe*, vol. 1, 291–306. Berlin: Mouton de Gruyter.
- Varol, Orhan. 2014. *İkidilli konuşuculara ait verilerdeki dilbilimsel görünümler*. Ankara: Ankara Üniversitesi. (Dissertation).
- Weil, Gotthold. 1953. Ein unbekannter türkischer Transkriptionstext aus dem Jahre 1489. *Oriens* 6. 239–265.
- Whorf, Benjamin Lee. 1944. The relation of habitual thought and behavior to language. *ETC: A Review of General Semantics* 1(4). 197–215.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.



# Chapter 10

## Dynamic properties of the heritage speaker lexicon

Nadine Zürn<sup>a</sup>, Mareike Keller<sup>a</sup>, Rosemarie Tracy<sup>a</sup> & Anke Lüdeling<sup>b</sup>

<sup>a</sup>University of Mannheim <sup>b</sup>Humboldt-Universität zu Berlin

Against the backcloth of current research on heritage speakers' linguistic knowledge and behavior, this chapter focuses on a domain very much in flux in any speaker: the lexicon. We investigate lexical resources and resourcefulness in written and spoken descriptions of the same event by heritage speakers in both their languages, and by monolingual speakers of English and German. Spoken and written reports are based on a filmed staged accident and were elicited in standardized situations manipulated in order to encourage use of either formal or informal registers. Our line of argument moves along the following research questions: First, how comparable are different speaker groups with respect to lexical inventory and lexical diversity, and what trends can be identified? Second: How do heritage speakers of German who were raised in the US and monolingual speakers of German compare with respect to German particle verbs regarding syntactic, morphological, semantic, phonological, and pragmatic properties? Third: What insight into lexical resources can be gained by studying performance-related phenomena (self-interruptions, self-repairs, filler particles, etc.)? Speaker group comparisons of lexical diversity and inventory are conducted via statistical modelling, whereas the particle verb analyses dealing with various interface phenomena are based on fine-grained qualitative analyses. Overall, our findings provide further evidence for tendencies towards explicitness and transparency discussed in heritage language research.



## 1 Introduction

Regardless of how researchers conceptualize the architecture of the mental lexicon, most will probably agree that it is the most dynamic subsystem of our overall linguistic knowledge. Once vocabulary growth speeds up in early childhood, our lexical repertoires continue to expand, but not necessarily in the languages we started out in, and certainly not just linearly. Words of our childhood are adjusted to target forms with respect to phonology, morphologically restructured, and recategorized with respect to contextual and cultural appropriateness.

Whether and to what extent the lexicon of a first language (L1) develops across a speaker's lifetime depends on many factors, including contact with and properties of other languages acquired from birth or later on, with each potentially influencing the others. Essentially then, the mental lexicon is a moving target, principally "on the go", with new words discoverable at any time and either holistically adopted or analyzed according to available productive word formation and inflection processes. The investigation of heritage languages (henceforth HLs), whose speakers often feel that their majority language (henceforth MajL) is the more proficient and dominant one, provides us with a natural laboratory for exploring how intricate word-related knowledge can be acquired in diverse acquisition scenarios.

What happens to immigrants' HLs in the long run, i.e. across generations in diasporic islands, has already been investigated for many language combinations. As for German as a heritage language, the adoption and adaptation of new vocabulary, especially freely importable discourse markers, leveling of irregular morphosyntactic paradigms, changes in argument structure, and word order have been identified as prominent outcomes of contact with English (Matras 1998, Muysken 2000, Fuller 2001, Clyne 2003, Boas 2009, Boas 2010, Muysken 2013, Putnam & Salmons 2013, Riehl 2014, Hopp & Putnam 2015, Stolberg 2015, Zimmer et al. 2020, etc.). Some of these long-term changes as well as convergence of similar forms, ample borrowing, and orthographic interaction have also been documented in first generation immigrants (Tracy & Lattey 2001, Clyne 2003, Schmid 2011, Keller 2014, Tracy 2022).

Complementary to these two research strands – long-term effects of language contact on immigrant languages in diaspora communities on the one hand and L1 change in first-generation immigrants on the other – this contribution focuses on second-generation immigrants. Our participants are early bilinguals, exposed to the HL within their family context. In some cases, contact with the HL is limited to communication with just one parent, as in the Tiny Language Island scenario discussed in Tsehayé et al. (2025 [this volume]). Where the minority

language can neither draw on a HL community outside the home nor on the educational system (as in mother-tongue classes, bilingual programs or foreign-language classes), its speakers may only rarely be exposed to functional varieties of their HL other than an informal-spoken register. As the background variables impacting quantity and quality of exposure can be very heterogeneous, the heterogeneity of linguistic outcomes, which is often mentioned in the literature, does not come as a surprise (Montrul 2006, Moreno-Fernandez 2007, Fairclough 2010, Polinsky 2018). However, it is important to recognize that similar inter-individual differences are noticeable in monolingual speakers (henceforth MSs) (Shadrova et al. 2021, Wiese et al. 2022, and other chapters in this volume).

Standardized data collection, corpus compilation and analysis took place in the context of the Research Unit *Emerging Grammars in Language Contact Situations* (RUEG), described in detail in Wiese et al. (2025 [this volume]), Klotz et al. (2024) and briefly in Section 3 below. The target were heritage varieties of Greek, Russian, and Turkish both in Germany and in the United States, as well as heritage German in the United States. We elicited the same type of data from monolingually raised majority language speakers in all countries, thereby minimizing the risk of attributing non-canonical HS utterances to language contact with one of the two MajLs, English or German.

In this contribution we report results based on a quantitative exploration of the HSs' lexical resources by means of lexical diversity calculations and lexical inventory assessments and identify group-specific patterns related to the lexicon. Further, two additional kinds of dynamics are addressed in qualitative analyses, both focusing on German particle verbs (PVs): First, we present canonical and non-canonical occurrences of PVs in order to capture patterns pointing towards innovation and change. Second, we ask what production phenomena, such as hesitations, filler items and overt repairs surrounding PVs reveal about word candidates considered at the moment of speaking and what they tell us about speakers' implicit judgement of the quality of their own utterances.

Despite the overall rise of interest in heritage languages and in what they contribute to general theories of learnability and language change, the question of how HSs use (non-)lexicalized forms in their productions still calls for an answer. We aim at contributing to closing this gap by analyzing the lexical items speakers resort to in situations where they are confronted with specific spoken and written tasks in both their languages. Our findings support the claims from previous literature that HSs, especially in contrast with second language learners, are “comfortable in experimenting with the lexicon of their language” (Polinsky 2018: 294–295) and that they display preferences for compositionality, semantic transparency, and explicitness (see Rakhilina et al. 2016, Pashkova et al. 2020).

Our argumentation proceeds as follows. Section 2 starts from a conception of the mental lexicon as tightly interconnected with all levels of grammar. This section also provides our rationale for selecting German particle verbs for scrutiny later on. Section 3 introduces the corpus and methodology. In Section 4, the empirical portion of this chapter begins with a quantitative analysis of the lexical inventory and lexical diversity. Section 5 narrows the focus to German PVs and illustrates subtle differences between HSs' and MSs' productions. Section 6 continues the exploration of PVs but shifts attention to production phenomena as additional ways in which speakers provide us with evidence for the lexical resources under their control. Section 7 summarizes findings, points out limitations and raises new questions.

## 2 The lexicon as a dynamic and interconnected resource

Current theories of the mental lexicon no longer consider it in isolation from the rest of the grammar or as a mere storage space for the non-productive and necessarily listed items, including multi-word idiomatic expressions. At the same time, many approaches go for “the lexicon all the way down”, with all kinds of meaningful units, from morpheme to clauses and even larger discourse chunks considered more or less unique form-function pairings in a gigantic construction (Goldberg 2005, Tomasello 2006, Bybee 2010; various contributions in Engelberg et al. 2011). However, the absence of consensus on how to best capture item-vs. rule-generated properties of natural languages, or, more specifically, of the lexicon, is irrelevant for our concern at this moment.

Following up on a metaphor by Jackendoff & Audring (2019), we conceive words to be “small bridges” across phonology, morphology, semantics, syntax, pragmatics, and, in writing, orthography. This means that word knowledge is inherently relational. As far as individual lexical items are tied to specific registers, different dialects or languages, they have to be marked accordingly. The bridging function of words across linguistic interfaces within each language and across languages, as well as the co-activation potential and competition of formally and/or semantically similar candidates, make lexical items highly susceptible to fluctuation. At the same time, these multiple connections provide speakers with a rich source for creativity and innovation (Degani et al. 2011, Prior et al. 2017, Rabinovich et al. 2018), which is particularly relevant from our perspective.

Researchers inquiring into the HS lexicon concluded that it does not match the repertoire and behavior of monolingual peers in size or age-adequate use (e.g. Montrul 2006, Polinsky 2018). Differences have also been identified with respect to lexical retrieval (Moreno-Fernandez 2007, Polinsky 2018). As stated re-

peatedly in the literature, reduced exposure beyond early childhood and decreasing relevance of the HL offer plausible explanations for differences between HSs and MSs. After all, in contrast with minority languages, the majority L1 does not have a status problem, hence is not questioned or threatened but supported by the education system.

One way of assessing lexical resources is measuring lexical diversity (LD). LD is considered “an important indicator of language learners’ active vocabulary and of how it is deployed” (Malvern & Richards 2002: 85) to communicate effectively and appropriately. Numerous studies operationalize LD as an indicator of proficiency and “a type of linguistic complexity” (Jarvis 2013: 95), either as a stand-alone measure or in combination with others, for instance lexical density and sophistication of expression (Bonvin et al. 2018, Gharibi & Boers 2019, Elabdali et al. 2022) in both spoken and/or written productions (Laufer & Nation 1995, Malvern & Richards 2002, Pennock-Speck & Clavel-Arroitia 2021). To this day, various kinds of LD measures have been applied to monolingual, to L2, as well as to HS data either for grouping speakers into proficiency categories (e.g. Kopotev et al. 2020) or with the aim to validate the appropriateness of this measure for comparing speaker groups, also with respect to different settings (Daller et al. 2003, Yu 2009, Hržica & Roch 2021, Petersen et al. 2021). For instance, in a study on LD in reports of younger and older HSs in comparison with monolingual peers, Gharibi & Boers (2019) found lower LD values in younger HSs compared to monolinguals and to older HSs. The authors attribute higher LD values in the latter to longer exposure time. On the whole, measures aiming at the assessment of a speaker’s lexicon only provide snapshots of a temporary state of lexical knowledge since that state is likely to change as a consequence of continued language exposure and use (Yu 2009, Czapka et al. 2021, Lambelet 2021), which is, in turn, connected to issues of the wider context and differences in status as minority or majority language (Treffers-Daller & Korybski 2016, Treffers-Daller 2019).

Variability in language dominance is also reflected in setting- or register-specific vocabulary. Van Gijsel et al. (2005), for instance, show an effect of register variation on lexical richness measured by type-token ratio, with lower values for informal settings compared to formal ones. This is in line with Alamillo (2019)’s findings on Spanish heritage and L2 speakers. Furthermore, topic familiarity, time pressure during production and self-confidence have been considered predictors of LD. Here, the intra-speaker comparison of spoken and written productions by Yu (2009) makes a better prediction of LD in spoken compared to written language productions, with overall similar levels of LD between both modes. Written tasks, which usually provide more time, yielded higher LD values, especially when subjects were familiar with the topic and felt more confident (Yu 2009: 250). Against this backdrop, Section 4 pursues the overarching research

question of what we can deduce from LD and LI measures in group comparisons, given the diverse acquisition contexts attested for heritage speakers.

As we later move from the quantitative assessments of lexical repertoires and the pros and cons of LD and LI measures to a very specific but theoretically complex type of verb, German particle verbs, some justification for this move is called for. Particle verbs are Janus-faced: On the one hand they behave like complex words, on the other hand like phrasal syntagmas. What are commonly considered PVs, such as *ankommen* (at-come, ‘arrive’) and *anrufen* (up-call, ‘phone’), do not form one homogeneous class. Their structural analysis is as controversial as it is intriguing. While some authors consider them words with strange properties, many analyze them as syntactic constructions – see e.g. Müller (2002), Lüdeling (2001), and Felfe (2012). As shown for example by Lüdeling (2001), it is even unclear what to count as a particle since various form classes behave similarly. As a theoretical discussion of PVs is far beyond the scope of this paper, our presentation of corpus data in Sections 5 and 6 is limited to undisputed cases, namely verbal particles homonymous with adverbs or prepositions, and we treat PVs as lexical entries without any further comments on their structural status. In the context of HL acquisition and maintainance, PVs are an interesting research object with respect to (a) syntactic distribution within clauses, (b) semantic function, and (c) differences in form due to their co-occurrence and amalgamation with deictic elements.

Example (1) gives a first impression of the material available in the RUEG corpus. The short passage contains seven clauses, main clauses (MC) and subordinate clauses (sc) counted separately. They are all functionally assertions, as expected in reports. The altogether ten verbs occur in one of two possible positions for verbs in German: finite verbs appear in second position in declarative main clauses, clause-finally in subordinate clauses.

- (1) [MC Ein Auto **is** in einem anderen Auto hinten **reingefahren** [sc weil das erste Auto für ein Hund schnell bremsen **musste**]]. [MC Das erste Auto **war** blau] [MC und das Auto [sc das hinten **reingefahren ist**] **war** weiss oder vielleicht grau]. [MC Niemand **sah verletzt aus**] [MC und jemand hat die Polizei **angerufen**].

‘One car rearended another car because the first car suddenly had to brake for a dog. The first car was blue, and the car that hit it from behind was white or maybe grey. Nobody seemed hurt and someone called the police.’ (USbi50FD\_fwD)<sup>1</sup>

---

<sup>1</sup>See Section 4, Table 2 for detailed information on how to interpret the speaker codes provided at the end of each example.

Non-finite verbs (infinitives, participles) and – crucial here – particles of PVs only appear in final position. In main clauses, finite base verbs and auxiliaries occur in second position. Hence, whenever a PV appears clause finally, whether non-finite in main clauses, or (non-)finite in subordinate clauses, the particle and the verb are adjacent and orthographically rendered without space, i.e., they look (and “feel”) like a word.

In addition to their distributional properties (continuous vs. discontinuous), PVs are semantically relevant in two ways. First, many of them are semantically intransparent and have to be learned as a whole. Others form patterns which can then be used for productive new formations. Second, we will see in later discussions of the way events are described (Section 5) that verbs encode manner of motion, with the particle providing path information with respect to direction and goal, thereby also often contributing to a change in aspect (Talmy 1988, Tenny & Pustejovsky 2000, Slobin 2003). *Fahren* in Example (2) is atelic (an activity in the sense of Vendler 1957). The particle verb *reinfahren* ‘crash into, hit’ in Example (3) is telic. The particle *rein* implies a goal (in this case *ihm* ‘him’, meaning the other driver and his car). The goal can also often be expressed by a full PP, as exemplified by *auf den Parkplatz* ‘into the parking lot’ in Example (4).<sup>2</sup>

- (2) an dem besagten tag **fuhr-en**        zwei autos auf dem        parkgelände  
on the said        day drive-PST.3PL two cars on the.DAT parking.area  
hintereinander  
behind.each.other  
‘on said day two cars drove behind each other in the parking area’  
(DEbi02FT\_fsD)
- (3) und er **fähr-t**        ihm        **rein**  
and he drive-PRS.3SG him.DAT into.VPART  
‘and he hits him’ (DEbi05FT\_iwD)

---

<sup>2</sup>Transcription conventions: Oral productions are transcribed according to project-internal transcription and annotation guidelines (<https://korpling.german.hu-berlin.de/rueg-docs/latest/annotations.html>). Mark-up irrelevant for the present discussion, (e.g. vowel length), has been removed. Pauses are marked by a hyphen in brackets. The spelling in written productions has been gently normalized to enhance readability and keep the focus on aspects relevant to this study.

Glossing: The glosses follow the basic principles of the Leipzig Glossing Rules (<https://www.eva.mpg.de/lingua/resources/glossing-rules.php>), but morphological boundaries and categories are only marked as far as they are relevant for our exposition. For verbal particles as the morphological feature in focus we introduced the label VPART. All items relevant for the discussion of production phenomena (hesitations, repetitions, self-corrections, etc.) are marked by a double asterisk (\*\*) in the gloss.

- (4) genau in dieser sekunde fuhr-en zwei autos auf den  
exactly in this second drive-PST.3PL two cars on the.ACC  
parkplatz  
parking.lot  
'exactly this second two cars entered the parking lot' (DEbi03FR\_isD)

While PVs expressing motion events are often fairly transparent, complexity arises because speakers can choose between morphologically different forms of the particle, such as *ein/rein/herein* (all meaning ‘in(-to)'). Sometimes choice is motivated by register parameters, in other cases choices have consequences for the expression (or comprehension) of argument structure (Härtl & Witt 1998). We saw in Example (3) that the PV *reinfahren* can be used with a dative argument. In (5), the same particle verb is used without an argument and with a slight shift in meaning: an overt locative argument (like a parking lot or a garage) of the particle is omitted but inferrable in shared non-verbal contexts. In Example (6), the particle *hinein* and the full PP *in diesen* ‘into this one’ are both used.

- (5) auf einmal fuhr ein weißes auto rein.  
at once drive.PST.3SG a white car into.VPART  
'Suddenly a white car drove in.' (DEbi52FT\_fsD)
- (6) Dieser fuhr nach der Voll-bremsung des Mazda in diesen  
this drive.PST.3SG after the full-brake of.the Mazda in this-ACC  
hin-ein.  
there-in.VPART  
'That one ran into the Mazda after its emergency stop.' (DEbi34FR\_fwD)

Even though PV behavior is even more complicated, for our purpose here it suffices to say that particles can influence the argument structure of the base verb. Sometimes the particle satisfies an argument slot, sometimes it opens up an argument position, and sometimes it changes the argument function (see e.g. Stiebels & Wunderlich 1994, Lüdeling 2001, Zeller 2001, Müller 2002, Boas 2003, Felfe 2012).

Although our current focus is on adolescent and adult heritage speakers, findings on L1 acquisition are worth pointing out. Interestingly, the syntactic behavior of PVs in both continuous and discontinuous transparent constellations is no acquisition hurdle (Schulz & Tracy 2011, Tracy 2011, 1991). Separable telic particles are already part of children’s lexicon at the time when they only produce one-word utterances (*weg* ‘away’, *auf* ‘up, open’, *zu* ‘closed’, *rein* ‘into’), and they are present in early two-word combinations with or without their verbal

base, with or without deictic expressions, hence well before the appearance of finite V2-clauses with finite verbs and particles separated.

The early emergence and stability of PVs can be attributed to the confluence of the following: (a) their consistent position at the end of clauses; (b) their contextually relevant and transparent semantic content compared to the rest of the verb, which is often a semantic lightweight, compare *aufmachen* ‘to open’ (lit: ‘to make open’), *zumachen* ‘to close’ (lit: ‘to make closed’), and *wegmachen* ‘to remove’ (lit: ‘to make gone’); (c) when combined with their verbal base, particles bear word stress (*AUFmachen zumachen*); and finally, (d), as particles, they remain uninflected, hence consistent in form, apart from combining with deictic elements which, in turn, contribute important information with respect to event specifics.

Likewise, comprehension studies provide evidence for children’s early sensitivity to telicity, see Van Hout (2000) for Dutch, where particles are equally precocious, and for L1 and early L2 German (Schulz & Tracy 2011).<sup>3</sup> In conclusion, the early emergence and prominence of PVs in a child’s lexicon could explain which of the overall intricate features associated with particle verbs are mastered and remain resilient in HSs even though L1 exposure may decrease after early childhood.

### 3 Data and methodology

The corpus explored here is the outcome of a large comparative research initiative (Lüdeling et al. 2024, RUEG) investigating various HLs (Greek, Turkish, Russian) in Germany in comparison with the same HLs plus Heritage German in the United States. The core data gathered across all projects consists of reports elicited on the basis of a video stimulus showing a staged minor car accident (Wiese 2020, see also Wiese et al. 2025 [this volume]). Participants are monolingual and bilingual adolescents (age 13–19) and adults (age 20–37). As one of the joint research goals is the investigation of formality- and mode-specific linguistic repertoires, all participants were asked to relate what happened in the 40-second film clip to different imagined addressees, both in speech and in writing. Fictive addressees were the police, contacted via voicemail (formal-spoken) and written testimony (formal-written), as well as friends, again in a spoken voice message

---

<sup>3</sup>Similarly, the formal properties of PVs do not seem to be difficult for older learners of German as a Foreign Language, see Lüdeling et al. (2017).

(informal-spoken) and in a written WhatsApp message (informal-written).<sup>4</sup> Bilingual speakers performed all tasks in both their heritage and their majority language. During elicitation sessions and during casual encounters around these sessions, project members – each in charge of eliciting either the heritage or the majority language data – did not engage in code-switching. Therefore code-switches and borrowing on part of participants was not primed by interlocutors. The repeated elicitation of reports on the same event in four different settings makes it possible to investigate consistency of lexical and grammatical choices, frequency of occurrence and cooccurrence of lexical items, type and frequency of morphological processes, as well as the adherence to and extension of the semantic scope of lexical items. In addition it provides us with insights into sensitivity towards different varieties, i.e. the specific registers which are of common concern to all projects (see the other chapters in this volume and Tsehayé et al. 2021, Wiese et al. 2022, Pashkova et al. 2022).<sup>5</sup>

In order to allow for corpus searches targeting the morphological make-up of lexemes and the inclusion of performance phenomena, we decided to augment the existing corpus by additional annotations:

1. Manual annotation of all verb tokens in the complete German sub-corpus for lemma (associates the separated particles of particle verbs with their base verb), morphological type (simplex, prefix, particle), and syntactic function (lexical, modal, auxiliary, copular)<sup>6</sup>
2. Selective manual span annotation of production phenomena (hesitations, filler particles, interruptions, repetitions, repairs, etc.)

Although the RUEG data was not elicited specifically for investigating lexical inventories, the use of the same stimulus material across various conversational settings as well as across participants with varying language backgrounds makes it possible to analyze and compare lexical resources within and across speaker groups. Given the generally dynamic nature of lexical knowledge and the variability in exposure of HSs to registers, as well as differences in opportunity to

---

<sup>4</sup>While we realize that differentiating degrees of formality is a complex matter, the descriptors “formal” and “informal” here refer to carefully arranged, formal or informal elicitation contexts, with even elicitors dressed accordingly.

<sup>5</sup>The notion *register* – roughly: situationally and functionally conditioned variation – is complex and we cannot do it justice here (Biber & Conrad 2009, Egbert & Biber 2018, Matthiessen 2019, Lüdeling et al. 2022). In the RUEG context, we operationalize register by the four situations created to elicit the data. For more detail see Tsehayé et al. (2025 [this volume]).

<sup>6</sup>The annotation guidelines are available online at <https://korpling.german.hu-berlin.de/rueg-docs/standalone/verb-analysis/>.

use their HL, we expect differences between monolingually raised, majority and heritage speakers of a given HL, here German. More specifically, we expect a gradation effect in LD and LI size from monolingual speakers of a language, such as German, to majority speakers of German followed by heritage speakers of German, as well as differences between the situational and conversational settings in line with Van Gijsel et al. (2005), Yu (2009), or Alamillo (2019). For majority language use, such as English in the US, we do not predict a similar gradation pattern. Even though the analyses discussed in later sections specifically focus on HSs of German raised in the United States (USbiGer), Section 4 includes data from other speaker groups for selective comparison.

Overall, our comparisons include monolingually raised speakers of English (USmo) or German (DEmo) as well as HSs of Greek, Russian or Turkish dominant in English (USbiGreek, USbiRuss, USbiTurk) or dominant in German (DEbiGreek, DEbiRuss, DEbiTurk). Table 1 displays the number of speakers per speaker group, along with summary statistics on the token count calculated across the elicitations of all speakers per speaker group. A comparison of the mean token count across groups and languages indicates that the USbiGer group has the lowest mean token count of all German groups, while, as expected, their mean token counts are similar to the other US groups, which consist of further majority English as well as English monolingual speakers. The average token count ranges from 111.51 to 168.39 tokens per elicitation session, with considerable variation within all groups ( $SD = 55.45\text{--}89.16$ ), irrespective of acquisition type and language (see also Shadrova 2025).

## 4 Lexical diversity and inventory

Our analysis begins with lexical diversity (LD), a common measure to determine language dominance in bilinguals (Treffers-Daller & Korybski 2016) as well as to examine language proficiency (Malvern & Richards 2002, Jarvis 2013). Whereas LD measures have been applied to different types of language data, the present contribution uses retellings of events as a basis for assessing LD, which have been deemed useful for measuring lexical knowledge, also cross-linguistically (Simon-Cereijido & Gutiérrez-Clellen 2009: 321). To assess LD both in English and in German, we employed the Moving Average Type-Token Ratio (MATTR, Covington & McFall 2010), which is considered a suitable LD measure for short texts of the type available in the RUEG data (Zenker & Kyle 2021),<sup>7</sup> on lemmatized

---

<sup>7</sup>Analyses were also performed with the Measure of Textual Lexical Diversity (MTLD, McCarthy & Jarvis 2010), which do not show substantially different results.

Table 1: Summary statistics of speaker group & token count

| Language | Speaker Group | Group Size | Mean   | Median | SD    | Min | Max |
|----------|---------------|------------|--------|--------|-------|-----|-----|
| German   | DEmo          | 64         | 159.93 | 142.50 | 80.84 | 34  | 524 |
|          | USbiGer       | 36         | 111.51 | 96.00  | 56.90 | 34  | 312 |
|          | DEbiGreek     | 45         | 134.02 | 121.50 | 57.04 | 42  | 420 |
|          | DEbiRuss      | 61         | 168.39 | 146.50 | 83.85 | 43  | 682 |
|          | DEbiTurk      | 65         | 150.21 | 137.50 | 74.54 | 36  | 595 |
| English  | USmo          | 64         | 124.34 | 111.00 | 55.45 | 36  | 305 |
|          | USbiGer       | 34         | 131.80 | 120.00 | 59.33 | 46  | 318 |
|          | USbiGreek     | 65         | 128.48 | 117.00 | 60.85 | 40  | 412 |
|          | USbiRuss      | 65         | 149.26 | 133.00 | 89.16 | 38  | 880 |
|          | USbiTurk      | 59         | 143.46 | 130.00 | 66.73 | 37  | 446 |

tokens.<sup>8</sup> Preliminary descriptive analyses, the research design (i.e., the nature of the data), and theories about differences in LD between HSs and MSs (e.g., Bonvin et al. 2018, Gharibi & Boers 2019) lead to the formulation of three linear mixed-effects models with MATTR as the dependent variable and contrast-coded independent variables (Schad et al. 2019), which are explained in Table 2. The final model structures are given in Table 3.<sup>9</sup>

Models 1 and 2 evaluate the MATTR measure across all of the German (Appendix, Table 4) and the English data (Appendix, Table 5), respectively.<sup>10</sup> Additionally, Model 3 was set up to target both German and English only for the

<sup>8</sup>The lemmatized tokens include all content and function words, as well as repetitions or repairs but exclude hesitations and non-verbal material.

<sup>9</sup>All quantitative analyses discussed in the present section are available in the Open Science Framework project “Quantitative Analyses of the Lexical Diversity and Lexical Inventory of Heritage Speakers in the RUEG Corpus”. The analyses were implemented with R (R Core Team 2021) using the following packages: tidyverse (Wickham et al. 2019), lme4 (Bates et al. 2015), emmeans (Lenth 2025), sjPlot (Lüdecke 2024), MASS (Venables & Ripley 2002), hypr (Rabe et al. 2020), performance (Lüdecke et al. 2021), kableExtra (Zhu 2024), and ggpubr (Kassambara 2023).

<sup>10</sup>At this point, it is important to note that the IV “speaker type” is not based (solely) on a theoretical demarcation between different types of HSs. For analytical reasons, we distinguish the USbiGer group from the other DEbi and USbi speaker groups as we focus on this former subgroup, even though from an aquisitional perspective, the USbi and USbiGer speaker groups do not differ from each other apart from the respective HL. Hence, the “US” and the “USbi” designations exclude the USbiGer speaker group in subsequent analyses.

Table 2: Independent variables

|                  |   |
|------------------|---|
| Speaker type     | German HSs with English as the MajL (USbiGer)<br>majority language speaker (DE-/USbi),<br>monolingual speaker (DE-/USmo),   |
| Formality        | formal, informal  |
| Mode             | spoken, written   |
| Session          | first, second, third, fourth elicitation session  |
| Language (=lang) | German, English   |
| Language order   | MajL-HL, HL-MajL  |
| ID               | unique speaker identifier (e.g. USbi72FD) composed of...<br>...Elicitation country DE (Germany), GR(eece),<br>RU(ssia), TU(rkey), US(A)<br>...Acquisition type mo(nolinguist), bi(lingual)<br>...Age group 01–49 (adolescent), 50–99 (adult)<br>...Gender F(emale), M(ale), X (diverse)<br>...HL/L1 D (German), E(nglish), G(reek),<br>R(ussian), T(urkish) |

Table 3: Final model structures

---

```

1: mattr ~ speakertype * formality * mode + session + (1 | ID),
   data=German
2: mattr ~ speakertype * formality * mode + session + (1 | ID),
   data=English
3: mattr ~ lang + formality + mode + session + language_order + (1 | ID),
   data=USbiGer

```

---

USbiGer speakers, the group in focus, and thus includes the IVs “lang” instead of “speakertype” (Appendix, Table 6). Since there are multiple LD measures per speaker, the variable “ID” is included as a random factor in all models. The three models which are reported result from the statistical evaluation of assumption tests and pairwise comparisons between model structures with and without the interactions of interest. The conditional r-squared ( $R^2_c$ ) values, visible in the model summary tables on the LD measurements, reveal that the models explain between ~48% and ~56% of the variance.

The summary of Model 1 (Appendix, Table 4) on the LD in the German data shows significant simple effects for speaker type, formality, and mode. Figure 4

illustrates a clear gradation pattern between the three speaker types: DEmos display a higher LD compared to DEbis, while these two groups show a higher LD compared to the USbiGer group. Regarding formality and mode, the model indicates lower LD in the formal (opposite to Van Gijsel et al. 2005, Alamillo 2019) and the spoken (in line with Van Gijsel et al. 2005; contrasting Yu 2009), respectively. Furthermore, there is an interaction effect between formality and mode, with lower MATTR values consistently in the spoken condition compared to the written one, however, less of a difference between spoken and written in the formal as opposed to the informal condition. This effect is mainly driven by the written LD values which are considerably lower in the formal condition in comparison to the informal condition. The independent variable "elicitation session" does not show a significant simple effect on the MATTR. There is considerable intra-level variation for the DEmo and USbiGer speaker groups compared to the DEbi speaker group.

Similar to the findings of Model 1, Model 2 on the LD in the English data (summary in Table 5, Appendix) shows simple and interaction effects for formality and mode. The speaker group comparisons reveal a significant difference between the USbiGer and the US speaker groups, i.e. all other groups in this data set, while the difference between the USmo and USbi groups (here, excluding USbiGers) lacks statistical significance. However, this result might be misleading, as plot C in Figure 5 (Appendix) suggests similarities in LD between the USbiGer and the USmo speaker groups but not between the USmo and USbi speaker groups. This contradiction can be explained by the large intra-group variation, estimated by 95% confidence intervals. From this, we conclude that, despite the contradictory significance values, the USbiGer and the USmo groups have similar LD, whereas the USbi group shows a lower LD, but only by a slight margin.

Taken together, Models 1 and 2 show a clear gradation pattern between the three speaker groups in the German data, with the USbiGers showing the lowest LD values, whereas the English data indicates less gradation and more similarity between speaker types. The results of both models also highlight large intra-group variances for USbiGers, USmos, DEmos, calling into question the meaningfulness of LD as a group measure for both monolinguals and bilinguals in the varying situational and conversational settings.

Model 3, which only calculates values for the USbiGer group (summary in Table 6, Appendix), reveals significant simple effects of language, formality and mode but no significant simple effect for language order and elicitation session (Appendix, Figure 6). This confirms the significant effect of formality and mode present within the USbiGer group, with lower LD in the formal and the spoken condition, respectively. It also reveals that, as expected, the USbiGers as a group

demonstrate a higher LD in English, their majority language, than in German, their HL.

Summing up, the MATTR calculations as a proxy for LD show that the number of different lexemes used in the reports making up the RUEG data vary in relation to speaker type, formality and mode. With respect to HSs of German (ie. USbiGer), analyses indicate that their vocabulary used is less diverse than that of the other speaker groups in their HL but equally diverse in their dominant language. Furthermore, the HSs of German show similar patterns in LD in the varying situational and conversational settings, comparing their majority and heritage language as well as the other English and German speaker groups.

As a next step, we looked at the lexical items used by the different speaker groups to get a clearer idea of the lexical inventory (LI). To this end, we conducted distinct- and shared-lemma analyses of three word classes (adjectives, nouns, and verbs) on a descriptive level. First, the size of the LI for German and for English is compared by counting the types<sup>11</sup>, normalized by the number of speakers per group. Second, we focus on lemmas shared across speaker groups. These LI analyses are based on the same speaker groups as the LD analyses, and they target the same variables that appeared to be relevant with respect to LD, namely speaker type, language, formality and mode. Generally, the number of different lemmas a speaker uses has the same effect on both the LD and the LI measures used in this study. In other words, the two measures are positively correlated. Yet, they provide different insights into the speaker's lexicon. LD is influenced by the lexeme repetition rate (Jarvis 2013: 87), i.e. it decreases if lexemes are repeated within the specified text span or window. Hence, the LD measure allows us to quantify *how* speakers make use of the lexical inventory at their disposal. The LI value, in turn, is not affected by lexeme repetition and only gives insight into *how many* word types are actively used by a speaker (group) in the four reports and thus indicates that the speakers hold "at least this number of words" (Nation & Anthony 2016: 358) in their repertoire.

The comparison of LI size between speaker groups in the German sub-corpus (Figure 1, plot A) reveals the smallest LI for the USbiGer and DEbiTurk speaker groups, while the largest LI is observed for the DEmo and DEbiRuss speaker groups. The DEbiGreek speaker group lies in-between. Hence, the clear gradation pattern between the DEmo, DEbi and USbiGer speaker groups we saw for the LD analyses is repeated. Furthermore, a small LI in combination with a low

---

<sup>11</sup>By *types* we mean unique lemmas, by *tokens* all occurrences, or word forms, of a lemma or class of lemmas in the corpus (see Pustejovsky & Batiukova 2019: 10). For example, the current version of the German sub-corpus contains 31.933 occurrences of verbs, i.e. verb tokens. These are instantiations of 1331 different verbs on the lemma layer, i.e. verb types.

LD, as seen for the USbiGer group in the German data, indicates that the smaller set of lemmas the speakers have at their disposal is also used more repetitively compared to the other speaker groups. Looking at the three major word classes of nouns, verbs and adjectives separately, the LI calculations show differences between the speaker groups concerning adjectives and nouns. In contrast, the size of the verb inventory is rather consistent across the speaker groups. In this data, the number of distinct adjectives is generally the lowest, while the verb inventory is approximately of the same size or smaller than the noun inventory.

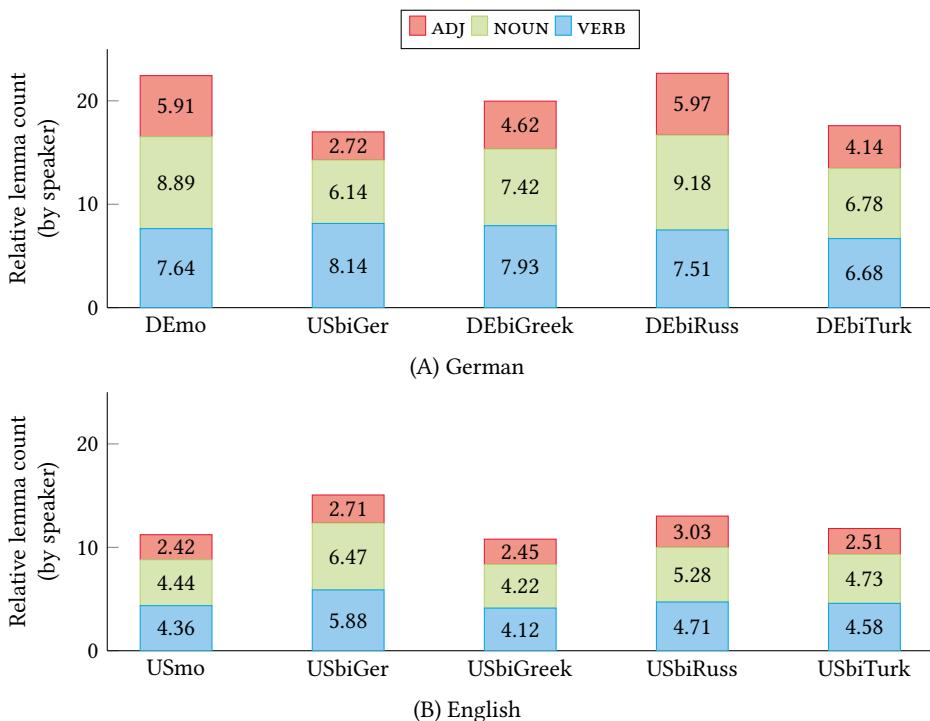


Figure 1: Size of the lexical inventory: Count of adjective, noun, and verb lemma types in the German (A) and English (B) sub-corpus

The English data (Figure 1, plot B) show quite a different distribution: There are differences between the speaker groups in relative frequencies across all categories, with the lowest relative frequencies in the USbiGreek, USmo, and USbiTurk speaker groups. The highest relative frequency is observed for the USbiGer speaker group, while the USbiRuss speaker group is in between, apart from the LI size of the adjectives, which is similar or higher than the one of the USbiGer group. This again contrasts with the results of the LD analyses which showed similar values for all speaker groups, especially the USmo and USbiGer speakers.

A comparison of the distribution between speaker groups in the English and German data (Figure 1, plots A and B) shows that the LI used for the reporting task is bigger in German compared to English for all categories. For the USbiGer speaker group, this difference can mainly be attributed to the LI size of the verb inventory, which is considerably lower in the English data. This may largely be caused by the language-specific lemmatization guidelines applied to the data. It mainly affects German particle verbs (as well as nouns) such as *davonrennen* or *vorbeirennen* and English phrasal verbs like *run off* or *run past* which due to differing orthographic conventions result in varying type counts if the basis for the lemma count is the orthographic word. We refrain from further quantitative cross-linguistic comparisons in this contribution due to this difference in lemmatization. Importantly, a larger LI value does not necessarily relate to the length of the elicited text (compare for instance the USbiRuss and DEbiRuss mean token counts in Table 1 and the respective LI values in Figure 1). Further, a larger LI value for a group does not necessarily imply that each individual speaker within this group uses more different lemmas than speakers of the other groups. A close look at the USbiGer speaker group shows, for instance, that the high average of different lemmas can be traced back to the high variety of nouns and verbs *within* the group. In other words, the USbiGer speakers are very heterogeneous with respect to their choice of lexical items, whereas, for instance, the USmo group shares more lemmas.

The observed heterogeneity within the groups concerning lexical choice suggests two avenues to fathom the dynamic properties of the HS lexicon: (a) a closer look at the speakers behavior within a group (see Sections 5 and 6); (b) an analysis of the intra- and inter-group lexical overlap, or “sharedness”.

We operationalize “sharedness” as the percentage of lemmas used by at least one speaker from group X and one speaker from group Y. Sharedness is positively correlated with LI size: The larger the LI of a group, the higher the chance for any lexeme in the inventory to overlap with a lexeme in the inventory of another group.

The sharedness calculations for the German data reveal that the USbiGer group shares the smallest number of lemmas with each of the other groups. (Appendix, Tables 7–10). Hence, the gradation pattern between the three language profiles, or speaker types, USbiGer, DEbi and DEmo, is consistent with the one obtained from the LD analysis. Furthermore, the heterogeneity of the USbiGer group becomes even more evident when looking at the within-group sharedness of lexemes: Apart from the percentage of shared adjectives, where almost all speaker groups show similar percentages, the USbiGer group demonstrates the lowest percentage of shared lemmas in general and specifically for nouns and verbs.

When we consider the sharedness of lemmas between the two levels of formality or mode (calculated as the percentage of lemmas shared at least once between the two levels of formality or mode, respectively) no clear patterns of sharedness *between* groups arise. However, *within* each speaker group, more lemmas, except for verbs, are shared between the two modes than between the two levels of formality (Figure 2). This suggests that speakers of all groups, including the USbiGers, select lexemes based more on formality than on mode. Additionally, particularly the number of nouns shared between the two levels of formality shows a negative correlation with LI size. In other words, a speaker group which uses larger number of nouns overall (see Figure 1, plot A), such as the DEmoGer and DEbiRuss group, uses more of these nouns in a setting-specific way, as opposed to, for instance, the USbiGer group with a smaller noun inventory and a higher percentage of shared noun lemmas between the formal and informal settings.

The proportion of shared lemma types between the different speaker groups in the English data (Appendix, Tables 11–14) is higher than in the German data. Whereas proportions for German range between 15% and 41%, those for English lie between 27% and 62%. The pairwise comparisons between the USbiGer, the other majority English, and the monolingual English speaker groups return no substantial differences concerning the numbers of shared items between and within groups, which is not surprising as the USbiGer speakers are majority English speakers and are thus of the same speaker type as the other USbi groups, apart from the analytical distinction made in this contribution.

With respect to formality and mode *within* each group (Figure 3), verbs are consistently shared most, followed by nouns and then adjectives. In contrast to the high variability observed in the German data, this holds true for all speaker groups. As in the German data, a higher number of lemma types, irrespective of category, is shared between the two modes than between the two levels of formality. Yet, the inverse relationship between sharedness and LI size cannot be observed in the English data.

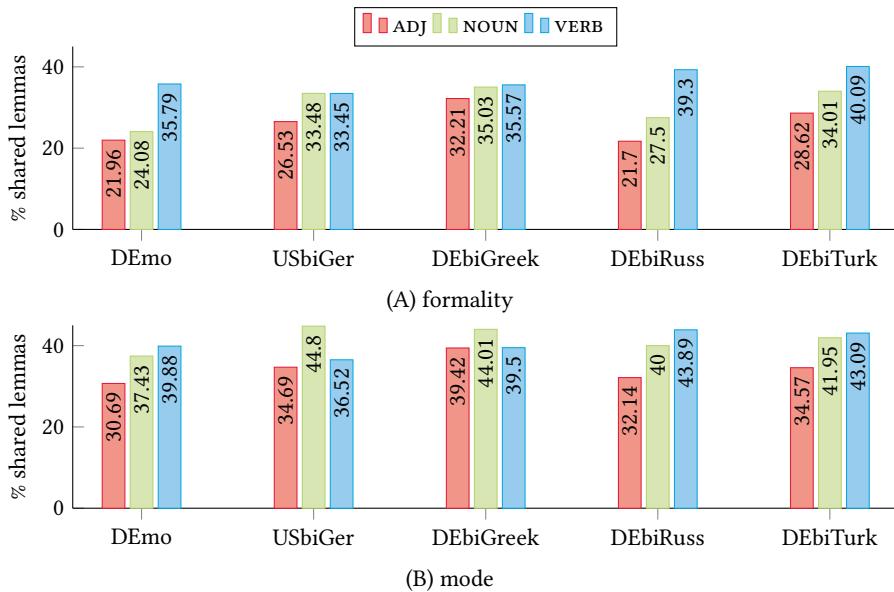


Figure 2: Adjective, noun, and verb lemma types in the German sub-corpus: Shared lemma types across formality (A) and mode (B)

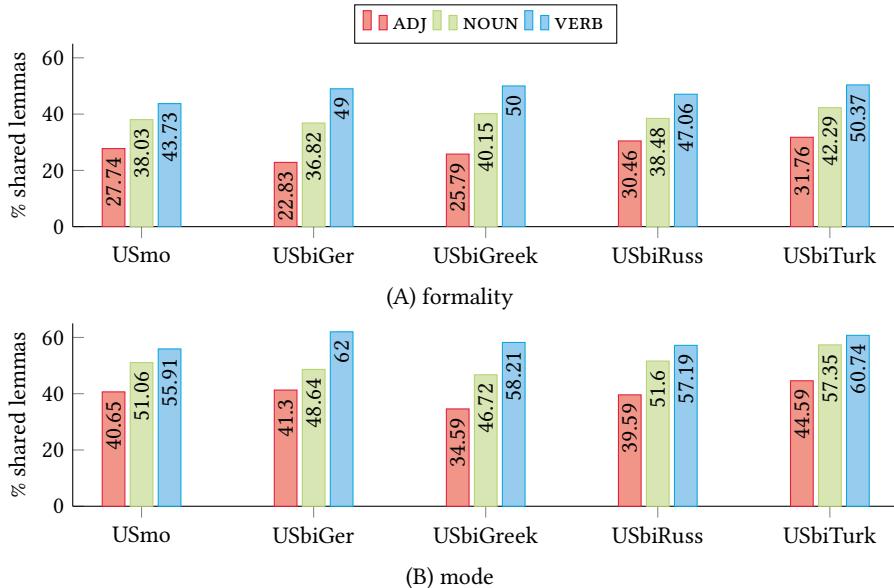


Figure 3: Adjective, noun, and verb lemma types in the English sub-corpus: Shared lemma types across formality (A) and mode (B)

This quantitative and descriptive glance at the LI shows that the number of different lemmas used in German appears to be larger than in English across all groups speaking the language. However, as discussed, this could primarily be due to differences in lemmatization conventions of morphologically complex lexemes between the English and German sub-corpora. In addition, a further mitigating factor is speaker group heterogeneity, specifically observed in the sharedness analysis for the USBiGer group. The particularly low number of shared lexemes *within* the USBiGer speaker group, as shown in Tables 7–10 (Appendix), suggests a high within-group variability that can be traced back to idiosyncratic rather than generalizable speaker group behavior.

Moreover, in both the German and English data, a larger number of lemmas is observed to be shared between modes than between levels of formality within groups, except for the verb category in the German data where the values are similar. This may be due to a between-elicitation priming effect since the spoken and written elicitation session within each formality setting were conducted consecutively, whereas the switch between formality settings was accompanied by separate introductions and a short break.

Our comparison of the three largest lexical classes showed that verbs are shared more often between groups than nouns or adjectives with similar inventory sizes between groups in the German but not in the English data. These findings are particularly interesting in light of Fridman & Meir (2023)'s research who conclude that "noun performance was more likely to diverge from the baseline, while verb performance followed a more monolingual-like trend" (Fridman & Meir 2023: 890), which supports Polinsky (2005)'s argument that "it is less 'costly' for an incomplete learner to lose a noun than it is to lose a verb" (Polinsky 2005: 430). To explore this finding further and to examine whether differences in the lexical repertoire result mainly from differences in lexical choice, we select verbs for a close-up analysis. In the following we compare heritage German and monolingual German speakers. In addition, we decided on starting with a narrow focus on an intricate phenomenon: German particle verbs.

## 5 Challenges in the inventory: German particle verbs

The aim of this section is to show how HSs of German handle the syntactic, semantic and pragmatic challenges posed by particle verbs (PVs) which were laid out in Section 2. We show how a quantitative analysis of the lexicon like the one presented in Section 4 can be augmented by qualitative explorations into a specific aspect of the lexicon to broaden our understanding of the details that characterize the vocabulary choices of heritage speakers of German.

As a preparatory step, we divided all verbs in the heritage German and the monolingual corpus into three groups, according to their morphological characteristics: simplex verbs, prefix verbs (i.e. verbs containing a non-separable prefix) and PVs with a separable particle. Of the 31.933 verb tokens in the German sub-corpus, only 16% are PVs – which is more than the 7% prefix verbs but very little compared to the 77% made up by simplex verbs. However, in terms of verb types, or lemmas, the 598 different PV types in the corpus make up 57%, compared to 271 simplex verbs (26%) and 178 prefix verbs (17%). There is no question, then, that they play an important role in verb inventories of our speakers.

The events in the stimulus lend themselves to coding options via particle verbs (see Section 2). Their description requires the identification of conceptual primitives (see Talmy 1972, Slobin 2003): figures, types of ground (street, parking lot, sidewalk), types of motion (driving, walking, running, rolling, falling, etc.) along various paths, with and without an already perceivable or imagined goal. For the sake of exposition we narrow our focus even more and turn to the most prominent German motion verb in the stimulus: *fahren* ('drive'). Despite this concentration our observations also apply to other verbs in the corpus. Driving events are central to the storyline of the video and account for a large number of types and tokens across speaker groups and communicative situations. Examples and numbers are all based on event descriptions from two subgroups of speakers, namely HSs of German in the US (USbiGer,  $N = 36$ ) and German monolinguals in Germany (DEmo,  $N = 64$ ). The data contain 138 tokens of *fahren* as a simplex verb, plus 369 tokens and 26 types of PVs with *fahren* as their base.

In the case of polysemous PVs, meaning can only be determined in context. Take the PV *anfahren*, which is among the top-five most frequently used PVs in both speaker groups. According to one of the major German dictionaries, Duden online, the verb *anfahren* has nine clearly differentiated subsenses, most of which cannot be inferred from the combined meanings of the particle and the verb. They are as diverse as 1. beginning to move, 2. approaching, 3. rearending, or 4. angrily shouting at someone. In the USbiGer data, *anfahren* is used primarily in the sense of approaching and entering new ground, such as turning into a street or a parking lot (Duden online, Subsense 2; Example 7). This is perfectly idiomatic and is also used by our DEMo speakers. However, in the DEMo data, *anfahren* is used much more often in the sense of accidentally hitting a target while driving a vehicle (Duden online, Subsense 6; Example 8). This sense is barely used by the HSs. Instead, we find the verb *treffen*, a loan translation of the English verb *hit*, as in Example (9).<sup>12</sup>

---

<sup>12</sup>As mentioned before, in order to preserve a clear focus, we decided against comments on non-canonical realizations like the auxiliary 'haben' in Example (7) and other non-canonical forms (case, gender) not at issue here.

- (7) als die zwei autos **an-ge-fahr-en** hatten  
when the two cars on.VPART-PTCP-drive-PTCP have.PST.3PL  
'when the two cars were approaching' (USbi71MD\_fsD)
- (8) vermutlich aus Angst den Hund **an-zu-fahr-en**  
probably for fear the.ACC dog on.VPART-to-drive-INF  
'probably for fear of hitting the dog.' (DEmo47MD\_fwD)
- (9) um den hund nicht zu **treff-en**  
for the.ACC dog not to hit-INF  
'so as not to hit the dog' (USbi71MD\_fwD)

Hence, even though a pure count of PVs suggests that both heritage speakers and monolinguals make frequent use of the same verb (see Section 4), there may be subtle but crucial differences on the semantic level, as we proceed to show.

One difference between monolingual and heritage speakers concerns the semantics of particles with respect to argument structure. The two most frequent particles combined with *fahren* are adverb-based *rein* in the USbiGer data and preposition-based *auf* in the DEmo data, along with its adverb-based variants *rauf* and *drauf*. However, both particles are used by either group. The subsense of the PV correlates in interesting ways with either NP complements expressing an affected object or PP adjuncts expressing detail on path of movement. When *reinfahren* is used in the sense of entering new ground, i.e. crossing a boundary, both USbiGer (Example 10) and DEmo speakers (Example 11) frequently add a PP complement. However, most of the time the PV *reinfahren* is used in the sense of rearending another car, a telic event and a momentary achievement in the sense of Vendler (1957). In these cases, DEmo speakers prefer dative NP complements (20 out of 31 clauses; Example 12), while USbiGer speakers almost exclusively choose PP complements (33 out of 35 clauses; Example 13).

- (10) das blaue auto was **in den parkplatz rein-ge-fahr-en**  
the blue car which in the.ACC parking.lot into.VPART-PTCP-drive-PTCP  
ist  
be.PRS.3SG  
'the blue car which drove into the parking lot' (USbi72FD\_isD)
- (11) und zwei autos sind **rein-ge-fahr-en** **in den parkplatz**  
and two cars be.PRS.3PL into.VPART-PTCP-drive-PTCP in the.ACC  
parking.lot  
'and two cars drove into the parking lot' (Demo38FD\_isD)

- (12) der eine is           **dem**   **ander-en** hinten  
      the one be.PRS.3SG the.DAT other-DAT in.back  
      **rein-ge-fahr-en**  
      into.VPART-PTCP-drive-PTCP  
      ‘and the one rearended the other one’ (DEmo19FD\_isD)
- (13) Und dann **fuhr**           das zweite auto **in das erste rein**  
      and then drive.PST.3SG the second car in the first into.VPART  
      ‘and then the second car rearended the first one’ (USbi04FD\_fwD)

The encoding of path information in German does not necessarily require a PV. It may also be expressed by a PP. When direction or goal of motion is already expressed by a simplex verb combined with a PP, adding a corresponding particle to the verb can be semantically redundant. At the same time, it is perfectly canonical in German to do so. It is quite interesting, therefore, that both speaker groups differ with respect to double marking in connection with motion events. In the DEmo data, simplex *fahren* together with a directional *auf*-PP is used to describe entering new ground, for example a parking lot (as in Example 14). Double marking critical subevents of motion with the help of the verbal particle *auf*- plus an *auf*-PP is used when the focus is on the endpoint of the motion event, i.e. when the event is telic, as in our texts describing the second car hitting the first one (Example 15). In other word, monolingual speakers use single versus double marking to convey subtle semantic differences in motion events.

- (14) ein Auto, was gerade **auf den Parkplatz ge-fahr-en** ist  
      a car which just on the parking.lot PTCP-drive-PTCP be.PRS.3SG  
      ‘a car which was just driving into the parking lot’ (DEmo88FD\_fwD)
- (15) Aufgrunddessen ist           **das hintere auf das vordere Auto**  
      Because.of.this be.PRS.3SG the back on the.ACC front car  
      **auf-ge-fahr-en**  
      on.VPART-PTCP-drive-PTCP  
      ‘For this reason, the car in back rearended the car in front’  
      (DEmo53FD\_fwD)

In the USBiGer data, the simplex *fahren*, together with a directional PP, is used to describe entering new ground, like the parking lot in the stimulus video. Most often, however, the PV *reinfahren* is used, along with the matching preposition

in (Example 16).<sup>13</sup> The same construction of a PV together with a directional PP is used for describing the act of rearending another car (Example 13).

- (16) Und es waren zwei autos die im pa/ parkplatz  
and it be.Pst.3PL two cars which in.DAT parking.lot  
**rein-ge-fahr-en** sind  
into.VPART-PTCP-drive-PTCP be.PRS.3PL  
'and there were two cars entering the parking lot' (USbi72FD\_fwD)

Examples like (16) suggest that in the USbiGer data, entering new ground and endpoint orientation of motion verbs are lexicalized in the same way. Looking at other PVs as well, double marking of direction and endpoint of motion is the preferred option in the USbiGer data. Since Pashkova et al. (2020) have shown evidence for increased explicitness in heritage speaker productions in their majority language, double marking in their heritage language may reflect this as well. However, in the case of German-English bilinguals, structural parallelism (*drive/run + PP*) and potential cross-linguistic effects must not be disregarded either.

Since previous research suggests that lexical choice according to register differentiation is particularly challenging, we now briefly turn to the use of PVs in specific communicative situations (formality, mode). The most frequent PV with the base *fahren* in the DEMo data, *auffahren* occurs most often in the formal-written setting (see 17). In the informal settings, the particle is often realized as *rauf*, evoking direction or target without overtly combining with a deictic argument, resulting in *rauffahren* (see 18).

- (17) auffahren: fw (40) > fs (28) > iw (16) > is (16)  
(18) rauffahren: is (30) > iw (18) > fs (14) > fw (6)

Even though the quantitative analysis of shared lemmas across different variables in Section 4 shows that the USbiGer speakers as a group are sensitive to formality distinctions in their selection of lexical items, differentiation according to formality and mode is not found for the most frequent *fahren* PV in the data; *reinfahren* is used across all four communicative situations for referring to the telic event of rearending a car. While this is perfectly acceptable in colloquial German, in more formal situations, like providing a witness report, *auffahren* would seem

<sup>13</sup>The form of German verbal particles derived from prepositions is usually exactly the same as the preposition. Only the preposition *in* changes to *ein-* when it is used as a verbal particle. The same holds for the deictic adverbials derived from the preposition *in* (*herein*, *hinein*, *rein*).

more appropriate. Having grown up in Germany, DEMo speakers are more likely to know that the PV *auffahren* is used specifically to describe the rearending of a car and that in official accounts accidents like this would be referred to as *Auf-fahrunfall*.<sup>14</sup> In contrast, many of our US-based HS participants are not likely to have encountered reference to many car accidents in German, especially not in formal contexts. Additionally, the frequent choice of *reinfahren* might be influenced by the speakers's dominant language, English: *Reinfahren* combines with the preposition *in*, which, as a homophonous diamorph, facilitates the transition between languages (see Clyne 1967, Muysken 2000: 133). Example (19) is a case in point.

- (19) als die zwei autos (-) die in einer (-) anderen straße  
   when the two cars \*\* which in a.DAT \*\* different.DAT street  
   fuhr-en (-) in ähm (-) the (-) parking lot ein-ge-bog-en  
   drive-PST.3PL \*\* in \*\* \*\* the \*\* parking lot in.VPART-PTCP-turn-PTCP  
   sind  
   be.PRS.3PL  
   ‘when the two cars which were driving along a different street turned  
   into the parking lot’ (USbi65MD\_fsD)

The speaker chooses the particle *ein-* together with the base *biegen*, a canonical construction in German, but the complement of the PP, encoding the goal of the directed motion, is realized in English. The preposition itself is unspecified for language: It can be German or English or both at the same time. Furthermore, the utterance contains short pauses and hesitation particles, hence exactly the type of production phenomena we turn to in Section 6.

This short illustration of the intricacies of German PVs that a speaker is faced with shows several minute differences between HSs and MSs which escape a quantitative assessment based on type and token counts. Given what is known about the early appearance of particles and particle verbs (see Tsehayé et al. 2025 [this volume]), it comes as no surprise that HSs have no problem with the most notorious syntactic feature of PVs, namely their distributional properties. Nevertheless, in the case of the specific motion verbs considered, individual heritage speakers arrive at slightly different conclusions than monolingual speakers of German with respect to particle choice, both in terms of meaning and register.

---

<sup>14</sup>Duden online lists rearending another car as the first subsense of *auffahren*, whereas this subsense is not listed at all for (*he-)reinfahren*. Nevertheless, whether or not a PV is listed in the Duden does not say much about actual use. As the formation of PVs is productive and often transparent, the Duden mostly lists them once a specific meaning has become lexicalized.

The following third section of our analysis looks at production phenomena with the question in mind what they reveal to us about moments of choice.

## 6 Producing particle verbs in real time

The speed with which speakers access their word store and within fractions of a second select from tens of thousands of available options those that fit an intended message is impressive. Thanks to efficient pro- and retroactive monitoring skills speakers can swiftly edit unintended messages and occasional slips of the tongue.<sup>15</sup> In the context of our current discussion, both the very fact of self-initiated interventions in specific places and the overt details in which they play out provide insight into speakers' "personal" view on their own messages.

Speakers' self-monitoring manifests itself in various performance phenomena besides the actual reparans (i.e. the correction): interruptions with and without hesitating, syllable lengthening, iterations, and sometimes speaker-specific fillers, such as tongue-clicks. Essentially, these phenomena are proliferous in everybody's unrehearsed speech (see Fromkin 1973, Hieke et al. 1983, Levelt 1989, Clark & Fox Tree 2002, Belz 2023). The downside is that the identification, transcription and annotation of relevant data involves painstaking attention to phonetic detail and acoustic measurement with respect to timing. As shown by Belz (2021), German hesitation particles alone – in the German data of the RUEG corpus orthographically transcribed as *äh* or *ähm* – occur in many phonetic shapes. Moreover, form is one thing, function another. Hence it is difficult, and may often be downright impossible, to unambiguously attribute a particular phonetic event to a specific challenge speakers face. Nevertheless, as we demonstrate here, along with other publications based on the RUEG corpus dealing with performance issues (Böttcher & Zellers 2023, Tracy & Gibbon 2023), self-initiated changes in utterances are revealing, especially with respect to what they tell us about lexical inventories. Similar to Levelt (1989)'s elicitation of speech errors in descriptions of paths taken through a visual array, our event narrations yield useful information on what is there to choose from.

In the oral RUEG narrations, regardless of minority or majority speaker status or language, overt and covert performance phenomena are attributable to various types of pro- and retroactive repairs of word selection or message construction,

---

<sup>15</sup>See Levelt (1989)'s Main Interruption rule and different motivations for reformulations and repairs. As he stresses, "[s]peakers can monitor for almost any output of their own speech" (Levelt 1989: 436).

and also to discourse-related motivations (e.g. change of topic, see other chapters of this volume). In the heritage German speaker data, we find an abundance of word iteration and modification in specific, predictable trouble spots related to challenges involving gender and case marking, in amalgamations of articles and prepositions, plural inflection in nouns, auxiliary choice, and participle morphology. None of these come as a surprise, given the eccentric, highly irregular nature of the paradigms involved. Some of these non-canonical features will be seen in the examples below, but we will not draw attention to them unless they are related to our immediate concern. While we maintain our main focus on German particle verbs, we now also include cues in their vicinity that provide insight into local troubleshooting.

Before we start on our analysis of the HSs productions, it must be pointed out that some of the monitoring phenomena discussed occur in monolingual speakers as well, as shown in Example (20). The speaker here replaces a partially uttered simplex verb of motion (*fahren*, ‘drive’, our model verb from Section 5), with a PV. The particle *lang*, a short form of *entlang* ‘along’ requires a deictic adverbial or object NP expressing path information concerning the region along which motion takes place. While reference to location is opaque, the obligatory argument position is filled, resulting in a syntactically well-formed clause conversationally adequate in informal contexts.

- (20) und dann sind        halt    (-) zwei autos **gefah/ da**  
       and then be.PRS.3PL simply \*\* two cars    \*\*        there  
**lang-ge-fahr-en**  
       along.VPART-PTCP-drive-PTCP  
       ‘and then two cars came driving along there’(DEmo57FD\_isD)

Next, consider the formal and informal reports from a German HS in Examples ((21)–(23)).

- (21) die war        grad ähm (-) [tcl] äh einkauf-en ge-gang-en    und ähm  
       she be.PST.3SG just \*\*    \*\* \*\*    \*\* shop-INF PTCP-go-PTCP and \*\*  
       woll-te        alles        in/ in-s        auto ähm **hin-tun**  
       want-PST.3SG everything \*\* in-the.ACC car    \*\*        there.VPART-put.INF  
       ‘she had just been shopping and wanted to put everything in the car’  
       (USbi03FD\_fsD)

- (22) und ähm (-) sie woll-te ihre sachen im äh (-) auto ähm (-) [tcl]  
 and \*\* \*\* she want-PST.3SG her stuff in.DAT \*\* \*\* car \*\* \*\* \*\*  
 äh (-) pa/ **hin-pack-en** (-)  
 \*\* \*\* \*\* there.VPART-pack.\*\*  
 ‘and she wanted to put her stuff in the car’ (USbi03FD\_isD)
- (23) äh dieses auto äh muss-te ga/ a/ auch ganz schnell stopp-en und  
 \*\* this car \*\* must-PST.3SG \*\* \*\* also really fast stop-INF and  
 is eigentlich ähm in-s erst/ erstes auto (-) äh  
 be.PRS.3SG actually \*\* in-the \*\* first car \*\* \*\*  
**hin-ge-fahr-n** äh **rein-ge-fahr-n**  
 there.VPART-PTCP-drive-PTCP \*\* into.VPART-PTCP-drive-PTCP  
 ‘this car had to stop really fast and actually drove to uhm bumped into  
 the first car’ (USbi03FD\_fsD)

We can identify lavishly distributed hesitation particles, reiterations and tongue-clicking leading up to proactively perceived troublespots. But what happens to the verbs? In the first two cases (Examples (21) and (22)), not going for a particle would actually have resulted in perfectly well-formed and contextually adequate expressions: *ins Auto tun* ‘put into the car’ and *ins Auto packen* ‘pack/load into the car’. However, the result is marginally (*hintun*) or more substantially (*hindenken*) odd, given the particular container, a car.

In the last case, the speaker produces what by anybody’s standard – and obviously by her own – is in need of repair (*hingefahren* → *reingefahren*). Interestingly, both written reports – controlled typing activities extending the time for corrections – of the same participant confirm her preference for the structure repaired to in Example (23), for one car hitting the other (Examples (24) and (25)).

- (24) Und dann ist ein zweites auto im parkplatz **ge-fahr-en**  
 And then be.PRS.3SG a second car in.DAT parking.lot PTCP-drive-PTCP  
 und ist im erstes auto **rein-ge-fahr-en**.  
 and be.PRS.3SG in.DAT first car into.VPART-PTCP-drive-PTCP.  
 ‘And then a second car drove into the parking lot and rearended the first  
 car’ (USbi03FD\_fwD)
- (25) ein auto muss-te ganz schnell stopp-en und ein anderes auto  
 a car must-PST.3SG really fast stop-INF and a different car  
 hat in-s ersten auto **rein-ge-fahr-en**  
 have.PRS.3SG in-the first car into.VPART-PTCP-drive-PTCP  
 ‘a car had to stop really fast and another car rearended the first car’  
 (USbi03FD\_iwD)

With hindsight it is unfortunate that the spectrum of RUEG data elicitation methods did not include tracking self-corrections and timing in written narratives. Nevertheless, the fact that participants repeatedly refer to the same scenes makes it possible to identify candidates for stable lexicalizations differing from conventional items. Examples (26) and (27) serve as cases in point. At first sight the non-canonical participle *hintergerannt* (lit. ‘behind-run’, intended ‘follow’, instead of the canonical *hinterhergerannt* ‘follow’) in Example (26) sounds like a one-shot speech error due to syllable elision. Yet, the same participle recurs in the formal-written scenario, supporting the assumption that *hinterrennen*, most likely strengthened by parallelism with the English *run after*, except for order, though innovative from a canonical perspective, Example (27) then constitutes an idiosyncratic conventionalized particle verb for this particular speaker.

- (26) ... ist ein hund vom rand der sträße (-) andere seite der/  
     be.PRS.3SG a dog from.DAT edge the.GEN street \*\* other side \*\*  
     dem ball **hinter-ge-rann-t**  
     the.DAT ball after.VPART-PTCP-run-PTCP  
     ‘a dog from the curb of the street, the other side, ran after the ball’  
     (USbi64MD\_fsD)
- (27) ... ist ein Hund von der andere seite der strass den ball  
     be.PRS.3SG a dog from the other side the.GEN street the.ACC ball  
     **hinter-ge-rann-t**  
     after.VPART-PTCP-run-PTCP  
     ‘a dog from the other side of the street ran after the ball’  
     (USbi64MD\_fwD)

Quite subtle cross-linguistic interactions can be seen in the following two examples. The particle verb *aufringen* (Example 28) is documented both in speech and in writing (we only quote the written version here). The particle *auf* is from German and the verb *ringen* is a borrowing from English *to ring* (to call). *Aufringen* may be morphologically possible but is not a verb used in German – it is a calque modelled after the English phrasal verb *ring up*. Once more, it is the repeated use of this verb which allows us to consider it an innovation enriching the speaker’s German repertoire.

- (28) **ring-t** mich auf  
     ring-PRS.2PL me.ACC up.VPART  
     ‘Ring me up’ (USbi77FD\_iwD)

Our final illustration is more intricate. In Example (29) the PV *vor sich hin-treiben* (in-front-of oneself there-drive), which in German can take a collocate PP, is arguably based on English *drive*. The PV construction augmented by ‘vor sich hin’ exists in German, but with the meaning of either actively chasing something or passively drifting along. What makes this instance particularly relevant is the speaker’s prompt attempt at a repair of the first calque by resorting to an existing but equally “off the mark” PV.

- (29) ... der einen ball **vor**      **sich**      **hin**      (-) ä:h hum/ (-)  
      who a-ACC ball in.front.of self.REFL there.VPART \*\* \* \* \*\* \*\*  
**vor**      **sich**      **hin-treib-t**      (-) ä:h oder  
      in.front.of self.REFL there.VPART-drive-PRS.3SG \*\* \* \* or  
**hin-spiel-t**  
      there.VPART-play-PRS.3SG  
‘who is driving uhm playing a ball in front of himself uhm playing’  
(USbi68MD\_isD)

The RUEG narratives provide a plethora of evidence for within-language and cross-linguistic networking, as can be seen in the spontaneous self-corrections presented here. With respect to German particle verbs, heritage speakers have all it takes in terms of the basic building blocks and combinatory principles, i.e. all of the morphological resources they need, and they struggle with details of choice if put on the spot, for example in a challenging experimental situation.

## 7 Discussion and conclusion

Heritage language research provides us with privileged access to studying which properties of early grammars remain stable when the languages of our childhood are sent to the backstage and exposure decreases. HS data also provide clues to what is likely to change, either due to dynamics of internal language change (such as regularization of irregular word forms), or as a consequence of intensive contact with a specific majority language (cf. Tsehay et al. 2025 [this volume]). Our contribution explored the lexicon of HSs, a domain of our linguistic competence which, regardless of speaker type, is highly dynamic: As we stated initially, the lexicon is a moving target.

For a quantitative assessment of lexical inventories we compared different groups of bilingual heritage speakers of Turkish, Russian, Greek and German in their English and German productions as well as the respective non-heritage monolingual speakers (Section 4). First of all, this revealed considerable intra-group

heterogeneity which can be attributed to significant differences in speaker background variables and to different interpretations of the elicitation tasks by the participants. Despite this heterogeneity, group comparisons show that specific subsets of the lexicon are not only comparable in size, *within* a language, but are also shared between the majority and monolingual speaker groups of either German or English. This is the case particularly for the verb inventory in the German data. This is not surprising in light of the findings on verb maintenance for HSs (Polinsky 2005, Fridman & Meir 2023) and further in terms of task demand, since the elicited descriptions relate the same events, while reference to the animate and inanimate protagonists involved is more diverse.

Additionally, the gradation pattern from MSs via MajS to HSs established by way of LD analysis confirmed our expectations based on previous research regarding reduced HL input and assumptions based on the available speaker metadata. The LD analyses without further qualitative assessment of the individual utterances do not support a tendency of HSs towards register leveling, in contrast to previous findings on HSs (e.g. Wiese et al. 2022) and LD in general (Van Gijsel et al. 2005, Yu 2009, Alamillo 2019). It is plausible that even though the diversity score (here MATTR) differs between levels of formality and mode, the lemma types embedded in that score do not, as we have seen for the most frequent particle verb *reinfahren* in the HS productions. This indicates that HSs may have a limited register-specific repertoire, yet are still able to differentiate between registers by using their resources in a diversified manner.

With respect to LD and LI analyses of data from different languages, a major methodological concern should be mentioned: In German and English contact situations, typological closeness not just creates descriptive challenges for transcribing and annotating the data. Local ambiguity makes automatic lemmatization difficult and requires considerable manual correction based on token-by-token-in-context decisions (see Wiese et al. 2025 [this volume]). However, in the RUEG corpus this detective work is supported by the availability of different texts on the same events produced by the same participants, allowing us to pursue questions relating to local problems (selecting from sets of highly similar particles, word searches, etc.) and individual coping strategies. In the case of particle verbs, our analysis shows how ambiguity due to different intended readings are only resolved by paying close attention to verbal contexts and to the sub-event of the accident focused on by the speaker.

The qualitative analysis of German particle verbs (Section 5) confirms our initial assumption that the very nature of the input material, i.e. the RUEG video stimulus, is well suited for eliciting verb bases referring to types and manner of motion which select for semantically relevant satellites: PPs, verbal particles,

and additional deictic elements satisfying argument positions. Verbal particles offer a substantial, though sometimes only minimally differing inventory of signs for identifying locations, paths, directions and goals, and for turning atelic processes into telic ones. In view of this multitude of formal-functional detail to be worked out in acquisition and managed in real-time tasks, the often-cited syntactically excentric status of combinations of verbs and particles seems downright insignificant. Our data provide no evidence that HSs struggle with the syntactic positioning of verbs and particles, no matter whether they are realized as one continuous string or split up between the left and right sentential bracket. We see, however, that HSs do not always choose the same lexical means as the MSs to convey meaning. We identified subtle differences between heritage and non-heritage speakers in terms of meaning shift and register. Moreover, we find that HSs, more so than monolinguals, tend to express specific subevents redundantly through both particles and prepositional phrases. This finding supports the hypothesis initially mentioned in Polinsky (2018: 294–295) that HSs tend to prefer compositionally transparent and explicit formulations.

The findings concerning the analysis of PVs in Section 5 are corroborated by our exploratory discussion of speech production in Section 6, which shows that particle verbs provide a good starting point for investigating local challenges due to minute contrasts in form, as in word-onsets such as *auffahren*, *rauffahren*, *drauffahren*. Both proactive signals of trouble and the direction and result of self-initiated change supply us with evidence for the individual lexical inventory and for the morphological tools needed for word formation. It cannot be overstated that all these performance phenomena are self-initiated, hence pointing to speaker awareness that alternative expressions were not just available but sometimes called for.

As mentioned in Section 2, particle verbs play an important part in children's early lexicons, and so does, for German-speaking children, the expression of telicity. It may well be the case that the bias towards redundant marking of path and goal discussed here echos child-directed registers. As discussed by Bryant (2018: 177), parents tend to go for the redundant marking of location or goal, which means via both particles and prepositional phrases: *Immer auf'n Tisch die Schalen draufwerfen* (lit. always on-the table the peels onto-throw, 'always throw the peels onto the table'). Also, as stated by Polinsky (2018: 291–328), majority speakers tend to consider the HS variety of the same language pragmatically peculiar and inadequate in view of a speaker's age. Unfortunately we did not meet our participants in early childhood and had no access to their parental baseline. Hypotheses concerning childhood input are thus waiting to be pursued in future studies.

Heritage speakers help us answer fundamental questions related to language learnability and maintainance: How can humans learn so much even under reduced input conditions and with the L1 under increasing pressure from a dominant, and possibly very similar, hence, distracting, language? As we have shown, heritage language speakers have an important part to play in solving puzzles related to acquisition, language change, and highly competent performance. Since HSs are not lost for words, as shown here, we can be optimistic.

## Acknowledgements

All authors gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the Research Unit *Emerging Grammars*, grant number 313607803, Project P11 – Lexicon. Additionally, Anke Lüdeling acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412 *Register*. Special thanks are owed to the RUEG team, our Mercator fellows, two anonymous external and our three internal reviewers Natalia Gagarina, Luka Szucsich and Sabine Zerbian, to RUEG cooperation partner Dafydd Gibbon for comments on earlier versions of this paper, to Elena Krotova for continuing support with data annotation and analysis in Python, as well as to our student research assistants Stella Baumann, Joshua Boivin, Anna Kuhn, Nils Picksak, Elena Unger for their tireless work from first data preparation to final proof-reading. We additionally thank Lea Coy for help with pre-publication formatting.

## Abbreviations

|                             |                                 |
|-----------------------------|---------------------------------|
| HS                          | Heritage Speaker                |
| HL                          | Heritage Language               |
| L1                          | First Language                  |
| L2                          | Second Language                 |
| LD                          | Lexical Diversity               |
| LI                          | Lexical Inventory               |
| MATTR                       | Moving Average Type-Token Ratio |
| MajL                        | Majority Language               |
| MS                          | Monolingual Speaker             |
| PV                          | Particle Verb                   |
| R <sup>2</sup> <sub>c</sub> | conditional r-squared           |
| RUEG                        | Research Unit Emerging Grammars |

## Appendix

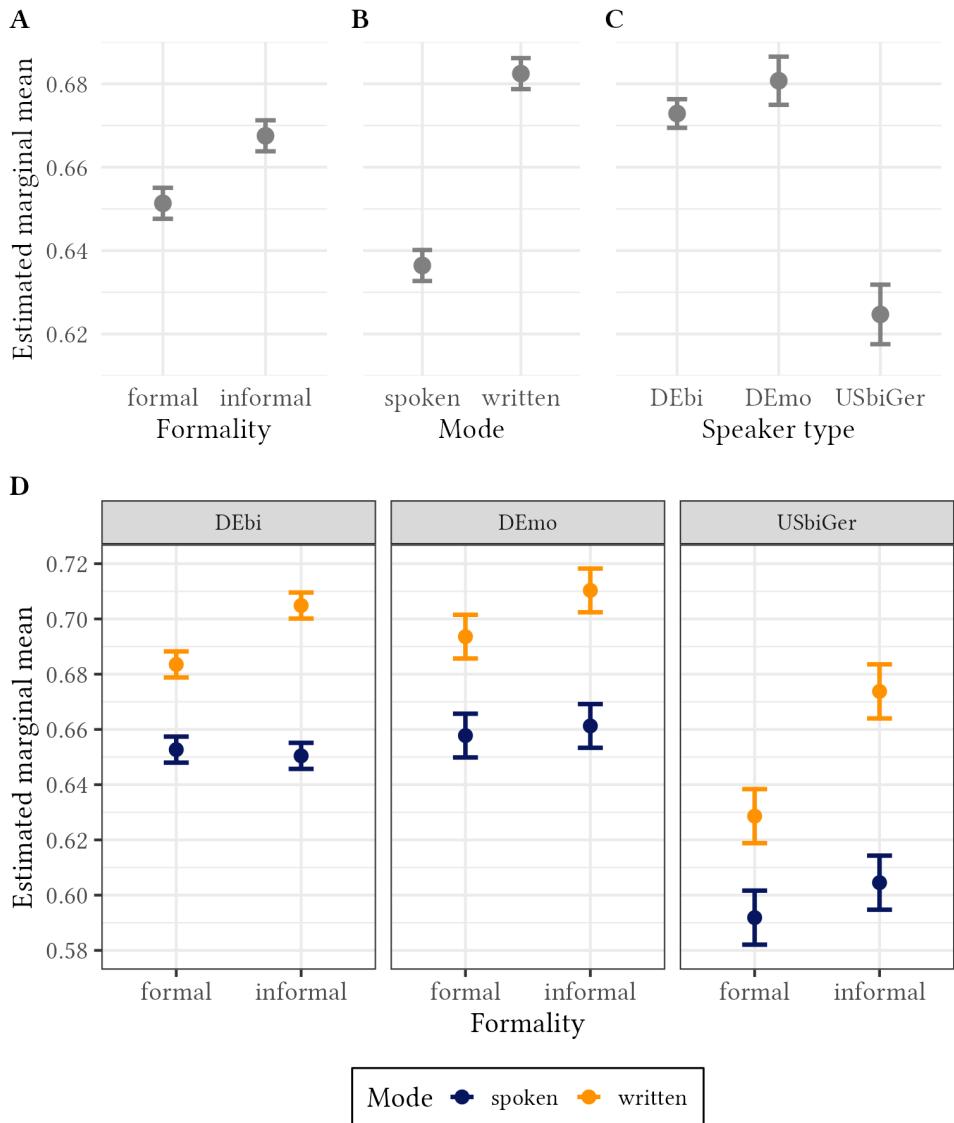


Figure 4: Model 1, predicted values of MATTR in the German data:  
Formality (A), Mode (B), Speaker Type (C), and the Interaction (D)

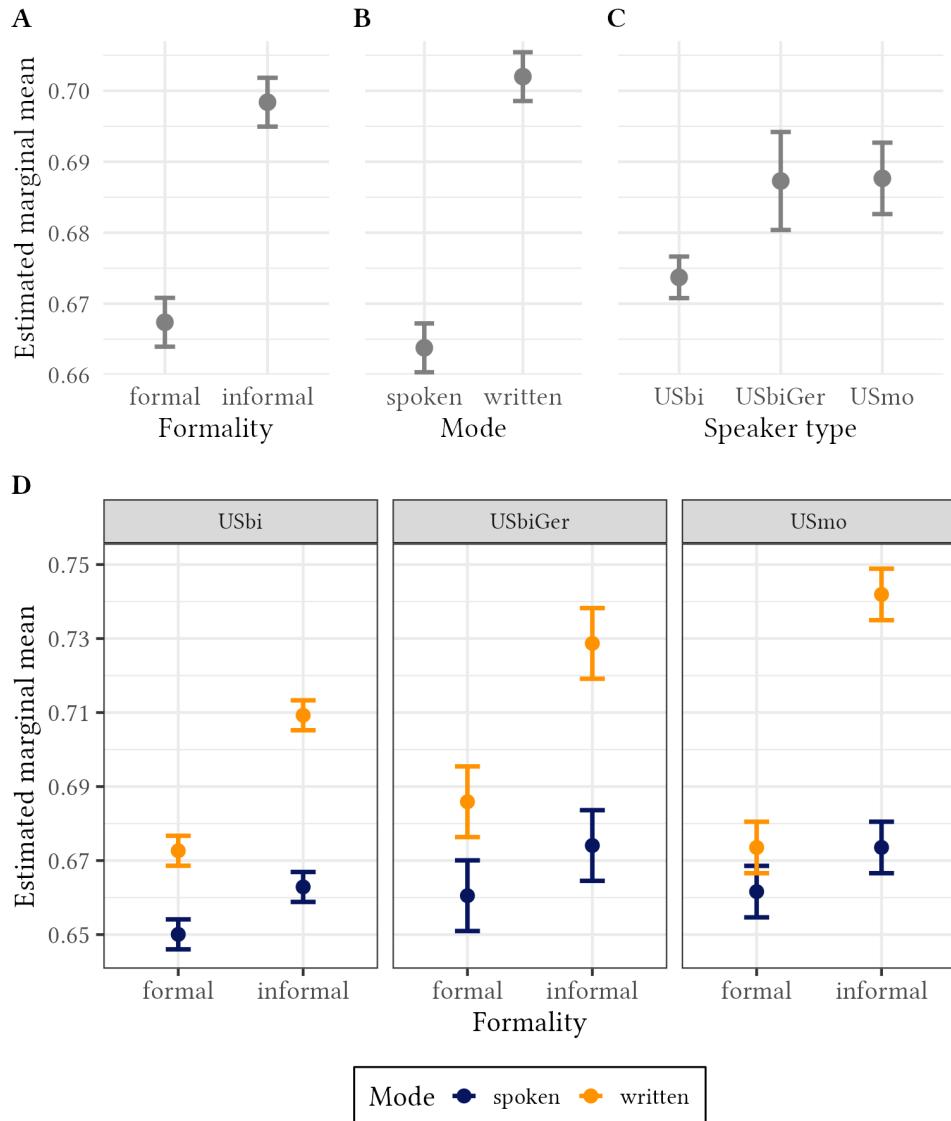


Figure 5: Model 2, predicted values of MATTR in the English data: Formality (A), Mode (B), Speaker Type (C), and the Interaction (D)

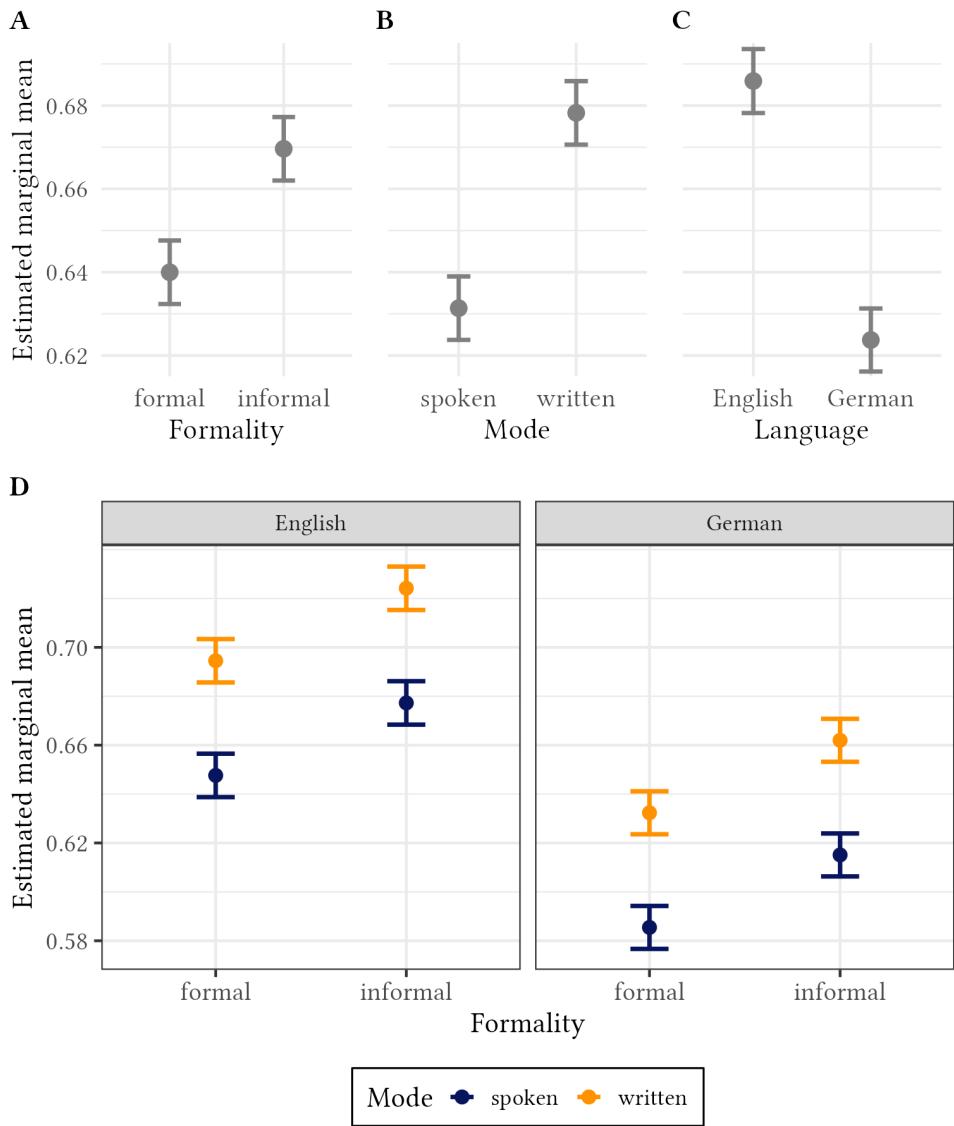


Figure 6: Model 3, Predicted Values of MATTR in the USBiGer Data:  
Formality (A), Mode (B), Speaker Type (C), and the Interaction (D)

Table 4: Model on the MATTR measurements in the German data

| Predictors  | $\beta$   | SE   | CI              | Statistic | p     | df  |
|---|-----------|------|-----------------|-----------|-------|-----|
| (Intercept)   | 0.66      | 0.00 | [ 0.65, 0.67]   | 202.94    | <0.01 | 246 |
| speaker type: USbiGer-DE <sup>a</sup>                 | -0.02     | 0.01 | [ -0.03, -0.01] | -3.53     | <0.01 | 246 |
| speaker type: DEMo-DEbi                               | 0.06      | 0.01 | [ 0.04, 0.07]   | 6.14      | <0.01 | 246 |
| formality: F-I  | 0.02      | 0.00 | [ 0.01, 0.02]   | 4.61      | <0.01 | 738 |
| mode: S-W   | 0.05      | 0.00 | [ 0.04, 0.05]   | 13.12     | <0.01 | 738 |
| session: 2-1  | 0.00      | 0.00 | [ 0.00, 0.01]   | 1.03      | 0.30  | 738 |
| session: 3-2  | 0.00      | 0.00 | [ -0.01, 0.01]  | -0.13     | 0.90  | 738 |
| session: 4-3  | 0.00      | 0.00 | [ -0.01, 0.01]  | 0.36      | 0.72  | 738 |
| speaker type: USbiGer-DE × formality: F-I             | 0.01      | 0.01 | [ 0.00, 0.02]   | 1.62      | 0.11  | 738 |
| speaker type: DEMo-DEbi × formality: F-I              | -0.02     | 0.01 | [ -0.04, 0.00]  | -1.91     | 0.06  | 738 |
| speaker type: USbiGer-DE × mode: S-W                  | 0.01      | 0.01 | [ -0.01, 0.02]  | 0.83      | 0.41  | 738 |
| speaker type: DEMo-DEbi × mode: S-W                   | -0.01     | 0.01 | [ -0.03, 0.01]  | -1.07     | 0.29  | 738 |
| formality: F-I × mode: S-W                            | 0.02      | 0.01 | [ 0.01, 0.04]   | 3.29      | <0.01 | 738 |
| speaker type: USbiGer-DE × formality: F-I × mode: S-W | 0.00      | 0.01 | [ -0.02, 0.02]  | -0.06     | 0.95  | 738 |
| speaker type: DEMo-DEbi × formality: F-I × mode: S-W  | -0.02     | 0.02 | [ -0.06, 0.02]  | -0.98     | 0.33  | 738 |
| Random effects  |           |      |                 |           |       |     |
| $\sigma^2$  | .00       |      |                 |           |       |     |
| $\tau_{00\text{ID}}$                                  | .00       |      |                 |           |       |     |
| ICC   | .38       |      |                 |           |       |     |
| N <sub>ID</sub>                                       | 246       |      |                 |           |       |     |
| Observations  | 984       |      |                 |           |       |     |
| Marginal R <sup>2</sup>                               | .210      |      |                 |           |       |     |
| Conditional R <sup>2</sup>                            | .509      |      |                 |           |       |     |
| AIC   | -2804.619 |      |                 |           |       |     |

<sup>a</sup>The DE group includes all monolingual and majority German speakers, DEMo and DEbi respectively.

Table 5: Model on the MATTR measurements in the English data

| Predictors   | $\beta$  | SE   | CI            | Statistic | p     | df     |
|--|----------|------|---------------|-----------|-------|--------|
| (Intercept)  | 0.68     | 0.00 | [ 0.68, 0.69] | 228       | <0.01 | 287    |
| speakertype: USbiGer-US <sup>a</sup>                 | 0.01     | 0.01 | [ 0.00, 0.02] | 2.67      | 0.01  | 287    |
| speakertype: USmo-USbi                               | 0.00     | 0.01 | [−0.02, 0.02] | −0.04     | 0.97  | 287    |
| formality: F-I                                       | 0.03     | 0.00 | [ 0.02, 0.04] | 9.40      | <0.01 | 861    |
| mode: S-W  | 0.04     | 0.00 | [ 0.03, 0.04] | 11.59     | <0.01 | 861    |
| session: 2-1   | 0.00     | 0.00 | [ 0.00, 0.01] | 1.00      | 0.32  | 861.45 |
| session: 3-2   | 0.00     | 0.00 | [−0.01, 0.01] | 0.50      | 0.62  | 861    |
| session: 4-3   | 0.00     | 0.00 | [−0.01, 0.01] | −0.23     | 0.82  | 861.46 |
| speakertype: USbiGer-US × formality: F-I             | 0.01     | 0.01 | [ 0.00, 0.02] | 1.67      | 0.10  | 861    |
| speakertype: USmo-USbi × formality: F-I              | −0.01    | 0.01 | [−0.03, 0.01] | −1.28     | 0.20  | 861    |
| speakertype: USbiGer-US × mode: S-W                  | 0.01     | 0.01 | [−0.01, 0.02] | 0.99      | 0.32  | 861    |
| speakertype: USmo-USbi × mode: S-W                   | 0.00     | 0.01 | [−0.02, 0.02] | −0.02     | 0.99  | 861    |
| formality: F-I × mode: S-W                           | 0.04     | 0.01 | [ 0.02, 0.05] | 5.54      | <0.01 | 861    |
| speakertype: USbiGer-US × formality: F-I × mode: S-W | 0.02     | 0.01 | [ 0.00, 0.04] | 1.68      | 0.09  | 861    |
| speakertype: USmo-USbi × formality: F-I × mode: S-W  | −0.03    | 0.02 | [−0.06, 0.01] | −1.46     | 0.15  | 861    |
| Random Effects                                       |          |      |               |           |       |        |
| $\sigma^2$   | .00      |      |               |           |       |        |
| $\tau_{00\text{ID}}$                                 | .00      |      |               |           |       |        |
| ICC  | .37      |      |               |           |       |        |
| N <sub>ID</sub>                                      | 287      |      |               |           |       |        |
| Observations   | 1148     |      |               |           |       |        |
| Marginal R <sup>2</sup>                              | .179     |      |               |           |       |        |
| Conditional R <sup>2</sup>                           | .479     |      |               |           |       |        |
| AIC  | -3404.25 |      |               |           |       |        |

<sup>a</sup>The US group includes all monolingual and majority English speakers, USmo and USbi respectively, minus the heritage German speaker group with English as the majority language.

Table 6: Model on the MATTR measurements in the USbiGer data

| Predictors                 | $\beta$ | SE   | CI             | Statistic | p     | df     |
|----------------------------|---------|------|----------------|-----------|-------|--------|
| (Intercept)                | 0.65    | 0.01 | [ 0.64, 0.67]  | 97.13     | <0.01 | 36.06  |
| lang: ENG-GER              | -0.06   | 0.01 | [-0.07, -0.05] | -9.96     | <0.01 | 248.00 |
| formality: F-I             | 0.03    | 0.01 | [ 0.02, 0.04]  | 4.75      | <0.01 | 244.26 |
| mode: S-W                  | 0.05    | 0.01 | [ 0.03, 0.06]  | 7.54      | <0.01 | 244.26 |
| session: 2-1               | 0.00    | 0.01 | [-0.02, 0.01]  | -0.40     | 0.69  | 244.26 |
| session: 3-2               | 0.01    | 0.01 | [-0.01, 0.03]  | 1.15      | 0.25  | 244.26 |
| session: 4-3               | 0.00    | 0.01 | [-0.01, 0.02]  | 0.49      | 0.62  | 244.26 |
| language order: H-M        | -0.02   | 0.01 | [-0.04, 0.01]  | -1.29     | 0.20  | 36.06  |
| <hr/>                      |         |      |                |           |       |        |
| Random Effects             |         |      |                |           |       |        |
| $\sigma^2$                 | .00     |      |                |           |       |        |
| $\tau_{00\text{ID}}$       | .00     |      |                |           |       |        |
| ICC                        | .32     |      |                |           |       |        |
| $N_{\text{ID}}$            | 36      |      |                |           |       |        |
| Observations               | 280     |      |                |           |       |        |
| Marginal R <sup>2</sup>    | .315    |      |                |           |       |        |
| Conditional R <sup>2</sup> | .535    |      |                |           |       |        |
| AIC                        | -725.44 |      |                |           |       |        |

Table 7: Percentage of all shared lemmas across speaker groups in the German sub-corpus

|           | DEmo  | USbiGer | DEbiGreek | DEbiRuss | DEbiTurk |
|-----------|-------|---------|-----------|----------|----------|
| DEmo      | 41.02 | 17.24   | 23.55     | 24.88    | 25.00    |
| USbiGer   |       | 35.29   | 20.32     | 18.25    | 18.85    |
| DEbiGreek |       |         | 42.38     | 24.46    | 25.35    |
| DEbiRuss  |       |         |           | 40.81    | 25.65    |
| DEbiTurk  |       |         |           |          | 44.41    |

Table 8: Percentage of shared adjective lemmas across speaker groups in the German sub-corpus

|           | DEmo  | USbiGer | DEbiGreek | DEbiRuss | DEbiTurk |
|-----------|-------|---------|-----------|----------|----------|
| DEmo      | 35.98 | 15.82   | 24.42     | 28.37    | 29.40    |
| USbiGer   |       | 36.73   | 18.60     | 14.93    | 19.16    |
| DEbiGreek |       |         | 38.94     | 26.27    | 27.20    |
| DEbiRuss  |       |         |           | 35.99    | 27.62    |
| DEbiTurk  |       |         |           |          | 38.29    |

Table 9: Percentage of shared noun lemmas across speaker groups in the German sub-corpus

|           | DEmo  | USbiGer | DEbiGreek | DEbiRuss | DEbiTurk |
|-----------|-------|---------|-----------|----------|----------|
| DEmo      | 37.61 | 20.24   | 28.45     | 32.67    | 33.07    |
| USbiGer   |       | 34.84   | 26.71     | 21.65    | 22.37    |
| DEbiGreek |       |         | 43.71     | 30.32    | 32.71    |
| DEbiRuss  |       |         |           | 38.04    | 33.64    |
| DEbiTurk  |       |         |           |          | 43.99    |

Table 10: percentage of shared verb lemmas across speaker groups in the German sub-corpus

|           | DEmo  | USbiGer | DEbiGreek | DEbiRuss | DEbiTurk |
|-----------|-------|---------|-----------|----------|----------|
| DEmo      | 48.88 | 24.72   | 38.46     | 37.65    | 36.54    |
| USbiGer   |       | 35.15   | 27.95     | 28.16    | 26.22    |
| DEbiGreek |       |         | 43.14     | 39.55    | 39.75    |
| DEbiRuss  |       |         |           | 48.03    | 40.92    |
| DEbiTurk  |       |         |           |          | 48.62    |

Table 11: Percentage of all shared lemmas across speaker groups in the English sub-corpus

|           | USmo  | USbiGer | USbiGreek | USbiRuss | USbiTurk |
|-----------|-------|---------|-----------|----------|----------|
| USmo      | 51.95 | 28.62   | 28.75     | 29.09    | 29.75    |
| USbiGer   |       | 52.15   | 28.94     | 27.32    | 28.62    |
| USbiGreek |       |         | 50.36     | 29.61    | 30.76    |
| USbiRuss  |       |         |           | 51.18    | 29.75    |
| USbiTurk  |       |         |           |          | 53.66    |

Table 12: Percentage of shared adjective lemmas across speaker groups in the English sub-corpus

|           | USmo  | USbiGer | USbiGreek | USbiRuss | USbiTurk |
|-----------|-------|---------|-----------|----------|----------|
| USmo      | 42.58 | 30.00   | 31.38     | 31.84    | 28.94    |
| USbiGer   |       | 44.57   | 32.11     | 27.31    | 30.43    |
| USbiGreek |       |         | 35.85     | 31.85    | 36.44    |
| USbiRuss  |       |         |           | 40.10    | 29.70    |
| USbiTurk  |       |         |           |          | 43.24    |

Table 13: Percentage of shared noun lemmas across speaker groups in the English sub-corpus

|           | USmo  | USbiGer | USbiGreek | USbiRuss | USbiTurk |
|-----------|-------|---------|-----------|----------|----------|
| USmo      | 52.11 | 39.23   | 39.15     | 39.96    | 41.10    |
| USbiGer   |       | 48.64   | 38.38     | 36.65    | 39.39    |
| USbiGreek |       |         | 52.55     | 42.82    | 42.89    |
| USbiRuss  |       |         |           | 51.90    | 44.65    |
| USbiTurk  |       |         |           |          | 52.69    |

Table 14: Percentage of shared verb lemmas across speaker groups in the English sub-corpus

|           | USmo  | USbiGer | USbiGreek | USbiRuss | USbiTurk |
|-----------|-------|---------|-----------|----------|----------|
| USmo      | 56.99 | 46.93   | 47.44     | 48.48    | 52.50    |
| USbiGer   |       | 59.50   | 48.57     | 45.40    | 46.42    |
| USbiGreek |       |         | 56.72     | 48.32    | 51.12    |
| USbiRuss  |       |         |           | 57.52    | 48.45    |
| USbiTurk  |       |         |           |          | 60.37    |

## References

- Alamillo, Rosalva. 2019. Lexical skills in the formal and informal writing of students of Spanish as a second language and as a heritage language: A comparative study. *Íkala, Revista de Lenguaje y Cultura* 24(3). 503–519. DOI: [10.17533/udea.ikala.v24n03a12](https://doi.org/10.17533/udea.ikala.v24n03a12).
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- Belz, Malte. 2021. *Die Phonetik von äh und ähm*. Berlin: Metzler. DOI: [10.1007/978-3-662-62812-6](https://doi.org/10.1007/978-3-662-62812-6).
- Belz, Malte. 2023. Defining filler particles: A phonetic account of the terminology, form, and grammatical classification of “filled pauses”. *Languages* 8(1). DOI: [10.3390/languages8010057](https://doi.org/10.3390/languages8010057).
- Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511814358](https://doi.org/10.1017/CBO9780511814358).
- Boas, Hans C. 2003. *A constructional approach to resultatives*. Stanford: CSLI Publications.
- Boas, Hans C. 2010. On the equivalence and multifunctionality of discourse markers in language contact situations. In Theo Harden & Elke Hentschel (eds.), *Partikelforschung*, 301–315. Tübingen: Stauffenburg.
- Boas, Hans Christian. 2009. *The life and death of Texas German*. Durham, NC: Duke University Press.
- Bonvin, Audrey, Jan Vanhove, Raphael Berthele & Amelia Lambelet. 2018. Die Entwicklung von produktiven lexikalischen Kompetenzen bei Schüler(innen) mit portugiesischem Migrationshintergrund in der Schweiz. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 23. 135–148.

- Böttcher, Marlene & Margaret Zellers. 2023. Hesitating with and without language heritage: Prosodic aspects of filler particles in the RUEG corpus. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, 2795–2799. Prague: The International Phonetic Association.
- Bryant, Doreen. 2018. Implikationen des Erst- und Zweitspracherwerbs: Postpositional ins präpositionale System – ein didaktisches Angebot. In Björn Rothstein (ed.), *Sprachvergleich in der Schule*, vol. 1, 157–187. Baltmannsweiler: Schneider Verlag Hohengehren.
- Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/s0010-0277(02)00017-3.
- Clyne, Michael C. 2003. *Dynamics of language contact: English and immigrant languages* (Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511606526.
- Clyne, Michael G. 1967. *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia*. The Hague: Nijhoff.
- Covington, Michael A. & Joe D. McFall. 2010. Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics* 17(2). 94–100. DOI: 10.1080/09296171003643098.
- Czapka, Sophia, Nathalie Topaj & Natalia Gagarina. 2021. A four-year longitudinal comparative study on the lexicon development of Russian and Turkish heritage speakers in Germany. *Languages* 6(27) (1). DOI: 10.3390/languages6010027.
- Daller, Helmut, Roeland van Hout & Jeanine Treffers-Daller. 2003. Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics* 24(2). 197–222. DOI: 10.1093/applin/24.2.197.
- Degani, Tamar, Anat Prior & Natasha Tokowicz. 2011. Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology* 23 (1). 18–28. DOI: 10.1080/20445911.2011.445986.
- Egbert, Jesse & Douglas Biber. 2018. Do all roads lead to Rome? Modeling register variation with factor analysis and discriminant analysis. *Corpus Linguistics and Linguistic Theory* 14(2). 233–273.
- Elabdali, Rima, Shira Wein & Lourdes Ortega. 2022. Can adult lexical diversity be measured bilingually? A proof-of-concept study. In David M. Palfreyman & Nizar Habash (eds.), *Bilingual writers and corpus analysis* (Routledge Studies

- in Applied Linguistics), 121–156. New York: Taylor & Francis. DOI: 10.4324/9781003183921.
- Engelberg, Stefan, Anke Holler & Kristel Proost (eds.). 2011. *Sprachliches Wissen zwischen Lexikon und Grammatik*. Berlin: De Gruyter. DOI: 10.1515/978310262339.
- Fairclough, Marta. 2010. Lexical availability in bilingual contexts: A preliminary study of Spanish lexicon in Houston. In *Paper presented at 5th International Conference of Hispanic Linguistics*.
- Felfe, Marc. 2012. *Das System der Partikelverben mit „an“*. Berlin: De Gruyter Mouton.
- Fridman, Clara & Natalia Meir. 2023. Lexical production and innovation in child and adult Russian heritage speakers dominant in English and Hebrew. *Bilingualism: Language and Cognition* 26(5). 880–895. DOI: 10.1017/s1366728923000147.
- Fromkin, Victoria (ed.). 1973. *Speech errors as linguistic evidence* (Iauua linguarum. Series maior 77). Berlin: Mouton.
- Fuller, Janet M. 2001. The principle of pragmatic detachability in borrowing: English-origin discourse markers in Pennsylvania German. *Linguistics* 39(2). 351–369. DOI: 10.1515/ling.2001.014.
- Gharibi, Khadijeh & Frank Boers. 2019. Influential factors in lexical richness of young heritage speakers' family language: Iranians in New Zealand. *International Journal of Bilingualism* 23(2). 381–399. DOI: 10.1177/1367006917728395.
- Goldberg, Adele E. 2005. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199268511.001.0001.
- Härtl, Holden & James Witt. 1998. Lokale Konzepte und Partikelverben in einem Modell der Sprachproduktion. *Zeitschrift für Sprachwissenschaft* 17(1). 3–34.
- Hieke, Adolf E., Sabine Kowal & Daniel C. O'Connell. 1983. The trouble with "articulatory" pauses. *Language and Speech* 26(3). 203–214. DOI: 10.1177/002383098302600302.
- Hopp, Holger & Michael T. Putnam. 2015. Syntactic restructuring in heritage grammars: Word order variation in Moundridge Schweitzer German. *Linguistic Approaches to Bilingualism* 5(2). 180–214. DOI: 10.1075/lab.5.2.02hop.
- Hržica, Gordana & Maja Roch. 2021. Lexical diversity in bilingual speakers of Croatian and Italian. In Sharon Armon-Lotem & Kleanthes K. Grohmann (eds.), *Language impairment in multilingual settings: LITMUS in action across Europe* (Trends in Language Acquisition Research 29), 99–130. John Benjamins. DOI: 10.1075/tilar.29.04hrz.

- Jackendoff, Ray & Jenny Audring. 2019. *The texture of the lexicon: Relational morphology and the parallel architecture*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198827900.001.0001.
- Jarvis, Scott. 2013. Capturing the diversity in lexical diversity. *Language Learning* 63(s1). 87–106. DOI: 10.1111/j.1467-9922.2012.00739.x.
- Kassambara, Alboukadel. 2023. *ggpubr: “ggplot2” based publication ready plots*. R package version 0.6.0. <https://CRAN.R-project.org/package=ggpubr>.
- Keller, Mareike. 2014. *Phraseme im bilingualen Diskurs: „All of a sudden geht mir ein Licht auf.“*. Frankfurt a.M.: Peter Lang Edition. DOI: 10.3726/978-3-653-04258-0.
- Klotz, Martin, Rahel Gajaneh Hartz, Annika Labrenz, Anke Lüdeling & Anna Shadrova. 2024. Die RUEG-Korpora: Ein Blick auf Design, Aufbau, Infrastruktur und Nachnutzung multilingualer Forschungsdaten. *Zeitschrift für germanistische Linguistik* 52(3). 578–592. DOI: 10.1515/zgl-2024-2026.
- Kopotev, Mikhail, Olesya Kisseelev & Maria Polinsky. 2020. Collocations and near-native competence: Lexical strategies of heritage speakers of Russian. *International Journal of Bilingualism* 28(6). 1135–1162. DOI: 10.1177/1367006920921594.
- Lambelet, Amelia. 2021. Lexical diversity development in newly arrived parent-child immigrant pairs: Aptitude, age, exposure, and anxiety. *Annual Review of Applied Linguistics* 41. 76–94. DOI: 10.1017/s0267190521000039.
- Laufer, Batia & Paul Nation. 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16(3). 307–322. DOI: 10.1093/applin/16.3.307.
- Lenth, Russell V. 2025. *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.11.0. <https://CRAN.R-project.org/package=emmeans>.
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation* (ACL-MIT Press Series in Natural Language Processing). Cambridge, MA/London: MIT Press. DOI: 10.7551/mitpress/6393.001.0001.
- Lüdecke, Daniel. 2024. *sjPlot: Data visualization for statistics in social science*. R package version 2.8.17. <https://CRAN.R-project.org/package=sjPlot>.
- Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Philip Waggoner & Dominique Makowski. 2021. performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software* 6(60). 3139. DOI: 10.21105/joss.03139.
- Lüdeling, Anke. 2001. *On particle verbs and similar constructions in German* (Dissertations in linguistics). Stanford: CSLI Publications.

- Lüdeling, Anke, Artemis Alexiadou, Aria Adli, Karin Donhauser, Malte Dreyer, Markus Egg, Anna Helene Feulner, Natalia Gagarina, Wolfgang Hock, Stefanie Jannedy, Frank Kammerzell, Pia Knoeferle, Thomas Krause, Manfred Krifka, Silvia Kutscher, Beate Lütke, Thomas McFadden, Roland Meyer, Christine Mooshammer, Stefan Müller, Katja Maquate, Muriel Norde, Uli Sauerland, Stephanie Solt, Luka Szucsich, Elisabeth Verhoeven, Richard Waltereit, Anne Wolfsgruber & Lars Erik Zeige. 2022. Register: Language users' knowledge of situational-functional variation. *Register Aspects of Language in Situation* 1(1). 1–58. DOI: 10.18452/24901.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehayé, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Lüdeling, Anke, Hagen Hirschmann & Anna Shadrova. 2017. Linguistic models, acquisition theories, and learner corpora: Morphological productivity in SLA research exemplified by complex verbs in German. *Language Learning* 67(S1). 96–129.
- Malvern, David & Brian Richards. 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19(1). 85–104. DOI: 10.1191/0265532202lt221oa.
- Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36(2). 281–332. DOI: 10.1515/ling.1998.36.2.281.
- Matthiessen, Christian M. I. M. 2019. Register in systemic functional linguistics. *Register Studies* 1(1). 10–41. DOI: 10.1075/rs.18010.mat.
- McCarthy, Philip M. & Scott Jarvis. 2010. MTLD, vocD-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42(2). 381–392. DOI: 10.3758/BRM.42.2.381.
- Montrul, Silvina. 2006. On the bilingual competence of Spanish heritage speakers: Syntax, lexical-semantics and processing. *International Journal of Bilingualism* 10(1). 37–70. DOI: 10.1177/13670069060100010301.
- Moreno-Fernandez, Francisco. 2007. Anglicismos en el léxico disponible de los adolescentes hispanos de Chicago. In Kim Potow & Richard Cameron (eds.), *Spanish in contact: Policy, social and linguistic inquiries*, 41–58. John Benjamins Publishing Company. DOI: 10.1075/impact.22.05mor.

- Müller, Stefan. 2002. *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German* (Studies in Constraint-Based Lexicalism 13). Stanford: CSLI Publications.
- Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16(4). 709–730. DOI: 10.1017/S1366728912000727.
- Nation, Paul & Laurence Anthony. 2016. Measuring vocabulary size. In Eli Hinkel (ed.), *Handbook of research in second language teaching and learning*, vol. 3, 355–368. New York: Routledge. DOI: 10.4324/9781315716893.
- Pashkova, Tatiana, Abigail Hodge & Shanley E. M. Allen. 2020. *Explicitness in heritage speakers' majority English productions*. Paper presented at DGfS-Jahrestagung. Hamburg.
- Pashkova, Tatiana, Wintai Tsehayé, Shanley E. M. Allen & Rosemarie Tracy. 2022. Syntactic optionality in heritage language use: Clause type preferences of German heritage speakers in a majority English context. *Heritage Language Journal* 19(1). 1–41. DOI: 10.1163/15507076-12340022.
- Pennock-Speck, Barry & Begoña Clavel-Arroitia. 2021. Analysing lexical density, diversity, and sophistication in written and spoken telecollaborative exchanges. *CALL-EJ* 22 (3). 230–250.
- Petersen, Jan Heegård, Gert Foget Hansen, Jacob Thøgersen & Karoline Kühl. 2021. Linguistic proficiency: A quantitative approach to immigrant and heritage speakers of Danish. *Corpus Linguistics and Linguistic Theory* 17 (2). 465–490. DOI: 10.1515/cllt-2017-0088.
- Polinsky, Maria. 2005. Word class distinctions in an incomplete grammar. In Dorit Diskin Ravid & Hava Bat-Zeev Shyldkrot (eds.), *Perspectives on language and language development: Essays in honor of Ruth A. Berman*, 419–434. Boston: Springer US. DOI: 10.1007/1-4020-7911-7\_30.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Prior, Anat, Tamar Degani, Sehrab Awawdy, Rana Yassin & Nachshon Korem. 2017. Is susceptibility to cross-language interference domain specific? *Cognition* 165. 10–25. DOI: 10.1016/j.cognition.2017.04.006.
- Pustejovsky, James & Olga Batiukova. 2019. *The lexicon* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/9780511982378.

- Putnam, Michael T. & Joseph C. Salmons. 2013. Losing their (passive) voice: Syntactic neutralization in heritage German. *Linguistic Approaches to Bilingualism* 3(2). 233–252. DOI: 10.1075/lab.3.2.05put.
- R Core Team. 2021. *R: A Language and environment for statistical computing*. Version 4.3.0. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rabe, Maximilian M., Shravan Vasishth, Sven Hohenstein, Reinhold Kliegl & Daniel J. Schad. 2020. Hypr: An R package for hypothesis-driven contrast coding. *The Journal of Open Source Software* 5(48). 2134. DOI: 10.21105/joss.02134.
- Rabinovich, Ella, Yulia Tsvetkov & Shuly Wintner. 2018. Native language cognate effects on second language lexical choice. *Transactions of the Association for Computational Linguistics* 6. 329–342. DOI: 10.1162/tacl\_a\_00024.
- Rakhilina, Ekaterina, Anastasia Vyrenkova & Maria Polinsky. 2016. Linguistic creativity in heritage speakers. *Glossa: A journal of general linguistics* 1(1). DOI: 10.5334/gjgl.90.
- Riehl, Claudia M. 2014. *Mehrsprachigkeit: Eine Einführung* (Einführung Germanistik). Darmstadt: WBG.
- Schad, Daniel J., Shravan Vasishth, Sven Hohenstein & Reinhold Kliegl. 2019. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language* 110. DOI: 10.1016/j.jml.2019.104038.
- Schmid, Monika. 2011. *Language attrition*. Cambridge: Cambridge University Press.
- Schulz, Petra & Rosemarie Tracy. 2011. *Linguistische Sprachstandserhebung: Deutsch als Zweitsprache (LiSe-DaZ): Language test for children with German as a second language*. Göttingen: Hogrefe.
- Shadrova, Anna. 2025. No three productions alike: Lexical variability, situated dynamics, and path dependence in task-based corpora. *Open Linguistics* 11(1). 20240036. DOI: 10.1515/opli-2024-0036.
- Shadrova, Anna, Pia Linscheid, Julia Lukassek, Anke Lüdeling & Sarah Schneider. 2021. A challenge for contrastive L1/L2 corpus studies: Large inter- and intra-individual variation across morphological, but not global syntactic categories in task-based corpus data of a homogeneous L1 German group. *Frontiers in Psychology* 12. DOI: 10.3389/fpsyg.2021.716485.
- Simon-Cereijido, Gabriela & Vera F. Gutiérrez-Clellen. 2009. A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics* 30(2). 315–337. DOI: 10.1017/s0142716409090134.

- Slobin, Dan. 2003. Language and thought online: Cognitive consequences of linguistic relativity. In Dedre Gentner & Susan Goldin-Meadow (eds.), *Language in mind: Advances in the investigation of language and thought*, 157–191. Cambridge: MIT Press.
- Stiebels, Barbara & Dieter Wunderlich. 1994. Morphology feeds syntax: The case of particle verbs. *Linguistics* 32. 913–968.
- Stolberg, Doris. 2015. *Changes between the lines: Diachronic contact phenomena in written Pennsylvania German*. Berlin/New York: De Gruyter Mouton.
- Talmy, Leonard. 1972. The basis for a crosslinguistic typology of motion/location, part I. *Working Papers on Language Universals* 9. 41–116.
- Talmy, Leonard. 1988. Force dynamics in language and cognition. *Cognitive Science*. 49–100.
- Tenny, Carol & James Pustejovsky (eds.). 2000. *Events as grammatical objects: The converging perspectives of lexical semantics and syntax*. Stanford: CSLI Publications.
- Tomasello, Michael. 2006. Konstruktionsgrammatik und früher Erstspracherwerb. In Kerstin Fischer & Anatol Stefanowitsch (eds.), *Konstruktionsgrammatik: Von der Anwendung zur Theorie* (Stauffenburg Linguistik 1), 19–37. Tübingen: Stauffenburg-Verlag.
- Tracy, Rosemarie. 1991. *Sprachliche Strukturentwicklung: Linguistische und kognitionspsychologische Aspekte einer Theorie des Erstspracherwerbs*. Tübingen: Narr.
- Tracy, Rosemarie. 2011. Konstruktion, Dekonstruktion und Rekonstruktion: Minimalistische und (trotzdem) konstruktivistische Überlegungen zum Spracherwerb. In Stefan Engelberg, Anke Holler & Kristel Proost (eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik*, 397–428. Berlin: De Gruyter. DOI: 10.1515/9783110262339.397.
- Tracy, Rosemarie. 2022. Gemischtsprachiges Sprechen: Formen, Funktionen, Dynamik. In Csaba Földes & Thorsten Roelcke (eds.), *Handbuch Mehrsprachigkeit*, 399–428. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110623444-018.
- Tracy, Rosemarie & Dafydd Gibbon. 2023. The beat goes on: A case study of timing in heritage German prosody. In Michael Beißwenger, Eva Gredel, Lothar Lemnitzer & Roman Schneider (eds.), *Korpusgestützte Sprachanalyse: Linguistische Grundlagen, Anwendungen und Analysen* (Studien zur Deutschen Sprache 88), 261–283. Tübingen: Narr.
- Tracy, Rosemarie & Elsa Lattey. 2001. Language contact in the individual: A case study based on letters from a German immigrant in New Jersey. In Per Sture Ureland (ed.), *Global Eurolinguistics: European languages in North America – migration, maintenance and death*, 413–433. Tübingen: Niemeyer.

- Treffers-Daller, Jeanine. 2019. What defines language dominance in bilinguals? *Annual Review of Linguistics* 5(1). 375–393. DOI: 10.1146/annurev-linguistics-011817-045554.
- Treffers-Daller, Jeanine & Tomasz Korybski. 2016. Using lexical diversity measures to operationalize language dominance in bilinguals. In Jeanine Treffers-Daller & Carmen Silva-Corvalan (eds.), *Language dominance in bilinguals: Issues of measurement and operationalization*, 106–133. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9781107375345.006.
- Tsehaye, Wintai, Tatiana Pashkova, Rosemarie Tracy & Shanley E. M. Allen. 2021. Deconstructing the native speaker: Further evidence from heritage speakers for why this horse should be dead! *Frontiers in Political Science* 12. 1–14. DOI: 10.3389/fpsyg.2021.717352.
- Tsehaye, Wintai, Rosemarie Tracy & Johanna Tausch. 2025. Inter- and intra-individual variation: How it materializes in Heritage German and why it matters. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 141–177. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775165.
- Van Gijsel, Sofie, Dirk Speelman & Dirk Geeraerts. 2005. A variationist, corpus linguistic analysis of lexical richness. In *Proceedings of Corpus Linguistics 2005*.
- Van Hout, Angeliek. 2000. Event semantics in the lexicon-syntax interface: Verb frame alternations in Dutch and their acquisition. In Carol Tenny & James Pustejowsky (eds.), *Events as grammatical objects: The converging perspectives of lexical semantics and syntax*, 239–282. Stanford: CSLI Publications.
- Venables, William N. & Brian D. Ripley. 2002. *Modern applied statistics with S*. 4th edn. New York: Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>.
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2). 143–160. DOI: 10.2307/2182371.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo & Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). DOI: 10.21105/joss.01686.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.

- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagaria, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 33–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775159.
- Yu, Guoxing. 2009. Lexical diversity in writing and speaking task performances. *Applied Linguistics* 31(2). 236–259. DOI: 10.1093/applin/amp024.
- Zeller, Jochen. 2001. *Particle verbs and local domains* (Linguistics Today 41). Amsterdam: John Benjamins. DOI: 10.1075/la.41.
- Zenker, Fred & Kristopher Kyle. 2021. Investigating minimum text lengths for lexical diversity indices. *Assessing Writing* 47. DOI: 10.1016/j.asw.2020.100505.
- Zhu, Hao. 2024. *kableExtra: Construct complex table with “kable” and pipe syntax*. R package version 1.4.0. <https://CRAN.R-project.org/package=kableExtra>.
- Zimmer, Christian, Heike Wiese, Horst J. Simon, Marianne Zappen-Thomson, Yannic Bracke, Britta Stuhl & Thomas Schmidt. 2020. Das Korpus Deutsch in Namibia (DNam): Eine Ressource für die Kontakt-, Variations- und Soziolinguistik. *Deutsche Sprache* 48(3). 210–232. <https://nbn-resolving.org/urn:nbn:de:bsz:mh39-100570> (27 April, 2023).



## **Part III**

# **Dynamics of pragmatic structure**



# Chapter 11

## Information packaging and word order dynamics in language contact

Oliver Bunk<sup>a</sup>, Shanley E. M. Allen<sup>b</sup>, Sabine Zerbian<sup>c</sup>, Tatiana Pashkova<sup>b</sup>, Yulia Zuban<sup>c</sup> & Erica Conti<sup>a</sup>

<sup>a</sup>Humboldt-Universität zu Berlin <sup>b</sup>University of Kaiserslautern-Landau

<sup>c</sup>University of Stuttgart

Word order is one of the linguistic resources speakers use to express specific meanings and present information. The literature considers information structure a major driving force behind word order and word order variation. The chapter argues that this is particularly true in language contact settings, enhancing the emergence of new word order patterns. We review studies that were conducted in the RUEG group, covering a) language-specific phenomena (referent introduction in English, V3 clauses and the placement of modal particles in German, and object-verb/verb-object patterns in Russian) and b) a cross-linguistically available construction – left dislocation in English, German, and Russian.

### 1 Introduction

In a world of increasingly dynamic global migration and communication, language contact has become an important and pervasive phenomenon in linguistic research. The interaction of different languages can result in the emergence of new linguistic forms and the shaping of language use and communication. One area of particular interest in studying language contact is the relationship between information packaging and word order. Information packaging involves two key concepts: information structure and information status. Information structure refers to how speakers encode and present information in a sentence



and includes notions such as focus and topic (see Krifka & Musan 2012). Information status means the status of referring expressions as referentially and/or lexically given or new in discourse (see Baumann 2006). Both concepts affect the packaging of information in spoken and written texts in order to respond to the immediate communicative needs of the interlocutor (see Chafe 1976).

The impact of language contact on information packaging and word order is a topic of ongoing research and debate. Information structure and status have been identified as one of the main causes of word order variation in languages with relatively free word order, such as German (Musan 2002), Russian (Jasinskaja 2016), and Turkish (Özsoy 2019), as well as languages with fixed word order, such as English (Ward & Birner 2004). Across different languages, word order has been shown to be influenced by specific information packaging principles including the given-before-new principle, the end-focus principle, the end-weight principle, and the complexity principle (see Hilpert 2021).

For example, according to the given-before-new principle (Halliday 1967, Hilpert 2021), discourse-given referents tend to appear before discourse new referents across different languages, such as English and Russian (Haviland & Clark 1974, Bock & Irwin 1980, Arnold et al. 2000, Slioussar 2007, 2011). In Russian, this pattern may lead to word orders with inversion ((XP) V S) or dislocation (XP S V), differing from the default subject-verb-object (SVO) pattern (e.g., King 1995, Kallestinova 2007, Bailyn 2012). In German, constituents can scramble within the clause to change their information status (e.g., Musan 2002). English uses specific syntactic constructions for highlighting and focusing. For example, locative inversion and existential *there* structures highlight the introduction of new information postverbally (Ward & Birner 2004).

The differences in word order flexibility and the different strategies in conveying information syntactically are particularly interesting from a cross-linguistic perspective, and language contact settings can be expected to enhance our understanding of the dynamic relationship between word order and information packaging. Various studies point to different factors influencing word order variation in bilinguals. For particular word order patterns, some studies suggest cross-linguistic influence as the explanatory factor (Polinsky 2018). However, for other phenomena, there is no evidence of such cross-linguistic influence. Rather, bilingual speakers seem more dynamic in language use than monolingual speakers (Wiese et al. 2022). This claim is based on the idea that exposure to multiple languages and linguistic systems allows bilinguals to draw on a wider range of linguistic resources, resulting in a more flexible and adaptive approach to language use (Wiese & Rehbein 2016). In particular, bilinguals may be more likely to adapt information packaging and word order in response to language contact,

as they can switch between different linguistic systems and draw on a range of linguistic resources to encode information, leading to either new emerging word order patterns or more frequent use of noncanonical patterns. In this chapter, we delve into the relation between information packaging, encompassing information structure and status, and word order, which seems to be dynamic in language contact situations and bilingual speakers.

This chapter presents studies within the RUEG group<sup>1</sup> investigating word order phenomena related to information packaging in English, Russian, and German. We consider different contact settings, namely 1) settings in which these languages are used as majority languages by monolinguals and by speakers of different heritage languages (i.e., Germany, the US, and Russia) or 2) settings in which two of these languages, German and Russian, are used as heritage languages (i.e., Germany and the US). We start by discussing the empirical basis of the analyses and findings for specific phenomena in English (referent introduction), German (V3 and modal particles), and Russian (word order in main/embedded clauses and OV/VO patterns). Then, we turn to the dynamics of information packaging from a cross-linguistic perspective, discussing a noncanonical word order pattern that occurs in all three languages, namely left dislocation constructions. To our knowledge, no study has systematically looked at the use and structure of left dislocation in different languages across different speaker groups (i.e., monolingual and bilingual) and different communicative settings. Thus, our study is one of the first to attempt such an endeavor and provides a broad perspective on noncanonical variation at the interface of word order and information packaging. Finally, we summarize our results and discuss implications for further research.

## 2 Word order variation in English, German, and Russian

English, German, and Russian exhibit various word order patterns related to information packaging. The languages differ concerning word order flexibility, making these language contact settings particularly interesting. As a language with flexible word order, Russian has a default SVO pattern, but all other alternations are possible, fulfilling information structural needs (e.g., Sirotinina 2003, Slioussar 2007, 2011). German is an SOV language with a V2 constraint, leading

---

<sup>1</sup>The Research Unit *Emerging Grammars in Language Contact Situations: A Comparative Approach* (FOR 2537, 2018–2024, <https://hu.berlin/RUEG>), funded by the DFG, investigated linguistic dynamics in monolingual and multilingual speakers' repertoires. The present article presents results from the subproject P8 (PIs: Shanley E. M. Allen, Oliver Bunk, Sabine Zerbian), focussing on information structural dynamics in contact situations.

to finite V2 in declarative main clauses and finite verb last in subordinate clauses. English is strict SVO with residual V2. In the following sections, we present findings pronounced in bilingual speakers, indicating that this group is particularly prone to word order variation triggered by information packaging.

## 2.1 Data basis

The empirical basis for the analyses is the RUEG corpus<sup>2</sup> (Lüdeling et al. 2024). The data were elicited using one common experimental set-up (Wiese 2020), which allows for a systematic comparison of language in different communicative situations. Participants were asked to picture themselves as a witness to a car accident that they had just watched in a short video clip. They were tasked to describe the accident in four communicative situations, differing in mode (spoken vs. written) and formality (formal vs. informal). Table 1 illustrates these four contexts.

Table 1: Communicative situations simulated in the Lang-Sit set-up  
(Wiese 2020)

|         | Formal                        | Informal                            |
|---------|-------------------------------|-------------------------------------|
| Spoken  | voice recording to the police | WhatsApp© audio message to a friend |
| Written | witness report to the police  | WhatsApp© text message to a friend  |

The corpus comprises data from adolescent and adult speakers from different countries and with different language biographies. Speakers were from Germany, Greece, Russia, Turkey, or the US. They were either monolingually or bilingually raised.<sup>3</sup> Monolinguals spoke the majority language of their respective countries, i.e., German, Greek, Russian, Turkish, or English. Bilingual speakers were tested in Germany and the US. In Germany, bilingual speakers were speakers of the majority language (German) and Greek, Russian, or Turkish as their heritage language. In the US, bilingual speakers were speakers of the majority

<sup>2</sup><https://hu.berlin/RUEG-corpus>

<sup>3</sup>We consider monolingually-raised speakers as speakers who grew up with one language as their family language. Bilingually-raised speakers are considered speakers who grew up with another family language in addition to the majority language. These speakers were either born in Germany, Greece, Russia, Turkey, or the US with the respective majority languages or migrated to these countries before the age of four years.

language (English) and German, Greek, Russian, or Turkish as their heritage language. Table 2 gives an overview of the subcorpora (RU-RUEG = Russian subcorpus, EN-RUEG = English subcorpus, DE-RUEG = German subcorpus; h=heritage, mo=monolingual), including the numbers of speakers per group.

Table 2: Numbers of speakers in the different subcorpora

| Speaker group    | Number of speakers |
|------------------|--------------------|
| RUEG-RU          |                    |
| mo-Russian (RUS) | 67                 |
| h-Russian (GER)  | 60                 |
| h-Russian (US)   | 69                 |
| RUEG-EN          |                    |
| mo-English (US)  | 64                 |
| h-German (US)    | 34                 |
| h-Greek (US)     | 65                 |
| h-Russian (US)   | 65                 |
| h-Turkish (US)   | 59                 |
| RUEG-DE          |                    |
| mo-German (GER)  | 64                 |
| h-German (US)    | 36                 |
| h-Greek (GER)    | 45                 |
| h-Russian (GER)  | 61                 |
| h-Turkish (GER)  | 65                 |

While the RUEG corpus was the basis for all of our investigations, we used additional corpora for German, to conduct comparative analyses across different language contact scenarios. We provide information on these corpora in the respective sections.

## 2.2 Referent introduction in English

Even though English word order is canonically SVO, it can be changed according to pragmatic constraints, resulting in specific information packaging constructions (Huddleston 2002). For example, the given-before-new principle (Biber et al. 2021, Hilpert 2021) discussed in the introduction leads to a conflict for new subjects in English: a new subject referent has to be placed first because of SVO

word order, but ideally, it would be placed closer to the end of the clause because the referent is new.

To resolve this tension, speakers can use several non-SVO constructions that put the original SVO subject referent after the finite verb, for instance, existential *there* (1a), presentational *there* (1b), locative inversion (1c), and passivization (1d). The referent might remain the syntactic subject (1c) or take a different syntactic role – a notional subject in (1a) and (1b) or an object of a preposition in (1d).

- (1) Original clause: A **woman** was unloading her groceries on the other side of the street.
  - a. There was a **woman** unloading her groceries on the other side of the street.
  - b. There stood a **woman** unloading her groceries on the other side of the street.
  - c. On the other side of the street was a **woman** unloading her groceries.
  - d. <I saw some groceries.> They were unloaded by a **woman**.

There is some evidence that bilingual speakers use information packaging constructions differently compared to monolingually-raised English speakers, possibly because of cross-linguistic influence. For example, speakers of Singapore and Jamaican English use fewer existential *there* constructions than speakers of monolingual varieties (Winkle 2015).

Our pilot findings in majority English showed differences and similarities between bilinguals and monolinguals. Unlike monolinguals, bilinguals used locative inversion for new subjects. Both groups used existential *there* to a similar extent.

Based on these previous findings, Pashkova et al. (submitted) focused on the syntactic structures that are used for the introduction of new subjects in majority English. We asked if English bilinguals use more non-SVO constructions for the introduction of new subjects as compared to English monolinguals. We hypothesized that English bilinguals with heritage Russian and heritage Turkish, languages with flexible word orders, would use more non-SVO structures in English, possibly because of the cross-linguistic influence from the heritage languages. To evaluate the hypothesis, we compared the syntactic structures used for new vs. given subjects produced by 82 English bilinguals (40 heritage Russian and 42 heritage Turkish) vs. 40 English monolinguals. Each subject was annotated for its information status (new vs. given) and for the syntactic structure it was used in. Subsequently, the syntactic structures were divided into SVO-type structures (SVO and copular clauses without inversion) and non-SVO-type

structures (existential and presentational *there*, locative and non-locative inversions, right and left dislocations, passives, questions, *it-* and pseudo-clefts). The results of the analysis indicated no difference between the frequency of use of non-SVO structures by English bilinguals and monolinguals. In contrast, there was a difference between new and given subjects in the frequency of non-SVO structures: new subjects were more likely to appear in such a structure than given subjects. There was no interaction between the information status of the subject and speaker group (English bilinguals vs. monolinguals). These results show that all speakers in our sample, regardless of their bilingualism, preferred to use non-SVO structures for new subjects more than for given subjects. The data indicate that new subjects appear more frequently in constructions referring to information status (i.e., non-SVO), and both bilinguals and monolinguals similarly follow this trend. Cross-linguistic influence, as suggested by other studies, does not seem to play a major role. Rather, dynamics in information packaging appear to influence word order patterns in both speaker groups.

### 2.3 V3 and modal particles in German

In contrast to English, German is a more flexible language, allowing for a range of word order patterns. German is generally considered an SOV language with a V2 constraint, placing the finite verb in the second position in declarative clauses. Various constituents, such as subjects, objects, or adverbials, can occupy the pre-verbal position. V2 is considered a rigid constraint in German by the vast majority of the literature, and deviations from V2 are often claimed to be ungrammatical (see 2):

- (2) \* Gestern Johann hat getanzt.  
           yesterday Johann has danced  
       ‘Yesterday, Johann danced.’ (Roberts & Roussou 2002: 137)

However, V3 patterns occur systematically in everyday language use, as (3) exemplifies:

- (3) danach er lässt den ball fallen  
       after.that he lets the ball fall  
       ‘after that, he drops the ball’ (RUEG corpus, DEBi58MT\_isD)

These instances show an adverbial-subject-finite verb linearization (but see Sluckin 2021 for other V3 orders, such as adverbial-adverbial-finite verb orders, sparsely occurring) and are not only reported for German (see Wiese 2013), but

also for other V2 languages in language contact settings. V3 occurs in urban contact dialects in Sweden (see Kotsinas 1992), Denmark (see Quist 2000), Norway (see Opsahl 2009), and the Netherlands (see Meelen et al. 2020). It is also documented in contexts where Germanic languages are spoken as minority languages, including German in Namibia and the US (see Tracy & Lattey 2010, Sewell 2015, Wiese & Müller 2018), heritage Norwegian (see Alexiadou & Lohndal 2018, Kinn & Larsson 2021, 2022), heritage Low German (see Rocker 2022), heritage Swedish (Kinn & Larsson 2021, 2022), heritage Danish (Kühl & Petersen 2018) and heritage Icelandic (Arnbjörnsdóttir et al. 2018). V3 is also reported for monolingual German speakers (see Schałowski 2017, Wiese & Müller 2018, Bunk 2020). Wiese & Müller (2018) find that multilingual speakers use more V3 structures than monolinguals, indicating that word order variation concerning the V2 constraint is particularly dynamic in this speaker group.

Wiese et al. (2022) looked at the distribution of V3 sentences across speaker groups in the RUEG corpus. Based on the corpus version 0.4.0 (Wiese et al. 2021), the study highlights that most V3 sentences were produced by bilingual speakers in Germany and the US. However, V3 also occurred in monolingual speakers. A closer look at the speakers in Germany revealed that bilingual speakers with Turkish as their heritage language produced the vast majority of the V3 sentences. Wiese et al. (2022) argue that these findings speak against contact-linguistic transfer because Greek and Russian tend to have SVO, which might more easily allow for V3, while Turkish is predominantly SOV, making transfer to V3 in German less likely compared to the other two languages. However, for heritage German in the US, cross-linguistic effects might be at play, as these speakers produced V3 that also involves non-subjects (see Wiese et al. 2022). In a follow-up study using the much larger database of the corpus version 1.0 (Lüdeling et al. 2024), we found a similar distribution of V3 (see Bunk & Rocker in preparation).

Information structure plays a crucial role in the emergence and use of German V3 patterns (see Wiese 2009, Schałowski 2017, Wiese & Müller 2018, Bunk 2020). The adverbial-subject-finite verb linearization is closely tied to a “frame-setter > (aboutness) topic > comment” order and this pattern appears to hold in non-verbal contexts, i.e., contexts in which participants are asked to retell a story using only toys and word cards (see Wiese 2020) as well as in second language speakers (Bunk & Gamper forthcoming). Wiese & Rehbein (2016) argue that bilinguals more frequently use V3 since they are exposed to more linguistic variation, leading to a less strict inventory of grammatical rules, which they apply more productively to form noncanonical patterns. However, V3 also occurs in monolingual speakers, and psycholinguistic evidence suggests that these speakers a)

process V3 as an integral part of German grammar and b) judge adverbial-subject-finite verb orders as more grammatical than adverbial-object-finite verb, arguing for a representation of the former but not the latter structure (see Bunk 2020). Taken together, these studies point to a cognitive preference for “frame-setter > (aboutness) topic > comment” orders as a way of information packaging.

From these previous findings, it seems that V3 spotlights the interaction of information structure and word order. Concurrently, it sheds light on the external interface of syntax and discourse (see Sorace 2011). The initial adverbial functions as a frame-setter at the information structural level, as a discourse marker at the discourse structuring level (see Schalowski 2017), or takes both functions simultaneously (see Bunk 2020). Bunk & Rocker (in preparation) investigate the distribution of V3 types (frame-setting V3 vs. discourse-connecting V3 vs. ambiguous cases) in the RUEG corpus, the Kiezdeutsch-Korpus (KiDKo, Wiese et al. 2010), the DNAm corpus (Zimmer et al. 2020), and a collection of V3 sentences from heritage Low-German in the US (Rocker 2022), focusing on spoken language. While RUEG comprises data from the US and Germany, KiDKo contains data from multilingual and monolingual adolescents from Germany. DNAm includes data from adult and adolescent speakers of the German-speaking community in Namibia. All corpora slightly differ in size and number of speakers, however normalized numbers allowed for a comparison of the data. The corpora permit investigating different contact scenarios with different statuses of German: 1) as a majority language (Germany), and 2) as a minority/heritage language (US and Namibia). While in the US, German is on the decline and influenced by majority English, the speech community in Namibia still uses German as a vital language. We were interested in potential differences in the impact of other (majority) languages due to these diverging statuses of German.

The data were annotated for the type of adverbial (frame-setter, discourse-connector, ambiguous) in all corpora. We found that frame-setting and discourse-connecting V3 structures were used at similar rates in the RUEG and DNAm corpus and that only a few cases were ambiguous. KiDKo exhibited only a few discourse-connecting V3 sentences, while most were framesetting or ambiguous. We argued that differences between KiDKo and RUEG/DNam are due to contextual factors, such as the oppositions between dialogue vs. monologue and free conversation vs. narration. However, we found that (*und*) *dann* ('(and) then') is the preferred adverbial in V3 across all corpora (see Wiese & Müller 2018 for a detailed discussion of *dann* ('then') in KiDKo, see Sewell 2015 for the prevalence of temporal adverbs licensing V3 in heritage German in Wisconsin German). We were also interested in the role of prosody in disambiguating frame-setting from discourse-connecting adverbials in V3, and found that prosodic boundaries might

provide a disambiguating cue for the interlocutor (Bunk & Rocker in preparation, Zerbian, Zuban, et al. 2025 [this volume] for a detailed summary).

Another case where information structure and discourse structure impact word order, particularly in the peripheries, is modal particles (henceforth MPs). MPs indicate “to the hearer the mood or attitude of a speaker” (Bross 2012: 183). Bunk et al. (2024) focus on the MPs *eben* and *halt*, investigating their distribution across speaker groups, function, and syntactic structure.

*Eben* and *halt* are often considered synonyms (e.g., Hentschel 1986, Diewald 2007), marking a proposition as definite and irrevocable, indicating irreversibility and resignation (Helbig 1988). However, others acknowledge subtle differences in meaning (see Thurmair 1989, Thielmann 2015, Blühdorn 2019). MPs are typically restricted to the so-called “middlefield”, the position between a finite and non-finite verb or verbal parts (Höhle 1986). Both finite and non-finite verbs and verbal particles form the “left sentence bracket” and “right sentence bracket” surrounding the middlefield. Table 3 provides an example of MPs in their canonical position in the middlefield.

Table 3: MPs in their canonical position in the middlefield

| Prefield                            | L. verb bracket | Middlefield                                 | R. verb br.                 | Postfield |
|-------------------------------------|-----------------|---|-----------------------------|-----------|
| Der Mann<br>the man                 | hat<br>has      | <b>halt/eben</b> den Ball<br>MP/MP the ball | in der Hand.<br>in his hand |           |
| ‘The man has the ball in his hand.’ |                 |   |                             |           |

In this study, we investigated *halt* and *eben* in the RUEG corpus, KiDKo, and the DNAm corpus, again, to tease apart the influence of the societal macro context and the status of German on the use of the two MPs. The corpora comprise data from Germany (RUEG, KiDKo), the US (RUEG), and Namibia (DNAm).

In line with previous studies on the distribution of *halt* and *eben* (e.g., Elspaß 2005), our data show that all speech communities prefer *halt* over *eben*, except for bilingual speakers of heritage German in the US. Here, we only encountered one occurrence of *halt* and no occurrence of *eben*. Interestingly, we found several cases where both MPs appear at the edge of the sentence and not in the middlefield. Peripheral *halt* and *eben* were more frequently used by multilingual than monolingual speakers, where we found only one instance of noncanonical particle placement in the right periphery. Thurmair (2020) argues that *halt* can only appear in the right periphery but not in the left periphery, losing its function as a MP and rather functioning as a discourse particle, toning down the importance of the preceding information. Imo (2008) finds *halt* in the left periphery,

where it still functions as a MP. In our data, *halt* was used as both discourse and modal particle in the left and the right periphery, predominantly frequent in the multilingual speaker groups. Thus, our data indicate that multilinguals not only use *halt* more frequently in a noncanonical position but also with a wider functional spectrum, including the highlighting of important information. However, there might also be sociolinguistic factors at play. While Namibia considers multilingualism the norm, Germany and the US are characterized by monoglossic ideologies. The absence of MPs in the US might indicate a strong influence of English as majority language due to these monoglossic ideologies. Previous studies indicate that these differences in macro contexts lead to different types of contact linguistic varieties (Wiese et al. 2022) and linguistic structures (Bunk et al. 2024, forthcoming). Thus, an external factor of linguistic variation such as different language contact settings might lead to different linguistic structures as an external factor of linguistic variation.

## 2.4 Word order in Russian

Of the three languages considered here, Russian is the most flexible concerning word order. Russian is reported to have a basic SVO order in pragmatically-neutral contexts (i.e., in broad focus contexts). However, Russian allows remarkable word order variation that is highly governed by information structure/status (Švedova 1980, Kovtunova 2002, Sirotinina 2003, Kallestinova 2007, Slioussar 2007, 2011, 2012).

Several studies investigate word order in heritage Russian in Germany. However, these studies do not explicitly investigate the impact of information structure. For example, Brehmer & Usanova (2015) report that bilinguals differed from monolingually-raised speakers of Russian by producing significantly more V-final linearizations in both main and embedded clauses. The authors explain the results in embedded clauses by referring to the influence from the majority language German, which has finite V-final in embedded clauses, whereas the results in main clauses might be due to pragmatic unmarking, in other words, V-final orders being used in a wider range of contexts (for details see Brehmer & Usanova 2015).

In contrast to Brehmer & Usanova (2015), Zuban et al. (2021) and Martynova et al. (2025) found that bilinguals with heritage Russian in Germany were similar to monolingual speakers of Russian regarding the choice of different word orders in both main and embedded clauses, not taking information structural factors into consideration. Zuban et al. (2021) investigate word orders of subject, verb, and (direct or indirect) object (i.e., SVO, OVS, SOV, OSV, VOS, VSO) of 16

adolescent bilinguals residing in the US and Germany and 8 age-matched monolingual speakers of Russian in the Russian subcorpus of the RUEG corpus. The data were manually annotated for clause type (main/embedded), verb type (e.g., auxiliary, copula, finite, gerund, infinitive), and word order pattern. The word orders included in the analysis contained a nonoblique subject, a finite verb, and an object (either direct or oblique). The overall dataset consisted of 783 clauses. The study revealed that bilinguals with heritage Russian in Germany were similar to monolingual speakers regarding their word order repertoire and word order distribution in both main and embedded clauses.

Martynova et al. (2025) examined the choice of OV/VO orders by 24 adolescent bilingual speakers in the US and Germany and monolingual speakers of Russian. The overall dataset consisted of 1,010 clauses. The data were manually annotated for clause type (main/embedded), verb type (e.g., auxiliary, copula, finite, gerund, infinitive), word order pattern (either OV or VO), and object realization (nominal or pronominal). Contrary to Zuban et al. (2021) and Brehmer & Usanova (2015), the study focused on the position of the object in relation to the verb, such that the word orders included in the analysis were OV and VO with either a finite or a non-finite verb and at least one object (direct or oblique). In addition, the study by Martynova et al. (2025) explicitly considered the influence of object realization (noun/pronoun) on OV/VO choice.

The study by Martynova et al. (2025) revealed that bilinguals in Germany were similar to monolingual speakers of Russian. Further, clause type and object role influenced the choice of OV vs. VO order in the bilingual and monolingual speakers similarly. Specifically, and in accordance with the literature on Standard Russian, the probability of OV over VO orders significantly decreased in embedded clauses compared to main clauses. Furthermore, and again in line with the literature on Standard Russian, objects realized by pronouns were associated with OV orders, while objects realized by nouns were associated with VO orders in narrations of both speaker groups.

As already pointed out, the findings in Zuban et al. (2021) and Martynova et al. (2025) contradict those of Brehmer & Usanova (2015). These differences could be due to 1) different text types (semi-spontaneous narrations in different registers in Zuban et al. (2021) and Martynova et al. (2025); vs. written narrations in Brehmer & Usanova (2015)), 2) differences in data annotation, and/or 3) differences in the data grouping and analysis.

As for the influence of information packaging on word order of heritage speakers in Germany, some studies such as Zuban et al. (2021) and Brehmer & Usanova (2015) acknowledge that word orders produced by heritage speakers are not always used to express the expected information structure or status. However, the above-mentioned studies do not explicitly focus on information packaging and

do not provide any quantitative results to support their observations. In the following paragraphs, we discuss one study that has examined the expression of information status by heritage speakers of Russian in Germany (Zuban 2023).

In formal speech in monolingual Russian, in accordance with the given-before-new principle, given referents canonically occur before new referents and new referents appear postverbally and clause-finally (Slioussar 2007, 2011). In informal speech, this given-new order can be violated (e.g., Sirotinina 2003), leading to a noncanonical new-given order. Zuban (2023) examined the syntactic and prosodic expression of referents with the order new-given produced in the RUEG corpus. The study analyzes the data of 120 speakers of the following groups: heritage speakers of Russian in the US ( $N = 40$ : 20 adolescents and 20 adults), heritage speakers of Russian in Germany ( $N = 40$ : 20 adolescents and 20 adults), and monolingual speakers of Russian ( $N = 40$ : 20 adolescents and 20 adults). The study focuses (among other things) on two questions: 1) whether heritage speakers of both groups and monolingual speakers in Russia produce noncanonical orders of referents (i.e., new before given referents in one clause), and 2) whether the frequency of the order of referents new-given is similar across the three speaker groups.

The data were manually annotated for word order based on Bailyn (2012) and Villavicencio (2002). Information status of the 23 most frequent referents was manually annotated according to the RefLex scheme (Riester & Baumann 2017) and language-specific principles such as the position of a referent in a clause or the expression of a referent by lexical or morphosyntactic means. Four types of referents were annotated: new, bridging, unused, and given (see their definitions in Riester & Baumann 2017: 5). The study shows that heritage speakers in Germany and monolingual speakers produced referents with the order new-given, as shown in the “target” in (4):

(4) Context:

i: v tot moment idët čelovek s mjačom (-) gde (-) u nego iz ruki (-)  
otpuskaetsja mjač i vykatyvaetsja na dorogu

‘and: at that moment there is a person walking with a ball where the ball is dropped from his hand and rolls out onto the road’

Target:

# potom sobaka (-) uvidela mjač

|                  |     |                    |
|------------------|-----|--------------------|
| S <sub>new</sub> | V   | O <sub>given</sub> |
| then             | dog | saw                |
|                  |     | ball               |

‘then a dog saw the ball’

Expected word order:

potom mjač uvidela sobaka

$o_{\text{given}}$  V       $s_{\text{new}}$  (RUEG corpus, DEBi74MR\_isR)

Overall, heritage speakers in Germany produced 52 instances of new-given combinations out of 408 clauses with different discourse-new referents (i.e., new, unused, and bridging), while monolinguals had 22 combinations of new-given out of 292 clauses with different discourse-new referents. The numerical difference of new-given combinations between the two speaker groups was statistically significant, i.e., heritage speakers in Germany produced significantly more combinations of new-given referents than monolingual speakers.

In the US, research generally reports on the increase of SVO and the reduction of word order flexibility in productions of Russian-English bilinguals (e.g., Isurin 2005, Kagan & Dillon 2006, Polinsky 2006, Isurin & Ivanova-Sullivan 2008, Laleko & Dubinina 2018). However, Zuban et al. (2021) found that other factors, such as clause type, may modulate the reduction of word order flexibility and the increase of SVO. They found that bilingual speakers in the US were similar to Russian monolinguals in main clauses but not in embedded ones, where they predominantly produced SVO orders. Zuban et al. (2021) argue that these results are not caused by transfer since transfer effects should emerge in both main and subordinate clauses. Rather, higher complexity of embedded clauses compared to main clauses led heritage speakers to the increased use of SVO word order (see Zuban et al. 2021 for a detailed discussion).

Martynova et al. (2025) found that, like bilingual speakers in Germany, bilinguals in the US and monolingual speakers of Russian are similar regarding their preference for OV and VO orders. Furthermore, the two groups of speakers were similar concerning clause type and object realization (noun vs. pronoun). Again, the different results in Martynova et al. (2025) and Zuban et al. (2021) are most likely due to the differences in the investigated phenomenon.

Regarding information packaging, some studies in the US report that word orders produced by heritage speakers are sometimes “contextually inappropriate”, i.e., they do not always express the intended information structure or information status (e.g., Laleko & Dubinina 2018, Kisilev 2019). This terminology of “contextually inappropriate” is highly problematic as it considers the monolingual formal language to be a norm for heritage speakers although the latter usually do not have exposure to formal instruction in their heritage language (see Wiese et al. 2022 for a discussion). Several studies suggest differences between bilingual and monolingual speakers and hint at different dynamics regarding information packaging. For instance, “contextually inappropriate” dislocation was

found to comprise from around 12% up to 30% of all word orders with dislocation in the data of bilinguals while monolingual speakers were reported to always produce “contextually appropriate” dislocation (Laleko & Dubinina 2018: 203, Kisselev 2019: 164). Importantly, both bilinguals and monolinguals produced word orders with dislocation. With respect to new-given orders in heritage Russian in the US, we found that heritage speakers of Russian in the US produced referents with the order of constituents new-given (61 instances out of 437 clauses) significantly more frequently than monolingual speakers (see Zuban 2023, 2024). The results were thus in line with the findings from heritage Russian in Germany as reported above.

Taken together, the results from our studies provide valuable insights into the issue of word order choices in heritage Russian. At first glance, bilinguals prefer the same word orders as monolingual speakers (Martynova et al. 2025, Zuban et al. 2021). However, once information packaging is added to the equation, the differences between bilinguals and monolinguals become obvious (Zuban 2023, 2024). Concurrently, the results of Zuban (2023, 2024) confirm the predictions of the Interface Hypothesis, according to which phenomena that lie at the external interface of syntax and discourse are predicted to show increased variability under language contact (Sorace 2011). The exact reasons behind this variability are less clear. However, the constant exposure of bilingual speakers to the various linguistic structures of their majority language and heritage language as well as different quantity and quality of input in the heritage languages might lead to the outcomes of heritage speakers that differ from monolingual speakers (see Sorace & Serratrice 2009, Zuban et al. 2023).

This section has summarized studies conducted within RUEG, showing that word order phenomena are modulated through information structure and status. These studies indicate that monolingual and bilingual speakers produce non-canonical patterns for specific information packaging purposes, for example regarding referent introduction in English. Some phenomena even highlighted that the relationship between information structure and word order seems particularly pronounced in bilingual speakers, such as V3 orders and modal particles in German and the order of new and given referents in Russian.

These findings provide valuable insights into the relationship between information packaging and word order in English, German and Russian, discussing isolated syntactic phenomena, with the major unifying factor being the influence of information packaging on word order. As will be shown below, focussing on a pattern occurring across these languages allows for a more systematic analysis of the interaction of information structure and word order under language contact. Taking this perspective, we might be able to pinpoint cross-linguistic effects

as well as more general, language external factors that might impact the interaction of information packaging and word order. In the next section, we look at left dislocation (LD) constructions as a case in point for such an endeavor.

### 3 Left dislocation constructions: A comparative view

#### 3.1 LD in English, German, and Russian

Left dislocation (henceforth LD) is defined as “a construction in which a constituent (e.g., a noun, a full pronoun, etc.) that appears before/to the left of its predicate has, within the same sentence, a (nonreflexive) coreferential pronoun” (Duranti & Ochs 1979: 378).

The construction is reported for most documented languages and is thus considered a universal phenomenon (Lambrecht 2001, Westbury 2016). This makes LD constructions particularly revealing regarding the effects of language contact on syntax and information packaging, especially when the languages involved differ concerning word order flexibility and exhibit slight grammatical and functional differences in LD constructions (see Westbury 2016 for an extensive overview of grammatical and functional features of LD across different languages). Several studies in the RUEG group investigated LD constructions in Russian (Zerbian, Barabashova, et al. 2025), English (Pashkova et al. submitted), and German (Conti 2022, Sluckin & Bunk 2023), both in monolingual speakers and bilingual speakers. (5) exemplifies LD constructions in English, Russian, and German:

- (5) a. English  
    *My father, he's Armenian.* (Prince 1997: 2)
- b. Russian  
    *Moskv-a, ona gorodam mat'.*  
    Moscow-NOM she cities mother  
    ‘Moscow is the mother of cities.’ (King 1995: 103)
- c. German  
    *Die Brigitte die kann ich schon gar nicht leiden.*  
    the Brigitte she can I MP MP not like  
    ‘Brigitte, I don't like her at all.’ (Altmann 1981: 48)

LD constructions are a specific type of topic constructions where the dislocated constituent is topicalized through its initial position and resumption (see, among many others, Altmann 1981).

LDs (re)introduce or (re)activate discourse referents (see Westbury 2016) or promote topics (see Gregory & Michaelis 2001, Frey 2005). On the level of discourse pragmatics, they are used as floor-seeking devices (see Duranti & Ochs 1979). Even though there is no systematic analysis of LD in different communicative situations or registers, several studies have pointed out that LD tends to occur in informal (Geluykens 1992) and spoken language (Shaer & Frey 2004, Guryev et al. submitted). In spoken discourse, LD is particularly frequent in narrations and discourses with present interlocutors (Bousquette et al. 2021).

Similar LD types have been categorized using different terminology in different languages, causing confusion from a cross-linguistic perspective. For example, Frey (2005) indicates that the term “left dislocation” is misleading when comparing German with English. While in English, the resumptive occurs in the canonical position of the dislocated constituent in the following syntagma, in German, the resumptive appears right after the dislocated constituent. However, resumptives in German may also appear in other positions in the following syntagma. In this case, the construction is called “Hanging Topic (Left Dislocation) Construction” (HTLD; Altmann 1981, Selting 1993, Frey 2005). Frey (2005) refers to constructions such as in (6c) as “Contrastive Left Dislocations” (CLD). (6) illustrates the different constructions.

- (6)    a. Maria, I know her. (English LD)
- b. Maria, die kenne ich. (German CLD)
- c. Maria, ich kenne sie. (German HTLD)

Various subtypes of LD constructions differ regarding their grammatical structure. In German CLD, NPs, PPs, CPs, APs, and AdvPs can be dislocated (Dewald 2012). These constituents are resumed by a *d*-pronoun (*der*, *die*, *das*) or an adverbial (see Sluckin & Bunk 2023 on adverbial resumption). HTLD in German is less restricted than CLD concerning the resumptive. It can occur as a *d*-pronoun, personal pronoun, a phrase, or it can be absent (see Altmann 1981, Dewald 2012). HTLDs are usually restricted to left dislocated NPs or PPs (Selting 1993). Russian and English LD constructions typically involve NPs that are resumed by a subject or object pronoun (King 1995, Kallestinova 2007). LD constructions in the three languages also differ concerning prosodic realization. While the left dislocated element is usually separated from the rest of the clause by a pause in Russian (Bailyn 2012), English (Frey 2005), and German HTLD, German CLDs lack such a pause and are prosodically integrated in the following syntagma (Altmann 1981, Selting 1993, Dewald 2012). In our analysis, we subsume all these different constructions under the term “left dislocation” (LD) for English, Russian, and German. In order to account for grammatical and functional differences between the

constructions, we annotated LD constructions concerning different grammatical features using a joint annotation scheme (see Section 3.2).

Even though LD is a widespread construction in the languages of the world, only a few studies have systematically investigated LD from the perspective of language contact. Generally, studies indicate that bilingual speakers have “robust knowledge of LD constructions” and “behave like monolingual native controls as regards production, interpretation, and use” (Bousquette et al. 2021: 11). In their investigation of heritage Norwegian and heritage German in the US, Bousquette et al. (2021: 17) find that “specific constraints on LD appear to have been weakened”. Overall, however, bilinguals patterned with monolinguals. Similar results have been found for adverbial resumption in Germany in informal German and Kiezdeutsch (Sluckin & Bunk 2023). Both varieties display similar patterns, even though Kiezdeutsch generally allows for more word order variation than the German of monolinguals, for example concerning verb placement (see Wiese & Rehbein 2016), which we already discussed above.

For Canadian French and English, Nagy et al. (2003) report that anglophones in Montreal use a distinct type of LD (“subject doubling”) in their French, but the construction is also known from the varieties of Canadian English spoken in Ontario, giving it a “distinctly French flavor” (Tagliamonte & Jankowski 2019: 1). Usually, this construction is rare in English but frequent in French. While at first glance, this observation might indicate cross-linguistic effects, the construction is also found in speech communities with less direct influence from French. Thus, Tagliamonte & Jankowski (2019: 13) conclude that the construction does not come from French but that “French is influencing the use of English”.

For Russian, Laleko & Dubinina (2018) investigate different word order patterns in bilingual and monolingual speakers of Russian in the US. Their study looks at different word order patterns, though not specifically including the LD constructions described here. They found that bilinguals produced fewer non-canonical patterns than monolinguals, indicating decreased word order flexibility in Russian.

In the following, we present first results from our comparative study on LD in language contact and contextualize findings from ongoing or published studies on LD in English, German, and Russian in language contact to gain further insights into LD constructions from a cross-linguistic perspective.

### **3.2 Data annotation**

The empirical basis is the RUEG corpus as described above. LD constructions were identified in the RUEG corpus manually and annotated using a joint anno-

tation scheme allowing for systematic, cross-linguistic analyses. We thoroughly examined the corpus to identify all instances of LDs according to the definition provided at the beginning of Section 3. Subsequently, we proceeded to annotate each LD based on five features: 1. referent, 2. pronoun type (personal, possessive, partitive), 3. noun phrase type (simple noun phrase, noun phrase with a preposition, noun phrase with coordination, noun phrase with a relative clause), 4. presence of intervening material (present or absent), and 5. function (new introduction, reintroduction, set, clarification).

### 3.3 Findings

Overall, our data show that all speaker groups in Germany and the US use LDs, and the construction is almost always used in spoken contexts. However, the groups differ concerning the number of LDs used by particular groups in the respective countries. Differences also emerged for LD use in formal vs. informal communicative situations. Table 4 gives a general overview of the number of LDs normalized against the number of CUs.<sup>4</sup>

While the overall figures do not indicate major differences at first glance, our statistical analyses revealed interesting results. For Russian, bilinguals in both the US and Germany used a significantly higher number of LDs in total than monolinguals (see Zerbian, Barabashova, et al. 2025). This is in contrast to the literature reporting a decreased word order flexibility in bilinguals (e.g., Laleko & Dubinina 2018). In addition to the number of LDs, more bilingual speakers (again in both the US and Germany) than monolinguals produced LDs at all. Formality was not a significant factor for the groups investigated, i.e., there was no difference between formal and informal contexts, pointing to register leveling in these groups (see Özsoy et al. 2022).

In German, monolinguals and bilinguals did not differ significantly from each other. LDs were produced at similar rates. However, formality had a minor influence when comparing monolinguals and bilinguals speaking Turkish as a heritage language. German monolinguals produced fewer LDs in the informal than in the formal context, while bilinguals of heritage Turkish produced fewer in the formal than in the informal setting. These results were only marginally significant; further investigations with a larger data set might thus enlighten this aspect. Interestingly, bilinguals used a structure not yet attested in the literature. They occasionally (9 times) produced an LD structure with a personal pronoun, where a *d*-pronoun would be expected, as discussed in the literature. (7) provides an example:

---

<sup>4</sup>We define a communication unit (CU) as an “independent clause with its modifiers” (Hughes et al. 1997: 53).

Table 4: Overall frequencies of LDs across speaker groups

|                     | Left<br>dislocations | CUs  | LDs per CU |
|---------------------|----------------------|------|------------|
| <b>RUEG-RU</b>      |                      |      |            |
| mo-Russian (Russia) | 26                   | 2780 | 0.009      |
| h-Russian (Germany) | 90                   | 3374 | 0.030      |
| h-Russian (US)      | 103                  | 3222 | 0.030      |
| <b>RUEG-EN</b>      |                      |      |            |
| mo-English (US)     | 24                   | 2734 | 0.009      |
| h-German (US)       | 15                   | 1499 | 0.009      |
| h-Greek (US)        | 19                   | 2879 | 0.007      |
| h-Russian (US)      | 41                   | 3170 | 0.013      |
| h-Turkish (US)      | 30                   | 2864 | 0.010      |
| <b>RUEG-DE</b>      |                      |      |            |
| mo-German (Germany) | 19                   | 3727 | 0.005      |
| h-German (US)       | 2                    | 1633 | 0.001      |
| h-Turkish (Germany) | 20                   | 3627 | 0.006      |
| h-Russian (Germany) | 12                   | 3639 | 0.003      |
| h-Greek (Germany)   | 4                    | 2251 | 0.002      |

- (7) dieser mann er spielt mit äh ball  
 this man he plays with uhm ball  
 ‘this man, he plays with the ball’ (RUEG corpus, DEBi59MT\_isD)

Like LD in German, the number of LDs in English was similar between monolinguals and bilinguals – slightly above or under 1% out of all CUs. However, speaker groups differed in using LDs in formal vs. informal situations. We found different patterns of use between bilinguals with the heritage languages Greek and Turkish on the one hand and bilinguals with heritage German, heritage Russian, and monolinguals on the other hand. In the formal situations, speakers of the heritage languages Greek and Turkish used fewer LDs than monolinguals, while bilinguals with heritage German and heritage Russian did not differ from monolinguals. In the informal context, bilinguals with heritage Turkish produced more LDs than monolinguals, while bilingual speakers with heritage Greek, heritage German, and heritage Russian did not differ from monolinguals. Finally,

bilinguals with heritage Greek and heritage Turkish approached the formality distinction differently from English monolinguals: these bilinguals had slightly more LDs in the informal than in the formal setting, while monolinguals had a reverse pattern: they had slightly more LDs in the formal setting than in the informal one. Bilinguals with heritage German and heritage Russian did not differ from monolinguals.

### 3.4 Discussion

The study on LD in the three languages leaves us with valuable insights and many questions. For one, cross-linguistic influence might explain some of the results, such as the more frequent use of LD in heritage Russian compared to monolinguals, and the new pattern in German of replacing the resumptive *d*-pronoun with a personal pronoun. Bilingual heritage Russian speakers might use the noncanonical structure more frequently than monolingual English speakers since, in both their languages, LD is a plausible option for topic promotion or (re)introduction. The general flexibility of Russian word order compared to Turkish, Greek, and German might amplify the use of LD in such contexts. Statistical analyses are needed to further explore whether heritage speakers' majority language aligns with their heritage language regarding LD, in other words, whether in this case speakers use LDs more regularly.

Cross-linguistic transfer, however, does not explain the distributional patterns across different communicative contexts: while formality was not a relevant factor for heritage Russian in Germany, bilinguals with heritage Turkish produced slightly more LDs in informal than in formal contexts, while monolinguals produced more LDs in formal than in informal situations. Interestingly, we found a similar pattern in the US. Again, bilinguals with heritage Turkish (and heritage Greek) produced fewer LDs in formal contexts than monolinguals or bilinguals with heritage German and heritage Russian, who, in turn, produced fewer LDs in informal than in formal situations. This finding is rather surprising, as it concerns particular speaker groups. While for Russian, linguistic factors might be at play, as Russian word order flexibility appears to provide speakers with a wider range of grammar patterns, in heritage Greek and heritage Turkish, potential reasons for the distribution of LDs in formal and informal settings might be extralinguistic factors. In the final section, we suggest how sociolinguistic factors might provide further insights into LD use in language contact and that these factors might influence linguistic structure.

## 4 Conclusion and outlook

This chapter aimed to provide an overview of studies conducted within RUEG exploring the relationship between information packaging and word order in language contact settings. These studies suggest that both monolinguals and bilinguals make use of noncanonical structures that are influenced by information structure and status. In English, monolinguals and bilinguals used non-SVO to introduce new subjects, while in Russian, both groups used noncanonical new-before-given word orders. In German, noncanonical V3 and peripheral modal particles are used by both bilinguals and monolinguals and in all of these languages, LDs occur in both monolingual and bilingual groups.

Many of our studies also imply that bilingual speakers are flexible regarding word order and its interaction with information structure and status. In Russian, bilinguals use LDs more frequently, contrary to the decrease in flexibility in bilinguals with heritage Russian, as often claimed in the literature. In addition, they apply noncanonical new-before-given word order more frequently than monolinguals. Even though some previous studies did not find any differences in word order between monolinguals and bilinguals (e.g., Martynova et al. 2025), or these differences were not found across all the investigated conditions (e.g., Zuban et al. 2021), the picture changes when considering how new vs. given referents are presented. In German, V3 is used more frequently by bilingual speakers, and bilinguals place the modal particles *halt* and *eben* more frequently in the sentence peripheries, taking on the functions of discourse particles and highlighting important information.

Our data also indicate that bilingual speakers are not only more dynamic but tend to be more sensitive to communicative situations, as we have seen for the preference of LD in informal situations in most bilingual groups, except for Russian. These effects cannot be explained by cross-linguistic transfer, as we would expect differences in all communicative situations for specific speaker groups. Considering sociolinguistic factors might further illuminate these findings. Bunk (2024) found that many German bilingual speakers with heritage Turkish and heritage Russian produced fewer noncanonical patterns and more features associated with formal, standard-like German on several linguistic levels (lexicon, phonetics, discourse-pragmatics) in the formal situations compared to monolinguals. For example, speakers tend to articulate non-morphemic *-t* in the auxiliary *ist* ('is') more frequently than their monolingual peers. In spoken language, speakers often drop non-morphemic *-t* due to a regular phonological process. This variant, however, deviates from written standard German and may not be associated with formal language. Bunk (2024, *in press*) argues that a reason for

such patterns may be linguistic pressure that multilingual speakers experience in a societal macro context that is characterized by widespread monoglossic ideologies. Multilingual speakers might feel the need to align with the majority society linguistically to be accepted as members of that society and to avoid Othering. This conclusion is further supported by qualitative sociolinguistic interviews, in which speakers reported the need to excel over monolinguals regarding standard language use in order to be considered valid members of the German society Bunk (2024, *in press*). An example for such a notion, uttered by a bilingual speaker with Russian as their heritage language in Germany, is given in (8):

- (8) sagen wir mal, man sieht ausländisch aus, [...] man wird einfach anders wahrgenommen, wenn das Deutsch perfekt ist  
'let's say you look foreign, [...] you are simply perceived differently if your German is perfect' (CS\_hR, 3:56)

The interview data illustrate strong monoglossic ideologies, standard language ideology, and perceived linguicism as indicated in (8). These ideologies are present in both the US and Germany, which is typical for countries of the Global North (Lippi-Green 1997, Blackledge 2000). Hence, they might affect the use of noncanonical patterns, such as LDs, in bilingual speakers in these countries. However, this interpretation of the data does not explain why speaker groups differ from each other. Two possible factors come to mind: a) different experiences of discrimination due to socialization, and b) different ideologies towards multilingualism and the standard variety of the majority language between the speaker groups.

If the societal macro context plays a role, we would not expect fewer LDs in formal contexts than in informal contexts in bilinguals and more LDs in formal contexts than in informal contexts by monolinguals in societies that perceive multilingualism as normalcy, e.g., Namibia. In an ongoing study, Conti & Bunk (*in preparation*) compare the use of LDs in German in Germany, the US, and Namibia, investigating this question further. However, we found other patterns that support this idea in the use of modal particles (see Bunk et al. 2024 and Section 2.2), and bare NPs (Bunk et al. *forthcoming*). Additionally, Wiese et al. (2022) show that the societal macro context seems to influence the overall structure of languages strongly. Even though this perspective does not explain all of our data, it might explain the fewer occurrences of LDs in formal situations in some bilingual speakers of heritage Turkish and heritage Greek in the US and heritage Turkish in Germany compared to monolinguals. While more in-depth studies are needed to determine language ideologies that are present in the soci-

etal macro-context, such as through systematic discourse analysis of public discussions or policy papers, these findings indicate that linguistic ideologies might be an important factor to consider when analyzing linguistic structures in language contact, in particular regarding noncanonical variation.

The case of left dislocation shows that in addition to looking at language-specific phenomena, systematic cross-linguistic analyses of similar patterns are important to illuminate the interaction of information packaging and word order variation. We further suggest that integrating sociolinguistic factors such as the societal macro contexts and ideologies towards multilingualism into our models of grammatical variation might further deepen our understanding of linguistic dynamics in heritage speakers' language use.

## Acknowledgements

The research for this paper was funded by the DFG (Research Unit FOR 2537, P6 (394838878) and P8 (313607803)). We thank the editors of this volume, two anonymous external reviewers, and three internal reviewers (Cem Keskin, Luca Szucsich, and Rosemary Tracy) for their valuable feedback. For all remaining errors and shortcomings, the authors, of course, take full responsibility. We would also like to thank all research assistants involved in the projects: Daria Alkhimchenkova, Erica Conti, Amelie Ellerich, Isabell Furtkert, Yuliia Ivashchyk, Alexander Lehmann, Birte Pravemann, Sharon Rauschenbach, Guendaline Reul, Myrto Rompaki, Alina Schöpf, Madeleine Spitzer, Chris Allison, Ricarda Bothe, Mert Can, Ryan Carroll, Franziska Cavar, Leah Doroski, Mary Elliott, Hannah Lee, Mark Murphy, Mariia Naumovets, Simge Sargin Kisacik, Jasmine Segarra, Golshan Shafeebae, Selena Song, Shreya Srivastava, Fiona Wong and Charlott Thomas for their support. We additionally thank Lea Coy for help with pre-publication formatting. Finally, we thank the whole RUEG team and the Mercator fellows for numerous fruitful discussions, feedback, and great support.

## References

- Alexiadou, Artemis & Terje Lohndal. 2018. V3 in Germanic: A comparison of urban vernaculars and heritage languages. In Mailin Antomo & Sonja Müller (eds.), *Non-canonical verb positioning in main clauses* (Linguistische Berichte Sonderheft 25), 245–264. Hamburg: Buske.

- Altmann, Hans. 1981. *Formen der Herausstellung im Deutschen: Rechtsversetzung, Linksversetzung, freies Thema und verwandte Konstruktionen* (Linguistische Arbeiten 106). Berlin & Boston: De Gruyter.
- Arnbjörnsdóttir, Birna, Höskuldur Thráinsson & Iris Edda Nowenstein. 2018. V2 and V3 orders in North-American Icelandic. *Journal of Language Contact* 11(3). 379–412. DOI: 10.1163/19552629-01103002.
- Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76. 28–55.
- Bailyn, John F. 2012. *The syntax of Russian* (Cambridge syntax guides). Cambridge, UK: Cambridge University Press.
- Baumann, Stefan. 2006. Information structure and prosody: Linguistic categories for spoken language annotation. In Stefan Sudhoff, Denisa Lenertova, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.), *Methods in empirical prosody research* (Language, Context and Cognition 3). Berlin: De Gruyter. DOI: 10.1515/9783110914641.153.
- Biber, Douglas, Stig Johansson, Geoffrey N. Leech, Susan Conrad & Edward Finegan. 2021. *Grammar of spoken and written English*. Amsterdam & Philadelphia: John Benjamins. DOI: 10.1075/z.232.
- Blackledge, Adrian. 2000. Monolingual ideologies in multilingual states: Language, hegemony and social justice in Western liberal democracies. *Sociolinguistic Studies* 1(2). 25–45. DOI: 10.1558/sols.v1i2.25.
- Blühdorn, Hardarik. 2019. Modalpartikeln und Akzent im Deutschen. *Linguistische Berichte* 2019(259). 275–317. DOI: 10.46771/23660775002591.
- Bock, Kathryn J. & David E. Irwin. 1980. Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior* 19(4). 467–484. DOI: 10.1016/S0022-5371(80)90321-7.
- Bousquette, Joshua, Kristin M. Eide, Arnstein Hjelde & Michael T. Putnam. 2021. Competition at the left edge: Left-dislocation vs. topicalization in H-Germanic. In Arnstein Hjelde & Åshildedt Søfteland (eds.), *Selected proceedings of the 10th Workshop on Immigrant Languages in the Americas (WILA 10)*, 11–21. Sommerville, MA: Cascadilla Proceedings Project.
- Brehmer, Bernhard & Irina Usanova. 2015. Let's fix it? Cross-linguistic influence in word order patterns of Russian h-speakers in Germany. In Hagen Peukert (ed.), *Transfer effects in multilingual language development* (Hamburg Studies on Linguistic Diversity), 161–188. Amsterdam: John Benjamins.
- Bross, Fabian. 2012. German modal particles and the common ground. *Helikon: A Multidisciplinary Online Journal* (2). 182–209.

- Bunk, Oliver. 2020. „*Aber immer alle sagen das*“: *The status of V3 in German: Use, processing, and syntactic representation*. Humboldt-Universität zu Berlin. (Doctoral dissertation). DOI: 10.18452/22085.
- Bunk, Oliver. 2024. What does linguistic structure tell us about language ideologies? The case of majority language anxiety in Germany. *European Journal of Applied Linguistics: Language ideologies: Old questions, new perspectives*. Naomi Truan & Esther Jahns (eds.). 91–116. DOI: 10.1515/eujal-2023-0049.
- Bunk, Oliver. In press. The anxious heritage speaker? Language anxiety and insecurity in multilingual contexts. In Artemis Alexiadou, Claudio Scarvaglieri, Christoph Schroeder & Heike Wiese (eds.), *The construction of multilinguals as others: Do we practice what we preach?* Berlin: Language Science Press.
- Bunk, Oliver & Jana Gamper. Forthcoming. *The V2 myth: Evidence for a functional motivation of V3 in L2 German*.
- Bunk, Oliver & Maike H. Rocker. In preparation. *The status of initial adverbials in V3 sentences: A comparative corpus study on prosodic realization*.
- Bunk, Oliver, Antje Sauermann & Fynn Raphael Dobler. 2024. Sociolinguistic variation in Kiezdeutsch and Namdeutsch: The case of the modal particles *halt* and *been*. In James M. Stratton & Karen V. Beaman (eds.), *Expanding variationist sociolinguistic research in varieties of German*. New York: Routledge.
- Bunk, Oliver, Britta Schulte & Heike Wiese. Forthcoming. Bare NPs in German: Results from a comparative corpus study. In Siegwalt Lindenfelser, Angélica Prédiger & Renata Szczepaniak (eds.), *Deutsche Sprachminderheiten in der Welt: Empirische Studien zur Sprachvariation und Sprachideologie* (Germanistische Linguistik). Berlin: de Gruyter.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.
- Conti, Erica. 2022. *German left dislocations across communicative situations: A comparative corpus study*. Berlin: MA thesis.
- Conti, Erica & Oliver Bunk. In preparation. *Left dislocation in German varieties: A comparative view*.
- Dewald, Anika. 2012. *Versetzungssstrukturen im Deutschen*. Köln: Universität zu Köln. (Dissertation).
- Diewald, Gabriele. 2007. Abtönungspartikel. In Ludger Hoffmann (ed.), *Handbuch der deutschen Wortarten* (De Gruyter Lexikon), 117–142. Berlin & New York: De Gruyter.
- Duranti, Alessandro & Elinor Ochs. 1979. Left-dislocation in Italian conversation. In Talmy Givon (ed.), *Discourse and syntax*, 377–416. New York: Academic Press.

- Elspaß, Stephan. 2005. Zum Wandel im Gebrauch regionalsprachlicher Lexik. Ergebnisse einer Neuerhebung. *Zeitschrift für Dialektologie und Linguistik* 72(1). 1–51.
- Frey, Werner. 2005. Pragmatic properties of certain German and English left peripheral constructions. *Linguistics* 43(1). 89–129.
- Geluykens, Ronald. 1992. *From discourse process to grammatical construction: On left-dislocation in English* (Studies in Discourse and Grammar 1). Amsterdam & Philadelphia: John Benjamins. DOI: 10.1075/sidag.1.
- Gregory, Michelle L. & Laura A. Michaelis. 2001. Topicalization and left-dislocation: A functional opposition revisited. *Journal of Pragmatics* 33(11). 1665–1706. DOI: 10.1016/S0378-2166(00)00063-1.
- Guryev, Alexander, Victoriya Trubnikova & François Delafontaine. Submitted. *Left-dislocation constructions in Russian: A functionalist approach in terms of “macro-syntactic routines”*.
- Halliday, Michael A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of linguistics* 3(2). 199–244.
- Haviland, Susan E. & Herbert H. Clark. 1974. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* 13(5). 512–521. DOI: 10.1016/S0022-5371(74)80003-4.
- Helbig, Gerhard. 1988. *Lexikon deutscher Partikeln*. 1. Aufl. Leipzig: VEB Verl. Enzyklopädie.
- Hentschel, Elke. 1986. *Funktion und Geschichte deutscher Partikeln*: Ja, doch, halt und eben (Reihe Germanistische Linguistik 63). Tübingen: Max Niemeyer Verlag.
- Hilpert, Martin. 2021. Information structure. In Bas Aarts, April MacMahon & Lars Hinrichs (eds.), *Handbook of English Linguistics*, 229–248. New York: Wiley.
- Höhle, Tilman N. 1986. Der Begriff „Mittelfeld“: Anmerkungen über die Theorie der topologischen Felder. In Walter E. Weiss, Herbert E. Wiegand & Marga Reis (eds.), *Kontroversen, alte und neue: Akten des VII. Internationalen Germanisten-Kongresses Göttingen 1985*, 329–340. Tübingen: Niemeyer.
- Huddleston, Rodney. 2002. Syntactic overview. In Rodney Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 43–70. Cambridge: Cambridge University Press. DOI: 10.1017/9781316423530.003.
- Hughes, Diana L., LaRae McGillivray & Mark Schmidek. 1997. *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking Publications.
- Imo, Wolfgang. 2008. Individuelle Konstrukte oder Vorboten einer neuen Konstruktion? Stellungsvarianten der Modalpartikel halt im Vor- und Nachfeld.

- In Anatol Stefanowitsch & Kerstin Fischer (eds.), *Konstruktionsgrammatik II* (Stauffenburg Linguistik), 135–156. Tübingen: Stauffenburg-Verl.
- Isurin, Ludmila. 2005. Cross-linguistic transfer in word order: Evidence from L1 forgetting and L2 acquisition. In James Cohen, Kara T. McAlister, Kellie Rolstad & Jeff McSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism*, 1115–1130. Somerville, MA: Cascadilla Press.
- Isurin, Ludmila & Tania Ivanova-Sullivan. 2008. Lost in between: The case of Russian speakers. *Heritage Language Journal* 6(1). 72–104.
- Jasinskaja, Katja. 2016. Information structure in Slavic. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure* (Oxford handbooks in linguistics), 709–732. Oxford: Oxford University Press.
- Kagan, Olga & Kathleen Dillon. 2006. Russian heritage learners: So what happens now? *The Slavic and East European Journal* 50(1). 83–96. DOI: 10.2307/20459235.
- Kallestinova, Elena. 2007. *Aspects of word order in Russian*. Iowa City: University of Iowa. (Doctoral dissertation).
- King, Tracy Holloway. 1995. *Configuring topic and focus in Russian* (Dissertations in linguistic series). Stanford, CA: CSLI Publications.
- Kinn, Kari & Ida Larsson. 2021. Pronominale demonstrativer: Nye perspektiver fra norsk og svensk. *Oslo Studies in Language* 11(2). 201–224.
- Kinn, Kari & Ida Larsson. 2022. Pronominal demonstratives in homeland and heritage Scandinavian. eng. *Nordic journal of linguistics* 45(3). 281–309.
- Kisselev, Olesya. 2019. Word order patterns in the writing of heritage and second language learners of Russian. *Russian Language Journal* 69. 149–174.
- Kotsinas, Ulla-Britt. 1992. Immigrant adolescents' Swedish in multicultural areas. In Cecilia Palmgren, Karin Lövgren & Göran Bolin (eds.), *Ethnicity in youth culture*, 43–62. Stockholm: University of Stockholm.
- Kovtunova, Irina I. 2002. *Sovremennyj russkij jazyk: Porjadok slov i aktual'noe členenie predloženij* [Modern Russian language: Word order and actual division of the sentence]. URSS.
- Krifka, Manfred & Renate Musan. 2012. Information structure: Overview and linguistic issues. In Manfred Krifka & Renate Musan (eds.), *The expression of information structure* (The expression of cognitive categories), 1–44. Berlin & Boston: Mouton de Gruyter.
- Kühl, Karoline & Jan Heegård Petersen. 2018. Word order in American Danish declaratives with a non-subject initial constituent. eng. *Journal of Language Contact* 11(3). 413–440.

## 11 Information packaging and word order dynamics in language contact

- Laleko, Oksana & Irina Dubinina. 2018. Word order production in heritage Russian: Perspectives from linguistics and pedagogy. In Susan Bauckus & Susan Kresin (eds.), *Connecting across languages and cultures: A heritage language Festschrift in honor of Olga Kagan*, 191–215. Bloomington: Slavica.
- Lambrecht, Knud. 2001. Dislocation. In Gerold Ungeheuer (ed.), *Language typology and language universals*, 1050–1078. Berlin & Boston: Mouton.
- Lippi-Green, Rosina. 1997. What we talk about when we talk about Ebonics: Why definitions matter. *The Black Scholar* 27(2). 7–11.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehayé, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Martynova, Maria, Yulia Zuban, Luka Szucsich & Natalia Gagarina. 2025. OV/VO word order in heritage Russian: Is transfer at play? *Journal of Slavic Linguistics* 33. <https://ojs.ung.si/index.php/JSL/article/view/411>.
- Meelen, Marieke, Khalid Mourigh & Lisa Lai-Shen Cheng. 2020. V3 in urban youth varieties of Dutch. In András Bárány, Theresa Biberauer, Jamie Douglas & Sten Vikner (eds.), *Syntactic architecture and its consequences II: Between syntax and morphology*, 327–355. Berlin: Language Science Press. DOI: 10.5281/zenodo.4280657.
- Musan, Renate. 2002. Informationsstrukturelle Dimensionen im Deutschen: Zur Variation der Wortstellung im Mittelfeld. *Zeitschrift für germanistische Linguistik* 30(2). 198.
- Nagy, Naomi G., Hélène Blondeau & Julie Auger. 2003. Second language acquisition and “real” French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change* 15(01). 73–103. DOI: 10.1017/S0954394503151034.
- Opsahl, Toril. 2009. Wolla I swear this is typical for the conversational style of adolescents in multiethnic areas in Oslo. *Nordic Journal of Linguistics* 32(2). 221–244. DOI: 10.1017/s0332586509990059.
- Özsoy, A. Sumru (ed.). 2019. *Word order in Turkish* (Studies in Natural Language and Linguistic Theory 97). Berlin: Springer. DOI: 10.1007/978-3-030-11385-8.
- Özsoy, Onur, Kateryna Iefremenko & Christoph Schroeder. 2022. Shifting and expanding clause combining strategies in heritage Turkish varieties. *Languages* 7(3). 242. DOI: 10.3390/languages7030242.

- Pashkova, Tatiana, Hannah Lee, Mark Murphy & Shanley E. M. Allen. Submitted. *Left dislocations in heritage speakers' majority English in the US*.
- Polinsky, Maria. 2006. Incomplete Acquisition: American Russian. *Journal of Slavic linguistics* 14(2). 191–262.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Prince, Ellen F. 1997. On the functions of left-dislocation in English discourse. In Akio Kamio (ed.), *Directions in functional linguistics* (Studies in language companion series (SLCS)), 117–143. Amsterdam & Philadelphia: John Benjamins.
- Quist, Pia. 2000. Ny københavnsk “multietnolekt”: Om sprogbrug blandt unge i sprogligt og kulturelt heterogene miljøer [New Copenhagen “multi-ethnolect”: About language use among young people in linguistically and culturally heterogeneous environments]. *Danske Talesprog* 1. 143–211.
- Riester, Arndt & Stefan Baumann. 2017. *The RefLex scheme: Annotation guidelines* (SinSpeC: Working Papers of the SFB 732, no. 14). Stuttgart: Universitätsbibliothek der Universität Stuttgart.
- Roberts, Ian & Anna Roussou. 2002. The Extended Projection Principle as a condition on the tense dependency. In Peter Svenonius (ed.), *Subjects, expletives, and the EPP* (Oxford Studies in Comparative Syntax), 125–155. Oxford: Oxford University Press.
- Rocker, Maike H. 2022. *Variation in finite verb placement in heritage Iowa Low German: The role of prosodic integration and information structure*. State College: Pennsylvania State University. (PhD Thesis).
- Schalowski, Sören. 2017. From adverbial to discourse connective. Multiple pre-fields in spoken German and the use of *dann* ‘then’ and *danach* ‘afterwards’. In Heike Wiese, Heiko F. Marten, Philip Bracker & Oliver Bunk (eds.), *Arbeitspapiere Sprache, Variation und Migration: Studentische Arbeiten*, 1–45. Potsdam: Online Publication, Universität Potsdam. [https://www.uni-potsdam.de/fileadmin/projects/svm/Arbeitspapiere/No6\\_Schalowski\\_2017.pdf](https://www.uni-potsdam.de/fileadmin/projects/svm/Arbeitspapiere/No6_Schalowski_2017.pdf).
- Selting, Margret. 1993. Voranstellungen vor den Satz: Zur grammatischen Form und interaktiven Funktion von Linksversetzung und Freiem Thema im Deutschen. *Zeitschrift für germanistische Linguistik* 21(3). 291–319.
- Sewell, Alyson. 2015. Sociolinguistic and syntactic variation in Wisconsin German narratives. In B. Richard Page & Michael T. Putnam (eds.), *Moribund Germanic heritage languages in North America*, 224–250. Leiden: Brill. DOI: 10.1163/9789004290211011.

- Shaer, Benjamin & Werner Frey. 2004. “Integrated” and “non-integrated” left-peripheral elements in German and English. *ZAS Papers in Linguistics* 35(2). 465–502. DOI: 10.21248/zaspil.35.2004.238.
- Sirotinina, Olga Borisovna. 2003. *Porjadok slov v russkom jazyke [Word order in Russian]*. Izd. 2., stereotipnoe (Lingvisticheskoe nasledie XX veka). Moskva: Ed. URSS.
- Slioussar, Natalija. 2007. *Grammar and information structure: A study with reference to Russian*. Utrecht: LOT Netherlands Graduate School of Linguistics. (PhD Thesis).
- Slioussar, Natalija. 2011. Processing of a free word order language: The role of syntax and context. *Journal of Psycholinguistic Research* 40(4). 291–306. DOI: 10.1007/s10936-011-9171-5.
- Sluckin, Benjamin Lowell. 2021. *Non-canonical subjects and subject positions*. Berlin: Humboldt-Universität zu Berlin. (PhD Thesis). DOI: 10.18452/23715.
- Sluckin, Benjamin Lowell & Oliver Bunk. 2023. Noncanonical V3 and resumption in Kiezdeutsch. In Karen de Clercq, Liliane Haegeman, Terje Lohndal & Christine Meklenborg (eds.), *Adverbial resumption in verb second languages*, 327–354. Oxford: Oxford University Press. DOI: 10.1093/oso/9780197651148.003.0014.
- Sorace, Antonella. 2011. Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism* 1(1). 1–33. DOI: 10.1075/lab.1.1.01sor.
- Sorace, Antonella & Ludovica Serratrice. 2009. Internal and external interfaces in bilingual language development: Beyond structural overlap. *The international journal of bilingualism: Cross-disciplinary, cross-linguistic studies of language behavior* 13(2). 195–210.
- Švedova, Natalija J. 1980. *Russkaja grammatika*. Place: Moskva.
- Tagliamonte, Sali A. & Bridget L. Jankowski. 2019. Grammatical convergence or microvariation? Subject doubling in English in a French dominant town. *Proceedings of the Linguistic Society of America* 4(1). 1–15. DOI: 10.3765/plsa.v4i1.4514.
- Thielmann, Winfried. 2015. Halt. *Deutsche Sprache* 72(1). 1–10. DOI: 10.37307/j.1868-775X.2015.01.02.
- Thurmair, Maria. 1989. *Modalpartikeln und ihre Kombinationen*. Tübingen: Max Niemayer Verlag. DOI: 10.1515/9783111354569.
- Thurmair, Maria. 2020. Zur Syntax von *halt*: Eine Modalpartikel im Nachfeld? *Sprachwissenschaft* 45(2). 241–273.
- Tracy, Rosemarie & Elsa Lattey. 2010. “It wasn’t easy but irgendwie äh da hat sich’s rentiert, net?”: A linguistic profile. In Michaela Albl-Mikasa (ed.), *Dimensionen der Zweitsprachenforschung*, 53–73. Tübingen: Narr.

- Villavicencio, Aline. 2002. Learning to distinguish PP arguments from adjuncts. In Ming-Syan Cheng, Philip S. Yu & Bing Liú (eds.), *Proceeding of the 6th conference on Natural language learning - COLING-02*, vol. 20, 1–7. Association for Computational Linguistics. DOI: 10.3115/1118853.1118886.
- Ward, Gregory & Betty Birner. 2004. Information structure and non-canonical syntax. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 152–174. Wiley. DOI: 10.1002/9780470756959.ch7.
- Westbury, Josh. 2016. Left dislocation: A typological overview. *Stellenbosch Papers in Linguistics Plus* 50(1). 21–45. DOI: 10.5842/50-0-715.
- Wiese, Heike. 2009. Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. *Lingua* 119(5). 782–806. DOI: 10.1016/j.lingua.2008.11.002.
- Wiese, Heike. 2013. What can new urban dialects tell us about internal language dynamics? The power of language diversity. In Werner Abraham & Elisabeth Leiss (eds.), *Dialektologie in neuem Gewand: Zu Mikro-/Varietätenlinguistik, Sprachenvergleich und Universalgrammatik* (Linguistische Berichte Sonderheft 19), 207–245. Hamburg: Buske.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Anke Lüdeling, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Tracy Rosemarie, Christoph Schroeder, Luka Szucsich, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2019. *RUEG Corpus (Version 0.3.0) [Data set]*. Zenodo. DOI: 10.5281/zenodo.3236069.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Anke Lüdeling, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Tracy Rosemarie, Christoph Schroeder, Luka Szucsich, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2021. *RUEG Corpus*. Version 0.4.0. Zenodo. DOI: 10.5281/zenodo.3236068.

- Wiese, Heike, Oliver Bunk, Fynn Dobler, Ulrike Freywald, Sophie Hamm, Banu Hueck, Anne Junghans, Jana Kiolbassa, Julia Kostka, Marlen Leisner, Nadine Lestmann, Katharina Mayr, Tiner Özcelik, Charlotte Pauli, Gergana Popova, Ines Rehbein, Nadja Reinhold, Franziska Rohland, Sören Schalowski, Kathleen Schumann, Kristina T. Sommer & Emiel Visser. 2010. *KiDKo: Ein Korpus spontaner Unterhaltungen unter Jugendlichen im multiethnischen und monoethnischen urbanen Raum*. <http://kiezdeutschkorpus.de/en/>.
- Wiese, Heike & Hans Georg Müller. 2018. The hidden life of V3: An overlooked word order variant on verb-second. In Mailin Antomo & Sonja Müller (eds.), *Non-canonical verb positioning in main clauses* (Linguistische Berichte Sonderheft 25), 202–223. Hamburg: Buske.
- Wiese, Heike & Ines Rehbein. 2016. Coherence in new urban dialects: A case study. *Lingua* 172–173. 45–61. DOI: [10.1016/j.lingua.2015.10.016](https://doi.org/10.1016/j.lingua.2015.10.016).
- Winkle, Claudia. 2015. *Non-canonical structures, they use them differently: Information packaging in spoken varieties of English*. Freiburg: Albert-Ludwigs-Universität Freiburg. (Dissertation). <https://freidok.uni-freiburg.de/data/10600>.
- Zerbian, Sabine, Kristina Barabashova & Yulia Zuban. 2025. *Left dislocations in monolingual and heritage Russian in Germany and the U.S.* DOI: <https://doi.org/10.1163/15507076-bja10035>.
- Zerbian, Sabine, Yulia Zuban, Marlene Böttcher & Oliver Bunk. 2025. Intonation in heritage languages. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 413–448. Berlin: Language Science Press. DOI: [10.5281/zenodo.15775183](https://doi.org/10.5281/zenodo.15775183).
- Zimmer, Christian, Heike Wiese, Horst J. Simon, Marianne Zappen-Thomson, Yannic Bracke, Britta Stuhl & Thomas Schmidt. 2020. Das Korpus Deutsch in Namibia (DNam): Eine Ressource für die Kontakt-, Variations- und Soziolinguistik. *Deutsche Sprache* 48(3). 210–232. <https://nbn-resolving.org/urn:nbn:de:bsz:mh39-100570> (27 April, 2023).
- Zuban, Yulia. 2023. *Unexpected use of referents in heritage Russian*. Poster presentation at Voices in Contexts, Cologne, Germany, October 12-13. <https://osf.io/xem4a/>.
- Zuban, Yulia. 2024. *Word order choice, expression of information status, and global intonation patterns in heritage Russian*. Stuttgart: Universität Stuttgart. (Doctoral dissertation).
- Zuban, Yulia, Maria Martynova, Sabine Zerbian, Luka Szucsich & Natalia Gagrina. 2021. Word order in heritage Russian: Clause type and majority language matter. *Russian Linguistics* 45(3). 253–281. DOI: [10.1007/s11185-021-09246-1](https://doi.org/10.1007/s11185-021-09246-1).

Zuban, Yulia, Tamara Rathcke & Sabine Zerbian. 2023. Do different majority languages lead to different intonational grammars? A case study of yes/no-questions in heritage Russian. *Heritage Language Journal* 20(1). 1–43. DOI: 10.1163/15507076-bja10016.

# Chapter 12

## Intonation in heritage languages

Sabine Zerbian<sup>a</sup>, Yulia Zuban<sup>a</sup>, Marlene Böttcher<sup>b</sup> & Oliver Bunk<sup>c</sup>

<sup>a</sup>University of Stuttgart <sup>b</sup>Kiel University <sup>c</sup>Humboldt-Universität zu Berlin

Intonation in heritage languages is a much less researched linguistic domain than, for example, morphosyntax. This chapter reports on several studies investigating the intonation of Russian and German bilingual speakers, based on the data from the RUEG corpus, in their heritage language and/or in the majority language. Section 3 addresses overall intonational features in (heritage) Russian, relating to phrasing and frequency of pitch accents. Section 4 reports findings on intonation used for various phenomena at linguistic interfaces, such as intonation in (heritage) Russian yes/no questions (Section 4.1), for contrastive adjective focus in (heritage) Russian (Section 4.2) and in discourse linking in heritage and majority German (Section 4.3). Section 5 addresses the perception of accents by (heritage) Russian listeners. Together, the studies reveal that heritage language speakers and monolingually-raised speakers show differences in prosodic realization and perception, although such differences are more quantitative than qualitative in nature.

### 1 Introduction and background

Heritage language speakers (HSs) are speakers who grew up in a bi- or multilingual home with a minority language in addition to the majority language (ML) dominant in the larger society (e.g., Rothman 2009, Polinsky 2018: 116). On the one hand, HSs are reported to be similar to monolingually-raised speakers with respect to the phonetics and phonology of their heritage language (HL), which has been termed “phonetic advantage” (Polinsky 2018: 116). On the other hand, studies have shown consistent differences in the phonetic realization of, for example, Voice Onset Time or vowels by HSs as compared to monolingually-raised



speakers (see Polinsky 2018, Montrul 2015 for overviews). These fine phonetic differences might contribute to the clearly discernible accent or heritage accent (Chang 2021) that has likewise been reported for HSs. It has been confirmed in various empirical studies that listeners rate the speech of HSs in their HL as more accented than monolinguals' speech (and less accented than L2 speech) (e.g., Kupisch et al. 2014 on Italian, French; Flores & Rato 2016 on European Portuguese) although it remains unclear what the perceived accent is due to.

In addition to differences in the segmental domain, it can be expected that there will also be differences in the suprasegmental domain relating to stress, rhythm and intonation, when comparing HSs with monolingually-raised speakers (see Montrul 2015, Polinsky 2018). And indeed, previous studies find that HSs produce suprasegmental features differently from both monolingually-raised speakers and L2 learners (e.g., Chang & Yao 2016 for Mandarin tone; Dehé 2018 for yes/no questions in North American heritage Icelandic). It is even suggested that suprasegmental features are more salient contributors to a heritage accent than segmental ones (Chang 2021: 601), thus further motivating the study of intonational aspects.

There are several reasons why we expect to find differences in the domain of intonation. On the one hand, intonation has been found to be fully acquired late in first language acquisition; thus, for a child who grows up with two languages simultaneously at a young age (as the HSs in our studies did), there is considerable room for cross-linguistic interactions in prosody, both from the ML to the HL and from the HL to the ML. Cross-linguistic influence might show itself along different dimensions: it might refer to the phonetic realization of identical tunes, the intonational inventory of the languages, the semantic meaning associated with a certain intonational event, the distribution of tunes, or the frequency of use (as detailed in the L2 Intonation Learning Theory model, LILt, by Mennen 2015).

On the other hand, intonation in languages like English, German or Russian is determined both by structural constraints including a nuclear pitch accent on the last constituent, as well as by pragmatic constraints, e.g., nuclear accent on the focused constituent, deaccentuation of given constituents, or final tunes determined by sentence type. Therefore, intonation is an interface phenomenon. The Interface Hypothesis (Sorace 2011) states that linguistic phenomena at the interface with external domains, such as pragmatics, are prone to change in language contact situations. A recent study by Feldhausen & Vanrell (2024) on the syntactic and prosodic realization of different focus types produced by HSs of Peninsular Spanish with German as ML gives support to the Interface Hypothesis by showing that HSs differ from monolingual speakers of Spanish at both the syntax-discourse and the phonology-discourse interface.

Finally, Polinsky describes somewhat opposing tendencies in heritage language phonology. On the one hand, various studies “converge on the observation that some contrasts present in the input undergo leveling” and, thus, a loss of distinction (Polinsky 2018: 125). On the other hand, based on data on final devoicing by HSs of Polish with ML English, Polinsky (2018: 134) describes a “tendency to overemphasize the sound properties of the heritage language that set it apart from the dominant language”. She further claims that this is “typical of heritage speakers who are more proficient in Polish”, as deduced from their higher rate of code-switching (cf. Poplack 1980 for code-switching as a sign of more balanced bilingualism). She suggests that it is “a general side effect of bilingualism” (Polinsky 2018: 136) that the differences between the languages are exaggerated. However, the nature of such “side effects” is unclear. Sociolinguistic perspectives might help in uncovering such effects; for example, some speakers might produce specific features more distinctly to align with the more privileged variety of a language (see Bunk et al. 2025 [this volume]).

We thus expect to find considerable variation in intonation produced by HSs in both their HL and ML. The variation might reveal new patterns along various dimensions of the intonation system. These might be attributable to factors such as cross-linguistic influence or general effects of bilingualism, for example, related to processing, fluency or decreased language exposure or use.

This chapter presents several studies that illustrate the spectrum of intonational variation that we have found in the data of our corpus of spoken language, focusing primarily on Russian HSs but also including some information on German HSs (see also Pashkova et al. 2025 [this volume] for intonational aspects of English as a majority language).<sup>1</sup> It presents examinations of overall intonational features, specifically prosodic phrasing and the inventory and distribution of pitch accents in (heritage) Russian. It also reports the results of investigations of intonational features in specific linguistic structures, zooming in on the prosodic realization of selected sentences, phrases or words in heritage and monolingual Russian as well as bilingual and monolingual German. Intonational patterns are compared across HSs in different countries and across formality. By

---

<sup>1</sup>Wandrei (2019) analysed features of segmental phonology in four HSs of Russian (ML German, female, 16–18 years old), using data from the RUEG corpus. The following phonological features were analyzed auditorily: palatalization, realization of <r>, allophonic realizations of /x/, final devoicing, voicing assimilation, and aspiration of voiceless plosives. The results show that salient differences emerge only in the realization of <r>, although not in all speakers (Wandrei 2019: 32). Another interesting observation was individual speakers producing single lexical items, such as “okej” with a German [o'ke:] instead of a Russian pronunciation [akej] (Wandrei 2019: 34).

including and presenting results from productions of Russian by monolingually-raised speakers, our studies also make interesting empirical contributions for this speaker group. In the case of German as an ML, the chapter presents results from a study of phrasing in the left sentence periphery.

## 2 Methodology

All research reported in this study is based on data in the RUEG corpus. The relevant subcorpora and/or additional data are further specified in their respective sections.

The RUEG corpus (Lüdeling et al. 2024) was collected and annotated by the collaborative Research Unit *Emerging Grammars* funded by the German Science Foundation from 2018–2024. Naturalistic data were elicited by means of presenting participants with a video clip depicting a fictional car accident. Participants were asked to describe the accident, imagining themselves in different communicative situations which covered different modes (spoken/written) and formality (formal/informal). In the formal communicative situation, participants were asked to give an account of what they saw to the police. In the informal communicative situation, participants were asked to leave a message for a friend using WhatsApp (for details on methodology, see Wiese 2020). Data were elicited from HSs of Turkish, Greek, and Russian in Germany and the US, from HSs of German in the US, and from monolingually-raised speakers in Turkey, Greece, Russia, Germany, and the US. For the studies on intonation reported in this chapter, only the spoken Russian data are considered as well as data from heritage and monolingually-raised speakers of German.

### 2.1 Russian: Data and participants

The spoken Russian data can be accessed as the Russian Prosody subcorpus (RuPro).<sup>2</sup> It contains spoken Russian data of about 25k word tokens by a total of 40 monolingually-raised Russian speakers and 53 bilingual Russian HSs in the United States. The speakers fall into two age groups: adolescents (14–18 years; 20 mono, 22 bilingual) and adults (22–35 years; 20 mono, 31 bilingual). The corpus data of each participant are enriched with extensive metadata on their language background in both HL and ML (e.g., beginning of language acquisition, media use, language use with parents, self-assessment of language skills such as understanding, speaking, reading, and writing), socio-economic status (e.g., highest degree for adult participants; parents' profession for adolescent participants),

---

<sup>2</sup><https://korpling.german.hu-berlin.de/annis/>

and personality traits (along five dimensions: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience; Gosling et al. 2003).

All data were transcribed, normalized, and annotated for communication units (CU; Hughes et al. 1997) and part of speech. Three prosodic layers were added to the data of the spoken formal and informal modes. The prosodic layers show annotations for pitch accent placement on the word level, pitch accent type, and Intonation Phrase (IP) boundaries. Pitch accent placement and pitch accent type were determined auditorily and phonetically, based on examination of the F0 (low and high turning points). IP boundaries were annotated following the guidelines stated in Himmelmann et al. (2018). Six types of pitch accents (henceforth PA; L\*, H\*, L\*+H, L+H\*, H+L\* and H\*+L) and two final boundary tones (henceforth BT; L%, H%) emerged as relevant in these data. Sample contours of the three most frequent PAs are provided in Figures 1 and 2. An illustration of all PA labels can be found in Zerbian et al. (2024).

All H tones could be additionally upstepped (pitch range expansion compared to a preceding high tone) or downstepped (pitch range compression). Praat was used for phonetic annotation of the audio files (Boersma 2001).

Concerning exposure and active language use, HSs had a high involvement with Russian in their everyday life according to self-assessments (e.g., watching TV programs in Russian, writing in Russian-speaking social media; almost every HS could write and read Russian Cyrillic script) (for more details see Zerbian et al. 2025, Zuban 2024).

Studies consider proficiency in bilingual populations to be a multidimensional phenomenon that includes many different aspects (e.g., type-token ratio, ratio of main and subordinate clauses, frequency of word-internal code switching and repetitions, word length, speech rate, and number of empty and filled pauses; see Petersen et al. 2021, Montrul 2015 on HSs; see Iwashita et al. 2008, Christoffersen 2017 on L2 learners). We concentrated on fluency and quantified HSs' fluency by measuring their speech rate and the number of filled pauses. Speech rate was calculated as number of words per minute (as done in other studies on heritage Russian, e.g., Polinsky 2008, Laleko & Dubinina 2018). Mean speech rate was comparable between HSs (114 words/min) and monolingually-raised speakers (119 words/min) (Zuban 2024). This shows an overall high fluency of the HSs in their HL. (In)formality was an important factor, though: HSs showed a similar speech rate in both formal and informal communicative situations while monolingually-raised speakers spoke significantly faster in the informal situation compared to the formal one, maybe due to lower habitual ease in the latter. As a result, HSs

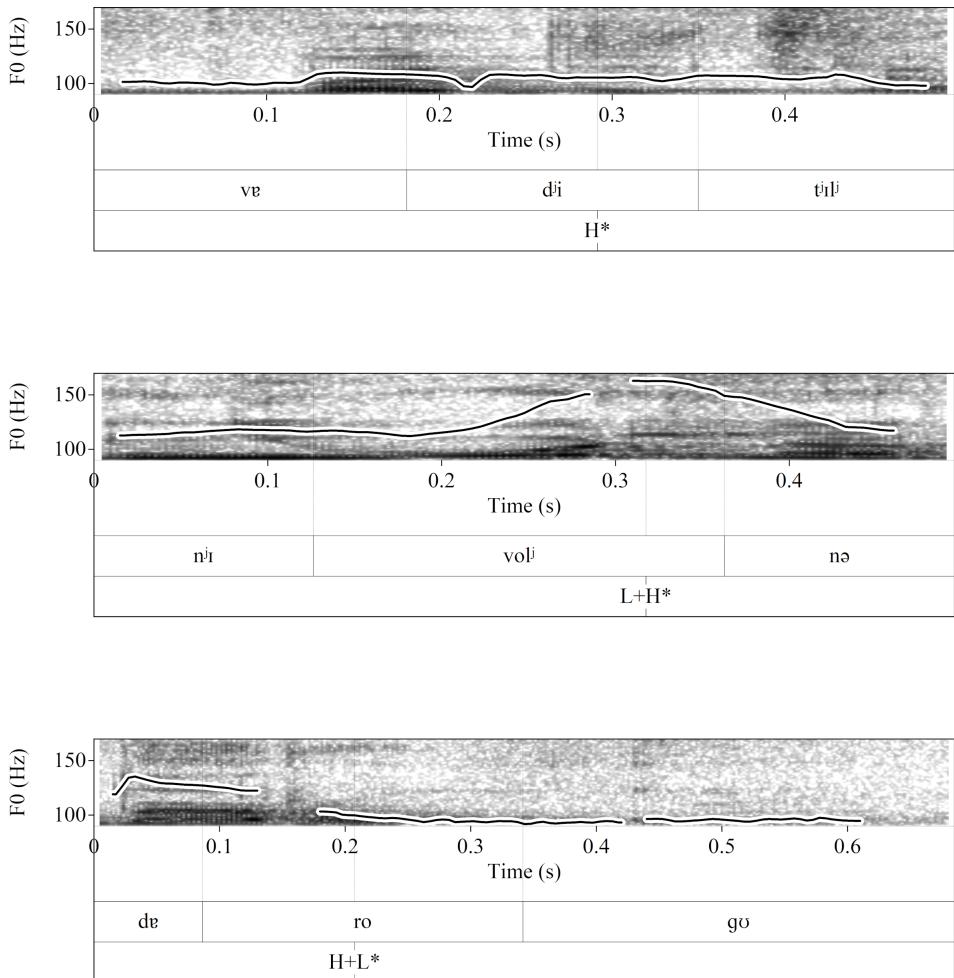


Figure 1: Sample contours of the three most frequent pitch accents in Russian (H<sup>\*</sup>, L+H<sup>\*</sup>, H+L<sup>\*</sup>)

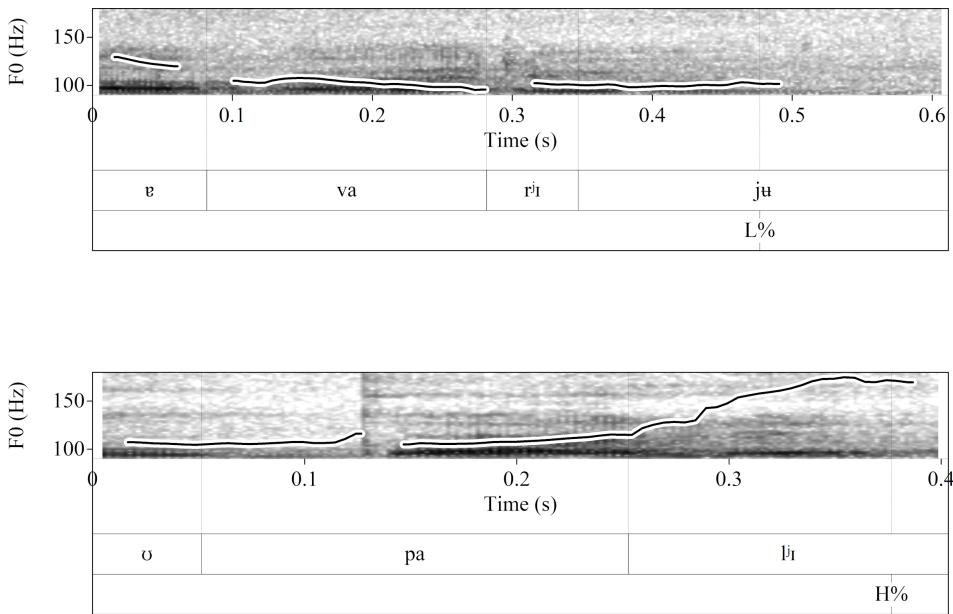


Figure 2: Sample contours of L% and H% in yes/no questions in (heritage) Russian (see Section 4.1)

and monolingually-raised speakers were similar to each other in the formal situations, but not in the informal ones.

Another indicator for fluency is the use of filled pauses like *um* in English (e.g., de Jong 2016). Böttcher & Zellers (2024) looked at the frequency of filled pauses in the RUEG corpus (version 1.0\_SNAPSHOT) including both heritage Russian in the US and Germany and monolingual Russian, as well as majority English and majority German. Statistical analyses reveal significantly more filled pauses in formal narrations and in the speech of bilingual speakers irrespective of their language. This indicates similarities in another aspect of fluency between HSs and monolingually raised speakers in formal situations, but not in informal situations.

The results concerning fluency, considering both speech rate and use of filled pauses, can be related to aspects of higher speech planning costs required in monitoring two languages and in formal situations. The monitoring of two languages and suppressing one in monolingual mode has been reported to be cognitively demanding (Kroll & Gollan 2014). Also, speech planning in formal contexts may take more effort, with higher pressure on language form and meaning, and might also include higher interpersonal insecurity compared to more intimate, informal situations (Tottie 2014, Staley & Jucker 2021 based on the use of filler particles).

## **2.2 German: Data and participants**

The spoken German data come from the German subcorpus of the RUEG corpus (RUEG-DE), and from a corpus on heritage Low German in the US (Rocker 2022). RUEG-DE (version 1.0-SNAPSHOT) contains a total of 156k word tokens by monolingually-raised German speakers ( $n = 64$ ; 32 adults, 32 adolescents), bilingual German speakers with the HLs Turkish ( $n = 65$ ; 33 adults, 32 adolescents), Russian ( $n = 59$ ; 30 adults, 29 adolescents), or Greek ( $n = 44$ ; 26 adults, 18 adolescents), and heritage German speakers in the US ( $n = 36$ ; 7 adults, 29 adolescents). All data were transcribed, annotated morphologically for part of speech, lemma, and function, and annotated syntactically for topological fields and syntactic dependencies. The topological field model analyzes German sentence structure into different fields that are occupied by specific elements (see, e.g., Drach 1937, Höhle 1986). Topological fields were annotated based on the annotation guidelines used for the Kiezdeutschkorpus (Wiese et al. 2010), which allow annotation of specific noncanonical phenomena in the left sentence periphery such as left dislocation constructions, hanging topics, and constituents such as adverbials that are followed by a non-verbal constituent. The corpus by Rocker (2022) on heritage low German in the US comprises data from 46 speakers and a total of 58 sociolinguistic interviews. The relevant data for the analyses were extracted from the corpora and prosodically annotated for phrase boundaries in order to analyze phrasing in the left periphery (see Section 4.3 for a detailed description).

## **3 Overall intonational features: Intonational phrases and pitch accents**

Heritage speakers and monolingually-raised speakers may differ with respect to overall intonational features; in other words, intonational features observed across entire narrations as opposed to specific linguistic constructions. Here we focus our research on two areas. First, we compare the intonation of HSs and monolingually-raised speakers on the frequency dimension (Mennen 2015), investigating whether intonational events such as pitch accents (PAs) are used equally frequently. Second, we asked whether prosodic units such as Intonational Phrases (IPs), which are defined by intonational contours, are similar in size in comparable narrations.

In an earlier case study, Comstock (2018) investigated intonation patterns of one HS of Russian in the US and two monolingual Russian-speaking journalists in two types of political interviews: affiliative and antagonistic. The affiliative

interviews were those that did not involve any controversy, but instead included topics that helped an interviewer establish friendly relationships with an interviewee (e.g., personal opinions on whether it is easy for foreign journalists to work in Russia). Comstock (2018: 268) classifies affiliative interviews as informal in nature since they resemble a conversation rather than a strict question-answer format. The antagonistic interviews had an element of conflict (e.g., the HS was asked to comment on a US presidential candidate who claimed that Russia was America's main enemy), and are therefore viewed as more formal by Comstock (2018: 268).

Comstock (2018) found that the HS differed from the two monolingual speakers of Russian in both interviews regarding the frequency of some intonational patterns. For instance, the monolingual speakers of Russian did not produce any monotonal PAs, while the HS produced on average 5.9 monotonal accents per IP in the affiliative interview and 5.2 monotonal accents per IP in the antagonistic interview (Comstock 2018: 220; 231; 248; 256). Furthermore, some prosodic patterns of monolingual interviewers were found to differ depending on the interview (e.g., the number of single-word IPs or the number of falling H+L\* nuclear accents) while prosodic patterns of the HS only differed in one aspect across the two interviews; namely, the number of fronted constituents was greater in the affiliative interview than in the antagonistic one (Comstock 2018: 267). Thus, this case study showed that the monolingual speakers and the HS differed in their intonation, and that type of interview mattered more for the intonation of monolingual speakers than for the intonation of the HS.

Similar to the study by Comstock (2018), we explored intonational differences between monolingual speakers and HSs of Russian, using the RuPro subcorpus. We explored differences along the systemic (e.g., inventory of tonal events) and frequency dimensions by extracting labels and intervals pertaining to intonational events (pitch accents (PA) and boundary tones (BT)) and units (Intonation Phrases (IP)) from the prosodic annotation of the data. Detailed descriptions are available in Zerbian et al. (2024) and Zuban (2024).

Regarding the inventory of PAs and BTs, no systemic difference emerged between HSs and monolingually-raised speakers as in both speaker groups the same kinds of PAs and BTs were attested in overall similar frequency. Here we report the group results for size of IPs and number of PAs to illustrate the difference in patterns of overall intonational features between the two groups.

To determine the size of intonational phrases, the overall number of IPs was extracted and divided by the total numbers of words to account for differences in the overall length of narrations. As shown in Table 1, HSs produce overall shorter IPs (i.e., fewer words per IP) than monolingually-raised speakers. A linear

mixed model was fit to the data and revealed a significant effect of formality ( $z = 5.97, p < 0.001$ ) and a significant interaction of formality and speaker group ( $z = -3.52, p < 0.001$ ). A post-hoc pairwise comparison showed that monolingual speakers produced shorter IPs in formal situations than in informal situations ( $p < 0.001$ ), and monolinguals differed from HSs by producing longer IPs in informal situations ( $p < 0.001$ ). HSs did not show any effect of formality on IP length. Thus, formality emerged as an important factor for IP length in monolingual speakers but not in HSs, in line with the results by Comstock (2018).

The number of PAs was extracted and divided by the number of content words on which a PA could be placed. A binomial generalized linear mixed effects model revealed that on average HSs produced more PAs on content words than monolingual speakers ( $z = 6.53, p < 0.001$ ), as also shown in Table 1.

Table 1: Overall intonational features in RuPro

|                    | words/IP | PAs/content words |
|--------------------|----------|-------------------|
| HSs ( $N = 53$ )   | 2.46     | 0.85              |
| formal             | 2.42     | 0.85              |
| informal           | 2.51     | 0.84              |
| Monos ( $N = 40$ ) | 2.77     | 0.77              |
| formal             | 2.55     | 0.77              |
| informal           | 3.00     | 0.75              |

The architecture of the RUEG corpus also allows for testing of whether age and gender are relevant factors, as they are reported to be crucial in language variation (Tagliamonte 2016, Nagy 2017). Zuban (2024) explored whether PA frequency was influenced by the factors age or gender in a reduced prosodically-annotated data set (40 HSs, 40 mono) that was later included in the RuPro corpus. However, age and gender were not found to be relevant for the number of PAs. Interestingly, Dehé & Rommel (2024) also did not find age to be a relevant factor in Heritage Icelandic (question) intonation.

To sum up, the results show that formality influences PA frequency in productions of monolingually-raised speakers, but not HSs. Further, HSs of Russian do not differentiate IP length according to formality. In informal situations, however, they produce IPs containing fewer words than monolingual speakers. At the same time, they produce more PAs on content words than monolingual speakers. These results are in line with previous studies that report the lack of register differentiation in HSs, possibly caused by the absence of formal instruction in the

HL (e.g., Schroeder et al. forthcoming, Wiese et al. 2022, Alexiadou et al. 2022, Comstock 2018). More frequent PAs produced by HSs have also been reported in studies by Zuban et al. (2023) and Zerbian et al. (2022) (presented in Sections 4.1 and 4.2), which have been interpreted as a general effect of bilingualism.

Together, the distribution of these intonational features might point to a different overall rhythm in the language. The result that the two speaker groups only differ in pitch accent frequency in the informal communicative situation is also in line with the fluency results by Böttcher & Zellers (2024). Thus, this effect of bilingualism can be linked to increased processing costs in bilingual speakers when monitoring two languages.

Similarly, the results for monolingual speakers can be connected to higher processing costs in formal situations with lower habitual ease, such as when speaking to a policeman (see Section 2 on methodology). A police report might be more detailed, require more careful choice of words, and require a longer planning period than a story presented to a friend, all leading to slower speech and more frequent accentuation in monolingual speakers.

## 4 Intonation at the interfaces

The intonation systems of HSs and monolingually-raised speakers can also be compared along the semantic dimension; in other words, how intonational events are used in different semantic/pragmatic contexts. Examples are intonational patterns used for the differentiation of sentence types (e.g., declarative versus question), or the expression of information structure or information status (e.g., PA for focus, deaccentuation for given information). Several studies investigate specific aspects of heritage language intonation, such as intonation in declaratives (e.g., Robles-Puente 2014, Colantoni et al. 2016), in questions (e.g., Dehé 2018), for focus (e.g., Gries & Miglio 2014, van Rijswyk et al. 2017, Kim 2019), or for a combination of those (e.g., Rao 2016).

In this section we will present the results of our studies on intonation used in yes/no questions (Section 4.1), contrastive focus (Section 4.2) and discourse linking (Section 4.3). Zuban (2024) investigates the syntax and intonation of referents in the order new-given produced by Russian HSs and monolingual speakers of Russian from the same corpus.

### 4.1 Intonation of yes/no questions

Zuban et al. (2023) looked at the prosodic realization of yes/no questions in Russian HSs in Germany and the US and monolingually-raised Russian speakers.

Yes/no questions are questions that require as the answer either “yes” or “no” (henceforth abbreviated YNQ). In Standard monolingual Russian, a frequent way of forming a YNQ is through intonation; in other words, declarative clauses and YNQs do not differ in their morphosyntax, but only in their prosodic realization (Bryzgunova 1980, Svetozarova 1998, Rathcke 2006). In a YNQ, the verb typically receives the main prominence, as opposed to the sentence-final argument in declaratives.

Moreover, the main accent is usually falling in declaratives and rising in YNQs (bitonal L<sup>\*</sup>+H or L+H<sup>\*</sup>; Rathcke 2006, Meyer & Mleinek 2006: 1619, Makarova 2003: 93). Constituents other than the verb are usually not accented (Rathcke 2006). The rising accent on the verb in Russian YNQs is followed by a low boundary tone at the end of the question if the nuclear accent is not realized on the utterance-final syllable.

In YNQs in English and German, however, the main prominence (monotonal L<sup>\*</sup>) remains on the sentence-final argument and the YNQs end in a rising boundary tone. Investigating heritage Russian in the US and in Germany therefore provides an ideal situation to test the influence of the dominant language. More specifically, Zuban et al. (2023) explored three questions: (1) how YNQs are produced by HSs of Russian (both in the US and in Germany) as compared to monolingual speakers in terms of location of accent placement (on subject, verb and object), pitch accent type (monotonal versus bitonal) and the choice of final boundary tone (L% versus H%); (2) whether speakers of heritage Russian show similar patterns concerning PA location, PA type and BT across different MLs; and (3) if the observed patterns can be attributed to a direct influence from the ML, such as through transfer, or if they should be characterized as innovations.

In their study, 20 adolescents in each of three groups took part: speakers of Russian in Russia (10 female, mean age=17.0, SD=0.58), HSs of Russian in the US (10 female, mean age=16.2, SD=1.53) and HSs of Russian in Germany (11 female, mean age=16.9, SD=0.87). These participants also provided data to the RUEG corpus, although the data for this study were collected in addition to the corpus data. The two populations of HSs share some characteristics: they are 2nd generation speakers who were either born in the US or Germany, respectively, or came there by the age of five. Furthermore, both countries have a large number of Russian-speaking immigrants (United States Census Bureau 2009, Statistisches Bundesamt 2022). Both groups of HSs were comparable in their self-assessed involvement with Russian.

Their task was to read out a list of scripted questions consisting of ten YNQs in total, divided into five transitive questions (SVO) and five intransitive questions (SV). All questions referred to the events of the fictional car accident that was

shown in the video that the participants had watched prior to the recording of the experimental sentences. Across the 60 participants, there were 600 YNQs.

All productions were annotated for PA location, PA type, and BT, using surface oriented phonetic labels. The results show that all speaker groups produce the most prominent PA on the verb (for a detailed description of the results including statistical analyses, see Zuban et al. 2023). Beyond that, HSs in both countries produce significantly more PAs on subjects and objects compared to monolingual speakers, although monolingual speakers do not fully deaccentuate other constituents as might have been expected based on the literature.

Regarding the PA type, all speaker groups have the same preference in using a monotonal H\* on subjects. Whereas monolingually-raised speakers also frequently use a L\* on subjects, HSs of Russian use H\* nearly exclusively, thus showing less variation of PA type.

For the PA on the verb, it is noteworthy that speakers of all groups use rising bitonal accents (L\*+H, L+H\*), with the alignment being partly determined by the segmental material available for the realization of the tones.

As for the use of boundary tones, the results show that both monolingual speakers and HSs predominantly produce salient low BTs in Russian YNQs. This is particularly noteworthy for the groups of HSs, for whom we had expected to find more final rises due to a possible influence of language contact. However, HSs in Germany and monolingual speakers produce high BTs (H%) mainly due to phonetic pressure; that is, when there is less segmental material available for the realization of a low BT following a rising nuclear pitch accent.

It is interesting to note that a group-level difference also emerges in another case of phonetic pressure, namely in the choice of the PA on the verb in SV sentences. Here, HSs in Germany group together with monolingual speakers, whereas HSs in the US show a slightly different pattern, as shown in Figure 3.

Overall, the results show that HSs produce YNQs in a way comparable to monolinguals in terms of the location and type of a PA on the verb, and also of the use of BTs. Influence from the dominant ML does not manifest itself directly. Nevertheless, HSs of both groups differ from monolinguals in the number of additional PAs on subjects and objects. Furthermore, HSs in the US differ from HSs in Germany and monolingually-raised speakers regarding tonal preferences and the handling of phonetic pressure when little segmental material is available. One important implication of these results is the existence of group-specific strategies to handle phonetic pressure. The prosodic strategies used to resolve phonetic pressure used by HSs in the US cannot be easily attributed to transfer from their ML; rather, these strategies should be considered innovations.

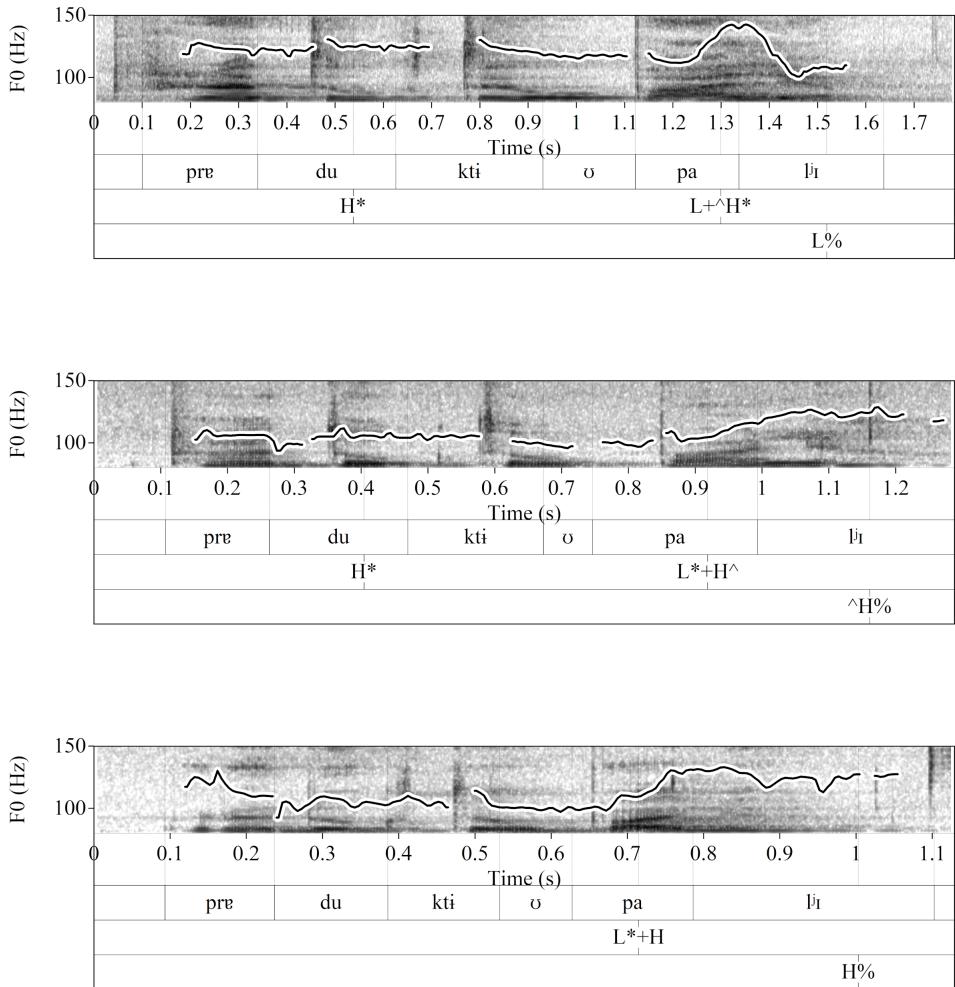


Figure 3: F0-contours of the SV sentence “Продукты упали?” (/pre'dukti v'spal'i/ ‘Did the groceries fall down?’), produced by a HS in the US (top panel); a HS in Germany (middle panel); and a monolingual speaker (bottom panel)

## 4.2 Contrastive focus

Zerbian et al. (2022) investigated the prosodic marking of contrastively focused adjectives in heritage Russian spoken in the US, using data from the RUEG corpus. The prosodic expression of focus and contrast within a noun phrase is a particularly interesting domain because languages differ in their prosodic marking of information structure not only at the sentence level but also at the level of noun phrases (e.g., Ladd 2008).

English and Russian share the option of prosodically marking contrastive adjectives through pitch accents. In a modified NP, such as *a blue car*, a PA on the noun (indicated by small capitals), as in (1a), can indicate a focus (indicated by brackets and subscript  $F$ ) only on the noun or on the entire NP. A PA on both adjective and noun, as in (1b), can indicate separate focus on the noun and on the adjective, focus on the NP, or focus only on the noun (with a prenuclear accent on the adjective in the latter two cases; Ladd 2008: 8). However, an accent on the adjective (and co-occurring deaccentuation of the noun), as in (1c), can only indicate narrow focus on the adjective. For Russian, the parallel has been reported as illustrated in (2) (King 1993: 57, Jasinskaja 2016: 4–5).

- (1) Prosody and interpretation in modified NPs in English
  - a. a white CAR = [a white car] $_F$ /a white [car] $_F$
  - b. a WHITE CAR = a [white] $_F$  [car] $_F$
  - c. a WHITE car = a [white] $_F$  car
- (2) Prosody and interpretation in modified NPs in Russian
  - a. krasnuju ZVEzdočku = [red star] $_F$ /red [star] $_F$   
red star
  - b. KRASnuju ZVEzdočku = [red] $_F$  [star] $_F$
  - c. KRASnuju zvezdočku = [red] $_F$  star

In addition to this similarity in prosodic focus marking shown in examples (1) and (2), English and Russian differ in the availability of additional syntactic means to indicate contrast. One example is the split construction in Russian, illustrated in (3b), in which a modifying adjective is separated from the head noun (Sekerina & Trueswell 2011: 281).

- (3) Split construction in Russian (from Sekerina & Trueswell 2011: 282)
  - a. Položite KRASnuju zvezdočku v Poziciju 4.  
put red.ACC.FEM star.ACC.FEM in position 4  
'Put the *red* star in position 4.'

- b. KRAŠnuju položite zvezdočku v Poziciju 4.  
red.ACC.FEM put star.ACC.FEM in position 4

Thus, the research question was how bilingual speakers of heritage Russian in the US produce NPs with a contrastively focused adjective in their two languages, namely heritage Russian and majority English. The data in Zerbian et al. (2022) form a subset of the RUEG corpus (Lüdeling et al. 2024), accessible as “RUEG-EnPro”, which contains the data from the spoken tasks of the elicitation sessions provided by 36 English monolingual speakers (24 adolescents, 12 adults), 40 Russian monolingual speakers (20 adolescents, 20 adults), 60 HSs of Russian in their ML English (30 adolescents, 30 adults) and 53 HSs of Russian in their HL Russian (31 adolescents, 22 adults). All bilingual Russian speakers were either born in the US or moved there before the age of 5. The data were transcribed and annotated for PA location and type.<sup>3</sup>

Because the car accident prompted a contrast between two differently coloured cars, a lexical search was performed on the corpus for the noun *car* immediately preceded by an adjective. The resulting hits were annotated manually for contrastive focus (Götze et al. 2007). The phrases containing contrastively focused adjectives (henceforth Adj<sub>CF</sub>+N) were extracted from the corpus and were manually annotated for presence/absence of PAs on their constituents.

The results concerning the frequency of occurrence revealed that fewer monolingual speakers of Russian produced Adj<sub>CF</sub>+N than HSs of Russian in their ML and HL and monolingual speakers English (for the detailed results, see Zerbian et al. 2022). Also, those monolingual speakers of Russian who did produce it, produced fewer instances than HSs of Russian in their ML and HL and monolingual speakers of English. Moreover, there seems to be a language-specific difference in that monolingual speakers and HSs of Russian produced Adj<sub>CF</sub>+N less frequently than HSs in their ML and monolingual English speakers. Across all four speaker groups, Adj<sub>CF</sub>+N is produced more frequently in the formal communicative situation than in the informal situation.

As for accentuation patterns in Adj<sub>CF</sub>+N, the results show that in the English data, the expected preference for a single PA on the adjective prevails across communicative situations and speaker groups. It can also be noted that bilingual speakers produce double accents (accent on both adjective and noun) more often than monolingual speakers. Across the mono- and bilingual speaker groups, there is a higher percentage of double accents in the formal communicative situation as compared to the informal situation.

---

<sup>3</sup><https://korpling.german.hu-berlin.de/annis/>

In the Russian data, the preference for a single PA on the adjective that would have been expected based on the literature only holds for monolingually-raised speakers in the informal situation (shown in Figure 5). HSs show a preference for double accents in both formal and informal communicative situations, exemplified in Figure 4. Double accents also seem to prevail in monolingually-raised speakers in the formal situation, although the number of monolingual speakers is too low for any strong claims. An increased use of double accents can be seen in the formal situation (only monolinguals) and overall in bilingual HSs, similar to the English data.

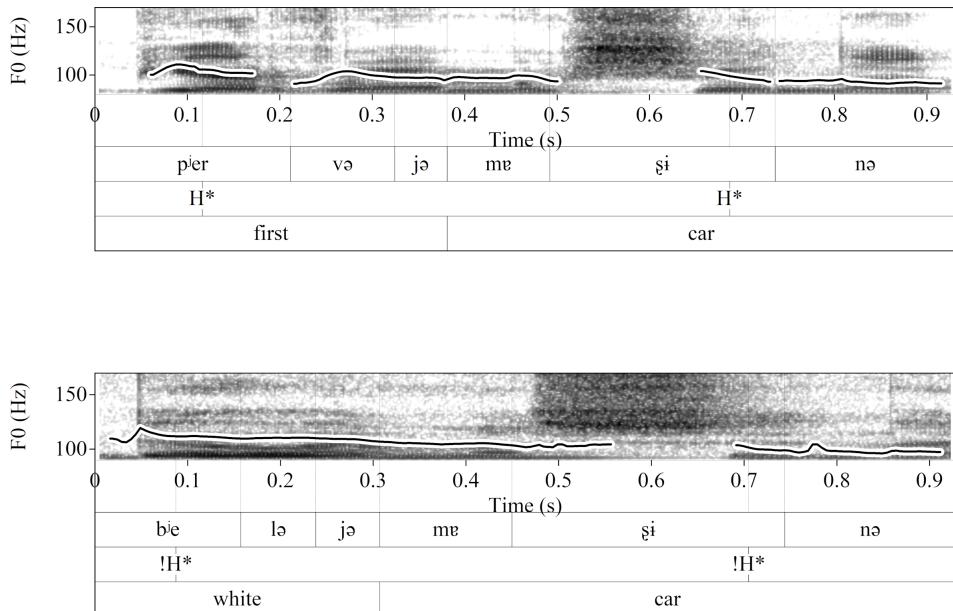


Figure 4: F0-contours of the contrastively focused adjectives “*первая*” (/p<sup>j</sup>ervəjə/ “first.NOM”) and “*белая*” (/b<sup>j</sup>eləjə/ “white.NOM”), produced by HSs in formal (top panel) and informal situations (bottom panel)

To sum up, the investigation of the corpus data reveals that Russian speakers (both heritage and monolingual) use the structure  $\text{Adj}_{\text{CF}} + \text{N}$  less frequently than English speakers (both mono- and bilingual), despite the reported parallel in terms of semantics and prosody. Also, English and Russian speakers differ in their accentuation pattern in  $\text{Adj}_{\text{CF}} + \text{N}$ . Russian HSs frequently use double accents in  $\text{Adj}_{\text{CF}} + \text{N}$ , a pattern that indicates a different information structure (see example 1b). Across English and Russian, double accents in  $\text{Adj}_{\text{CF}} + \text{N}$  occur more

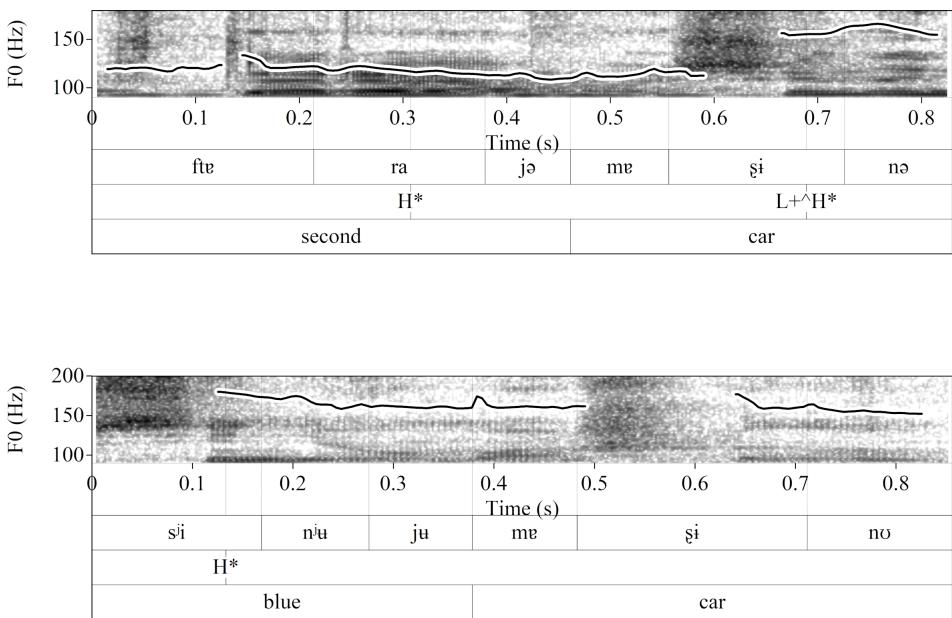


Figure 5: F0-contours of the contrastively focused adjectives “вторая” (/fte'rajə/ “second.NOM”) and “синюю” (/s̥in'juː/ “blue.ACC”), produced by monolingual speakers in formal (top panel) and informal situations (bottom panel)

frequently in formal than in informal situations, and more frequently in bilingual than in monolingual speakers.

One reason for why Adj<sub>CF</sub>+N are used less frequently in Russian than in English might lie in syntax: whereas English has a rigid SVO word order, Russian allows for word order changes for the expression of information structure, such as inversion, dislocation, and split constructions as in Example (3) (e.g., Jasinskaja 2016). Although we did not explicitly check word order for these constructions, other studies on word order in the RUEG corpus have attested word order variations in HSs of Russian, such as left dislocation (Zerbian et al. 2025; see also Bunk et al. 2025 [this volume]), inversion (OVS) and dislocation (SOV) (Zuban et al. 2021). Moreover, in a processing study, Sekerina & Trueswell (2011: 296) found prosody to be a weak cue to contrastiveness in Russian. In a production study, Kim (2019) also found prosody to be a weak cue to focus in the Spanish of monolingual speakers, but used more by HSs of Spanish in the US.

The increased use of double accents in bilinguals’ heritage Russian and in their English (though to a lesser extent) could again be interpreted as relating to increased processing costs when using two languages (see Section 3). However,

interpreting double accents as a general feature of bilingualism makes the prediction that we would find an increased use across all bilingual speaker groups. A comparative look at the other language constellations in the RUEG corpus, however, shows that the increased use of (double) accents is not found in all bilingual groups. For example, Böttcher (2021, 2022) investigated the prosody of contrastive adjectives in majority English by HSs of Russian and Greek. Whereas the use of double accents was frequently found in the English productions of Russian HSs, this was not the case for Greek HSs. Thus, the frequent production of double accents in our study cannot be related to transfer nor to bilingualism in general, but might be a group-specific pattern.

### 4.3 Intonation and discourse markers

The data collected in the RUEG corpus also allowed for an explorative investigation of the interplay of prosody, syntax, and discourse linking. Intonation can play an important and disambiguating role in discourse linking. The data from the RUEG project provide some first insights from the realization of V3 structures in mono- and bilingual German.

V3 sentences display an adverbial-subject-finite verb order and thus violate the V2 constraint (for a more detailed description of the pattern, see Bunk et al. 2025 [this volume]). An example is given in example (4).

- (4) V3 in German (DEbi58MT\_is)  
 und dann er lässt sein Ball einfach fallen  
 and then he lets his ball just fall  
 ‘and then he just drops his ball’

The structures have been described for various contact varieties in north western Europe (e.g., Germany, Sweden, Norway, Denmark, the Netherlands; see Wiese 2009, Kotsinas 1992, Opsahl 2009, Quist 2000, and Meelen et al. 2020, respectively) and HLs across the globe (e.g., US, Namibia; see Tracy & Lattey 2010, Wiese & Müller 2018). V3 is also reported for monolingual German; however, multilinguals appear to make use of V3 more frequently (Wiese & Müller 2018).

On the level of information structure, Wiese & Rehbein (2016) identify an order of frame-setter > topic > comment. Schalowski (2017) argues that the initial adverbial can be either frame-setting (as in 4) or discourse connecting (as in 5) while Bunk (2020) claims it can have both functions at the same time (as in 6). The ambiguous nature of the initial adverbial makes V3 structures particularly interesting to look at from the perspective of prosody, as prosody might give the

interlocutor additional cues as to the interpretation of the initial adverbial. In one of our studies we investigated the prosodic realization of V3 sentences in German and the role of prosody in the disambiguation of information structure (Bunk & Rocker in preparation).

- (5) Discourse connecting (DEbi18MT\_isD)  
auf jeden Fall ich bin dann einfach weggegangen  
in any case I am then just away.walked  
'Anyway I just walked away'
- (6) Frame-setting and discourse connecting (DEbi25MT\_is)  
und denn der eine war ziemlich schnell  
and then the one was pretty quick  
'And then one of them was pretty quick'

Previous studies have extensively investigated prosodic properties of discourse markers (DMs) and come to different conclusions. While a number of studies highlight that DMs are prosodically and syntactically not integrated into the following utterance (Jucker & Ziv 1998, Auer & Günthner 2005, Brinton 2010), others argue that there is no evidence that DMs have specific prosodic features, so they might be prosodically integrated or not (Imo 2012). Another strand of research indicates that prosodic integration differs between items and is related to the degree of grammaticalization of these items. Particular lexical items lose parts of their meaning as they grammaticalize/pragmatalize (see Traugott 1995, Auer & Günthner 2005). This development goes hand in hand with a loss of prosodic prominence, in particular stress (see Bybee et al. 1991, Hirschberg & Litman 1993, Wichmann 2011, Dehé & Stathi 2016). Wichmann et al. (2010) show that *of course* is syntactically more integrated into the following syntagma when it is used as a discourse marker, which is characterized by the lack of stress.

Only a few studies so far investigate the prosodic realization of V3 sentences. However, it seems that prosody can give us hints with respect to the status of the initial adverbial in V3 sentences. In addition, it might help us to better understand the mechanisms of grammaticalization in German in general, as V3 adverbials appear to spotlight this phenomenon. The preverbal area in V3 predominantly entails prosodically light material and does not bear accent (see Freywald et al. 2015, te Velde 2017b). Te Velde (2017a) states that the preverbal area is realized in one prosodic phrase, while Breitbarth (2022) finds that the vast majority of V3 structures exhibit a prosodic boundary between the initial constituent and the subject. However, both studies look at very specific lexical items. Te Velde (2017a)

restricts his analysis to temporal adverbials and Breitbarth (2022) excludes discourse organizing *dann* ‘then’, even though *dann* has been described to be particularly frequent in V3 sentences (Wiese & Müller 2018). Hence, the studies to date do not distinguish between frame-setting and discourse linking adverbials in V3 with respect to prosodic realization.

The empirical basis of our study were two corpora as detailed in Section 2.2: the German subcorpus of the RUEG corpus (Wiese et al. 2021), i.e., monolingual and bilingual speakers in Germany and bilinguals in the US, and the corpus by Rocker (2022), comprising data of heritage Low German in the US. The corpora served the purpose of including different contact scenarios in order to investigate V3 under different conditions of language contact.

The data were annotated for prosodic boundaries based on Himmelmann et al. (2018) and Breitbarth (2022). The presence of a boundary indicated prosodic separation while the absence indicated prosodic integration. Based on the observations concerning prosodic integration and function described above (i.e., the higher the level of prosodic integration, the more likely the discourse marker functions of the adverbial), we expected V3 adverbials to be more integrated if they were used as discourse markers (= no prosodic boundary) and less integrated if they were used as frame-setters (= prosodic boundary).

Even though figures were rather low (79 V3 sentences in RUEG, 180 V3 sentences in heritage Low German), Bunk & Rocker (in preparation) found the following tendencies that need further investigation with a larger data set or experiments. Overall, they found that the German speakers produced V3 with and without boundaries in all corpora. However, the vast majority of the structures did not have a prosodic boundary. If boundaries occurred, they tended to occur after frame-setting adverbials, not discourse connective adverbials. There were only very few cases where discourse connective adverbials were followed by a prosodic boundary. This was particularly true for the most frequent adverbial *dann* ‘then’. This result points to the development of *dann* ‘then’ into a discourse marker, as highlighted by prosodic integration. The patterns were similar across all groups; however, monolingual speakers in Germany used V3 less frequently than the other groups. HSs of German in the US almost always used *und dann* ‘and then’ instead of *dann* ‘then’ in discourse marking function, indicating that these two elements form a unit that is schematized and conceptualized as a grammatical or discourse-pragmatic pattern (see Haspelmath 2002). V3 structures in heritage Low German more often had a prosodic boundary. However, the majority of V3 sentences with initial (*und*) *dann* ‘(and) then’ were again produced without a prosodic boundary. In sum, the study showed that the default case

in our data seems to be prosodic integration (i.e., lack of a boundary) in V3 sentences, even though V3 is also produced with a boundary. These findings indicate that V3 adverbials are at different stages on a grammaticalization path. Particular items, such as (*und*) *dann* ‘(and) then’ are more grammaticalized than others, such as *auf einmal* ‘suddenly’. When boundaries were used, they marked frame-setting adverbials.

## 5 Perception of intonation by heritage speakers

Whereas production is often investigated in heritage language phonology, perception or comprehension, especially of prosody, is researched less often (Polinsky & Scontras 2020, footnote 6; but see Sekerina & Trueswell 2011 and the short summary in Polinsky 2018: 158–162).

HL listeners have been claimed to have a perceptual advantage (Polinsky 2018: 11) due to the exposure to the HL in childhood which gives them an advantage in phonological categorization and perception of contrasts, particularly in segmental phonology. Prosodic distinctions have been claimed to be much more subject to change and variation.

Prior work on the processing of prosody and word order in mono- and bilingual Russian listeners has attested only a weak effect of contrastive accent (Sekerina & Trueswell 2011). In addition, Sekerina & Trueswell (2011) found that it is relevant for the processing of prosody whether a semantic contrast is explicitly expressed or not. Polinsky (2018: 161f) suggests that “HSs have a selective prosodic advantage over native speakers: [they] are sensitive to stronger, more salient prosodic cues” but “[t]he prosodic defaults in heritage speakers may be more general than the defaults established by native speakers”.

In an online perception experiment, Zerbian & Barabashova (2023) tested the acceptability of different accent patterns in adjective+noun phrases with semantic focus on the adjective (Adj<sub>CF</sub>+N) by mono- and bilingual listeners of Russian. The sentence structure was SVO with a verb in active voice. Both naturally produced and lab-recorded Adj<sub>CF</sub>+N phrases were presented auditorily, as part of a discourse context which evokes semantic contrastive focus on the adjective (in bold), as in (7).

- (7) Èta belaja mašina vrezalas' v sinjuju mašinu.  
this white car crashed into blue car  
'This white car crashed into a blue<sub>CF</sub> car.'

Listeners had to evaluate how natural the presented  $\text{Adj}_{\text{CF}}+\text{N}$  phrases sounded. The accent patterns varied with respect to accent on the noun, on the adjective or on both. Moreover, the  $\text{Adj}_{\text{CF}}+\text{N}$  stimuli represented either the first or second noun phrase in the sentence (i.e., encoding the contrast explicitly only if occurring as the second noun phrase, as in (7)).

The aim was to investigate whether mono- and bilingual listeners rate different accent patterns in  $\text{Adj}_{\text{CF}}+\text{N}$  differently with respect to naturalness. Furthermore, it was asked whether the position (i.e., whether the noun phrase is mentioned first or second in a sentence) leads to a difference in the naturalness ratings. The ratings of accent patterns on  $\text{Adj}_{\text{CF}}+\text{N}$  were predicted to depend on three factors. First, they were predicted to depend on accent location such that adjective accents would be rated high on the naturalness scale by both mono- and bilingual listeners and clearly different in acceptability from accents on nouns. Second, based on production data (see Section 4.2) they were predicted to depend on the bilingualism of listeners such that double accents (i.e., on both adjective and noun) would be rated as more natural more often by bilingual than by monolingual listeners. And third, they were predicted to depend on the NP's position in the sentence such that double accents would be rated as more natural if they occurred on the first noun phrase of a sentence rather than on the second, because in the second noun phrase the contrast is linguistically explicit.

The online study consisted of two experimental parts. Experimental part 1 elicited naturalness ratings for  $\text{Adj}_{\text{CF}}+\text{N}$  with differing accent patterns, namely an accent on the adjective, on the noun or on both (double accent), using a scale from 1 (= not at all natural) to 7 (= very natural). Experimental part 2 explored the role of explicit linguistic contrast for naturalness ratings of  $\text{Adj}_{\text{CF}}+\text{Ns}$  as first and second occurrence in an utterance, respectively.

For both experimental parts, original productions of  $\text{Adj}_{\text{CF}}+\text{N}$  from the RUEG corpus were used (see Section 2.1 above). The narrations described an accident involving two differently coloured cars, thus providing the context for a semantic contrastive focus on the colour adjective. As a control condition, listeners were additionally presented with lab-produced recordings of  $\text{Adj}_{\text{CF}}+\text{N}$  phrases, with the same three accent patterns: on the adjective, on the noun, and on both.

Participants were recruited via the platforms Prolific and Mechanical Turk. Requirements specified via the platform included the US or Germany as location, Russian as L1, and Russian and English or German as fluent languages. Twelve participants were either born to Russian-speaking parents in the US or Germany or immigrated to the respective countries when they were younger than 5 years old and are thus considered Russian HSSs. Sixteen monolingually-raised Russian

speakers who lived in Russia and were recruited via social networks served as a comparison group.

With the lab-recorded stimuli, both listener groups rated an accent on the adjective (HSs: mean=5.4, mono: mean=6.0)<sup>4</sup> as more natural than a double accent (HSs: mean=4.0, mono: mean=3.6) or an accent on the noun (HSs: mean=3.9, mono: mean=4.2). For the stimuli taken from the corpus, heritage listeners rated accents on adjectives on average higher (mean=5.3) than double accents (mean=4.9) or accents on nouns (mean=4.8), though less clearly so than in the lab-recorded stimuli. For monolingual listeners, all accent patterns were rated as similarly highly natural.

As for the role of explicit linguistic contrast (whether  $\text{Adj}_{\text{CF}}+\text{N}$  was first or second occurrence), monolingual as well as heritage listeners judge accents on adjectives as more natural than double accents both in first and second position. However, heritage listeners rated double accents significantly more natural in first position (mean=4.7) than in second (mean=3.9). For monolingual listeners, the position did not matter.

The results for the lab-recorded stimuli were in line with the literature for both listener groups, namely that adjective accent is judged significantly more natural than an accent on the noun or double accent. Thus, both listener groups were shown to be sensitive to accent and its relation to information structure in Russian.

For the stimuli taken from natural productions, all accent patterns sounded equally (highly) natural, at least to monolingual listeners. The naturalness and greater variability of the data might have indirectly led to convergence of judgments in monolingual speakers. This might be even more so due to the generally weak load of accent in Russian (Sekerina & Trueswell 2011). Bilingual heritage listeners, on the other hand, might benefit from a perceptual advantage that leads to “better-than-native perception” (e.g., Chang 2016). HSs have more familiarity with variation and accents and might thus adapt more quickly to varied speech.

## **6 Discussion**

What emerges across our studies on intonation presented in this chapter is that we find evidence for various differences between monolingually-raised speakers and HSs. For intonation used at the linguistic interface with semantics and pragmatics, such as in contrastive focus or discourse linking, this is in line with the

---

<sup>4</sup>Further details on the spread of the data are given in Zerbian & Barabashova (2023).

Interface Hypothesis (e.g., Sorace 2011) which predicts phenomena at linguistic interfaces to show variation in language contact.

Variation is found along different dimensions of intonation (see the LILt model by Mennen 2015). For example, the different phonetic implementation of boundary tones in YNQs by Russian HSs in the US as opposed to Germany can be considered variation along the realizational dimension. The increased frequency of pitch accents across the productions in the corpus (Section 3) as well as in the additional experimental items of YNQs (Section 4.1) point to a difference along the frequency dimension. At the same time, the latter can also be considered variation along the semantic dimension as pitch accents occur on otherwise deaccented constituents and thereby might take on a different semantic meaning. The same reasoning can be applied to the placement of accents in modified noun phrases with contrastively focused adjectives (Section 4.2). The use of phrasing (Section 4.3) to potentially distinguish frame-setting adverbials from discourse-linking adverbials is another instance of the semantic dimension of intonation.

It is noteworthy that the frequency difference concerning the occurrence of pitch accents emerges across different data sets (Sections 3, 4.1, 4.2). Similarly, an effect of formality emerges across different data sets (Sections 3, 4.2). Zerbian et al. (2022) found that in the cases investigated HSs of Russian in the US accentuate both a contrastive adjective and noun. Monolingually-raised speakers of Russian were found to place a single pitch accent on the adjective in the informal situation, but not in the formal one where they produced double accents on both adjective and a noun. Thus, there is a possible formality effect for monolingually-raised speakers (note the few realizations though), but not for the HSs. The same trend was found for overall intonational features (Section 3): shorter IPs in formal than in informal settings in monolingually-raised speakers but a levelling of registers for this feature in HSs (Zerbian et al. 2024). Our findings are in line with other studies that report register-levelling processes in HSs in other linguistic domains (e.g., Colantoni et al. 2016, Alexiadou et al. 2022, Schroeder et al. forthcoming). Register-levelling might be due to a lack of exposure to formal registers. The potential influence of exposure to formal registers on prosodic features, such as pitch and speech rate, is detailed in Rao et al. (2020).

In the cases reported here, the observed differences cannot clearly be traced back to cross-linguistic influence from the majority language. For boundary tones in YNQs, HSs of Russian in the US seem to implement them closest to what has been reported for monolingually-raised speakers of Russian, although the latter (and HSs in Germany) were not found to strongly prefer that pattern themselves. Some HS groups might be more conservative in their linguistic features (e.g., Polinsky 2018: 129), overemphasizing features of the HL.

What also emerges across all studies on intonation is that HSs do not behave completely differently compared to monolingually-raised speakers. Rather, they show certain patterns – these could be innovative or conservative patterns – more frequently than monolingually-raised speakers. This is in line with considering heritage and monolingually-raised speakers as part of the same native speaker continuum (see Wiese et al. 2022), albeit situated at different places along this continuum.

Large-scale, multilingual heritage language corpora like the one of the Heritage Language Variation and Change Project (HLVC) (Nagy 2011) or the RUEG corpus provide opportunities to address questions of heritage language grammar beyond individual language pairings. For example, the RUEG corpus allows differentiating between the influence of transfer, bilingualism, or group-specific patterns in the linguistic phenomena under investigation. Thus, one direction of future research is to investigate whether general prosodic patterns observed with Russian HSs are also attested in other HS groups. If, for example, smaller IPs and an increase in pitch accents is a general feature of bilingual speakers, then we would expect to find it also in Greek and Turkish HSs as compared to monolingually-raised speakers. A first comparison along these lines for the intonation of AdjCF+N sequences by mono- and bilingual speakers of English does not suggest a unified pattern of increased double accents across different groups of HSs (see Section 4.2). Thus, this seems to make it a group-specific pattern in this particular instance, even if more accents are often reported for bilingual speakers. The RUEG corpus affords the unique situation to have comparable data available to test such hypotheses across different language pairings. The findings can inform models on the role of transfer, bilingualism and group-specific innovations in heritage language grammar.

Furthermore, the RUEG corpus also allows testing hypotheses about prosodic features across formal and informal communicative situations. Qualitative research will allow for a better understanding of some specific results of our studies. For instance, shorter IPs and more frequent accentuation in formal situations by monolingually-raised speakers of Russian might be connected to differences in the discourse structure in the two communicative situations.

The work reported in this chapter thus contributes to theories of language contact, such as the L2 Intonation Learning theory (LILT, Mennen 2015) or the Interface Hypothesis (Sorace 2011). The results help to identify stable and dynamic features of intonation in HSs. Furthermore, when HSs and monolinguals showed differences in intonation, possible sources behind these differences were discussed.

## Acknowledgments

The research reported in this chapter was funded by the German Research Foundation (DFG) as part of the Research Unit *Emerging Grammars in Language Contact Situations: A comparative approach* (FOR 2537) in project P8 (grant number: 313607803). We thank all research assistants who contributed to the work reported in this chapter either by data collection, data annotation and/or technical help. These are Yuliia Ivashchyk, Nash Whaley, Daria Alkhimchenkova, Erica Conti, Myrto Rompaki, Birte Pravemann. We further thank Lea Coy for help with pre-publication formatting. We are also grateful to the anonymous external reviewers and the RUEG-internal reviewers Johanna Tausch, Sofia Grigoriadou, and Martin Klotz for constructive feedback. Last but not least, a special thanks to the entire RUEG team and the Mercator fellows for fruitful discussions and guidance.

## References

- Alexiadou, Artemis, Vasiliki Rizou & Foteini Karkaletsou. 2022. A plural indefinite article in heritage Greek: The role of register. *Languages* 7(2). 115. DOI: 10.3390/languages7020115.
- Auer, Peter & Susanne Günthner. 2005. Die Entstehung von Diskursmarkern im Deutschen: Ein Fall von Grammatikalisierung? In Torsten Leuschner & Tanka Mortelmans (eds.), *Grammatikalisierung im Deutschen*, 335–362. Berlin: de Gruyter. DOI: 10.1515/9783110925364.335.
- Boersma, Paul. 2001. Praat: A system for doing phonetics by computer. *Glot International* 5(9/10). 341–345.
- Böttcher, Marlene. 2021. *The prosody of contrastive adjectives in mono- and bilingual English*. University of Stuttgart. (MA thesis).
- Böttcher, Marlene. 2022. A comparison of pitch accent patterns in contrastive adjective+noun structures in bilingual Englishes. *Proceedings of Speech Prosody, Lisbon*. 807–811. DOI: 10.21437/SpeechProsody.2022-164.
- Böttcher, Marlene & Margaret Zellers. 2024. Do you say *uh* or *uhm*? A cross-linguistic approach to filler particle use in heritage and majority speakers across three languages. *Frontiers in Psychology* 15. DOI: 10.3389/fpsyg.2024.1305862.
- Breitbarth, Anne. 2022. Prosodie, Syntax und Diskursfunktion von V>2 in gesprochenem Deutsch. *Deutsche Sprache* (1). 1–30. DOI: 10.37307/j.1868-775x.2022.01.02.

- Brinton, Laurel J. 2010. *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin: Walter de Gruyter.
- Bryzgunova, Elena. 1980. Intonacija. In Natalia J. Švedova, Nina D. Arutjunova, Alexander V. Bondarko, Vyacheslav V. Ivanov, Vladimir V. Lopatin, Igor S. Uluxanov & Fedot P. Filin (eds.), *Russkaja grammatika*, vol. 1, 96–122. Moscow: Akademija Nauka.
- Bunk, Oliver. 2020. „Aber immer alle sagen das“: The status of V3 in German: Use, processing, and syntactic representation. Humboldt-Universität zu Berlin. (Doctoral dissertation). DOI: 10.18452/22085.
- Bunk, Oliver, Shanley E. M. Allen, Sabine Zerbian, Tatiana Pashkova, Yulia Zuban & Erica Conti. 2025. Information packaging and word order dynamics in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 379–412. Berlin: Language Science Press. DOI: 10.5281/zenodo.1577518.
- Bunk, Oliver & Maike H. Rocker. In preparation. *The status of initial adverbials in V3 sentences: A comparative corpus study on prosodic realization*.
- Bybee, Joan L., William Pagliuca & Revére D. Perkins. 1991. Back to the future. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 2: Types of grammatical markers, 17–58. Amsterdam: John Benjamins.
- Chang, Charles B. 2016. Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition* 19(4). 791–809. DOI: 10.1017/s1366728914000261.
- Chang, Charles B. 2021. Phonetics and phonology of heritage languages. In Silvina Montrul & Maria Polinsky (eds.), *The Cambridge handbook of heritage languages and linguistics*, 581–612. Cambridge: Cambridge University Press. DOI: 10.1017/9781108766340.027.
- Chang, Charles B. & Yao Yao. 2016. Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal* 13(2). 134–160. DOI: 10.46538/hlj.13.2.4.
- Christoffersen, Katherine. 2017. Comparing native speaker ratings and quantitative measures of oral proficiency in IELTS interviews. *Estudios de Linguistica Inglesa Aplicada (ELIA)* (17). 233–250. DOI: 10.12795/elia.2017.i17.10.
- Colantoni, Laura, Alejandro Cuza & Natalia Mazzaro. 2016. Task-related effects in the prosody of Spanish heritage speakers and long-term immigrants. In Meghan E. Armstrong, Nicholas C. Henriksen & Vanrell Maria del Mar (eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields*, 1–24. Amsterdam: John Benjamins.

- Comstock, Lindy Burden. 2018. *Pragmatic accommodation and linguistic salience in U.S.-Russian political interviews*. University of California. (Doctoral dissertation).
- de Jong, Nivja H. 2016. Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching* 54(2). DOI: 10.1515/iral-2016-9993.
- Dehé, Nicole. 2018. The intonation of polar questions in North American (“heritage”) Icelandic. *Journal of Germanic Linguistics* 30(3). 213–259. DOI: 10.1017/s1470542717000125.
- Dehé, Nicole & Meike Rommel. 2024. The intonation of declaratives and polar questions in modern versus heritage Icelandic. In Rajiv Rao (ed.), *Heritage language phonology*, 259–279. Cambridge: Cambridge University Press.
- Dehé, Nicole & Katerina Stathi. 2016. Grammaticalization and prosody: The case of English *sort/kind/type of* constructions. *Language* 92(4). 911–946. DOI: 10.1353/lan.2016.0077.
- Drach, Erich. 1937. *Grundgedanken der deutschen Satzlehre*. Frankfurt am Main: Diesterweg.
- Feldhausen, Ingo & Maria del Mar Vanrell. 2024. Focus realization in heritage Spanish. The case of German-dominant speakers of peninsular Spanish. In Rajiv Rao (ed.), *Heritage language phonology*, 166–189. Cambridge: Cambridge University Press.
- Flores, Cristina & Anabela Rato. 2016. Global accent in the Portuguese speech of heritage returnees. *Heritage Language Journal* 13(2). 161–183. DOI: 10.46538/hlj.13.2.5.
- Freywald, Ulrike, Leonie Cornips, Natalia Ganuza, Ingvild Nistov & Toril Opsahl. 2015. Beyond verb second – a matter of novel information structural effects? Evidence from Norwegian, Swedish, German, and Dutch. In Jacomine Nortier & Bente A. Svendsen (eds.), *Language, youth and identity in the 21st century: Linguistic practices across urban spaces*, 73–92. Cambridge: Cambridge University Press.
- Gosling, Samuel D., Peter J. Rentfrow & William B. Swann. 2003. A very brief measure of the big five personality domains. *Journal of Research in Personality* 37. 504–528.
- Götze, Michael, Thomas Weskott, Cornelia Endriss, Ines Fiedler, Stefan Hinterwimmer, Svetlana Petrova, Anne Schwarz, Stavros Skopeteas & Ruben Stoel. 2007. Information structure. *Interdisciplinary Studies on Information Structure: ISIS* (7). 147–187.

- Gries, Stefan Th. & Viola G. Miglio. 2014. New information in naturalistic data is also signalled by pitch movement: An analysis from monolingual English/Spanish and bilingual Spanish speakers. *Complutense Journal of English Studies* 22(0). 11–33.
- Haspelmath, Martin. 2002. Grammatikalisierung: Von der Performanz zur Kompetenz ohne angeborene Grammatik. In Sybille Krämer & Ekkehard König (eds.), *Gibt es eine Sprache hinter dem Sprechen?*, 162–286. Frankfurt am Main: Suhrkamp.
- Himmelmann, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology* 35(2). 207–245. DOI: [10.1017/s0952675718000039](https://doi.org/10.1017/s0952675718000039).
- Hirschberg, Julia & Diane Litman. 1993. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics* 19(3). 501–530.
- Höhle, Tilman N. 1986. Der Begriff „Mittelfeld“: Anmerkungen über die Theorie der topologischen Felder. In Walter E. Weiss, Herbert E. Wiegand & Marga Reis (eds.), *Kontroversen, alte und neue: Akten des VII. Internationalen Germanisten-Kongresses Göttingen 1985*, 329–340. Tübingen: Niemeyer.
- Hughes, Diana L., LaRae McGillivray & Mark Schmidek. 1997. *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking Publications.
- Imo, Wolfgang. 2012. Wortart Diskursmarker? In Björn Rothstein (ed.), *Nichtflektierende Wortarten*, 48–88. Berlin: De Gruyter.
- Iwashita, Noriko, Annie Brown, Tim McNamara & Sally O'Hagan. 2008. Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics* 29(1). 24–49. DOI: [10.1093/applin/amm017](https://doi.org/10.1093/applin/amm017).
- Jasinskaja, Katja. 2016. Information structure in Slavic. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure* (Oxford handbooks in linguistics), 709–732. Oxford: Oxford University Press.
- Jucker, Andreas H. & Yael Ziv. 1998. *Discourse markers: Descriptions and theory*. Amsterdam: John Benjamins.
- Kim, Ji-Young. 2019. Heritage speakers' use of prosodic strategies in focus marking in Spanish. *International Journal of Bilingualism* 23(5). 986–1004. DOI: [10.1177/1367006918763139](https://doi.org/10.1177/1367006918763139).
- King, Tracy Holloway. 1993. *Configuring topic and focus in Russian*. Stanford University. (Doctoral dissertation).
- Kotsinas, Ulla-Britt. 1992. Immigrant adolescents' Swedish in multicultural areas. In Cecilia Palmgren, Karin Lövgren & Göran Bolin (eds.), *Ethnicity in youth culture*, 43–62. Stockholm: University of Stockholm.

- Kroll, Judith F. & Tamar H. Gollan. 2014. Speech planning in two languages: What bilinguals tell us about language production. In Matthew Goldrick, Victor Ferreira & Michele Miozzo (eds.), *The Oxford handbook of language production*, 165–181. Oxford: Oxford University Press.
- Kupisch, Tanja, Dagmar Barton, Katja Hailer, Ewgenia Klaschik, Ilse Stangen, Tatjana Lein & Joost van de Weijer. 2014. Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal* 11(2). 123–150. DOI: 10.46538/hlj.11.2.2.
- Ladd, D. Robert. 2008. *Intonational phonology*. 2nd edn. (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press.
- Laleko, Oksana & Irina Dubinina. 2018. Word order production in heritage Russian: Perspectives from linguistics and pedagogy. In Susan Bauckus & Susan Kresin (eds.), *Connecting across languages and cultures: A heritage language Festschrift in honor of Olga Kagan*, 191–215. Bloomington: Slavica.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehayé, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Makarova, Veronika. 2003. Intonation features in categorization of Russian sentence type. In Peter Kosta, Joanna Blaszcak, Frasek & Ljudmila Geist (eds.), *Investigations into formal Slavic linguistics*, vol. IV, 83–95. Berlin: Peter Lang.
- Meelen, Marieke, Khalid Mourigh & Lisa Lai-Shen Cheng. 2020. V3 in urban youth varieties of Dutch. In András Bárány, Theresa Biberauer, Jamie Douglas & Sten Vikner (eds.), *Syntactic architecture and its consequences II: Between syntax and morphology*, 327–355. Berlin: Language Science Press. DOI: 10.5281/zenodo.4280657.
- Mennen, Ineke. 2015. Beyond segments: Towards a L2 intonation learning theory. In Elisabeth Delais-Roussarie, Mathieu Avanzi & Sophie Herment (eds.), *Prosody and language in contact L2 acquisition, attrition and languages in multilingual situations*, 171–188. Berlin: Springer.
- Meyer, Roland & Ina Mleinek. 2006. How prosody signals force and focus: A study of pitch accents in Russian yes-no questions. *Journal of Pragmatics* 38(10). 1615–1635. DOI: 10.1016/j.pragma.2005.05.011.
- Montrul, Silvina. 2015. *The acquisition of heritage languages*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139030502.

- Nagy, Naomi. 2011. A multilingual corpus to explore geographic variation. *Rassegna Italiana di Linguistica Applicata* 43 (1-2). 65–84.
- Nagy, Naomi. 2017. Documenting variation in (endangered) heritage languages: How and why? In Kristine A. Hildebrandt, Carmen Jany & Wilson Silva (eds.), *Documenting variation in endangered languages* (Language Documentation and Conservation Special Publication 13), 33–64. Honolulu: University of Hawai'i Press.
- Opsahl, Toril. 2009. Wolla I swear this is typical for the conversational style of adolescents in multiethnic areas in Oslo. *Nordic Journal of Linguistics* 32(2). 221–244. DOI: 10.1017/s0332586509990059.
- Pashkova, Tatiana, Marlene Böttcher, Kalliopi Katsika, Sabine Zerbian & Shanley E. M. Allen. 2025. Majority English of heritage speakers. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 449–491. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775185.
- Petersen, Jan Heegård, Gert Foget Hansen, Jacob Thøgersen & Karoline Kühl. 2021. Linguistic proficiency: A quantitative approach to immigrant and heritage speakers of Danish. *Corpus Linguistics and Linguistic Theory* 17 (2). 465–490. DOI: 10.1515/cllt-2017-0088.
- Polinsky, Maria. 2008. Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization. *Heritage Language Journal* 6(1). 40–71.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Polinsky, Maria & Gregory Scontras. 2020. Understanding heritage languages. *Bilingualism: Language and Cognition* 23(1). 4–20. DOI: 10.1017/S1366728919000245.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en Español: Toward a typology of code-switching. *Linguistics* 18(7–8). DOI: 10.1515/ling.1980.18.7-8.581.
- Quist, Pia. 2000. Ny københavnsk "multietnolekt": Om sprogbrug blandt unge i sprogligt og kulturelt heterogene miljøer [New Copenhagen "multi-ethnolect": About language use among young people in linguistically and culturally heterogeneous environments]. *Danske Talesprog* 1. 143–211.
- Rao, Rajiv. 2016. On the nuclear intonational phonology of heritage speakers of Spanish. In Diego Pascual y Cabo (ed.), *Advances in Spanish as a heritage language* (Studies in Bilingualism 49), 51–80. Amsterdam: John Benjamins.

- Rao, Rajiv, Zuzanna Fuchs, Maria Polinsky & María Luisa Parra. 2020. The sound pattern of heritage Spanish: An exploratory study on the effects of a classroom experience. *Languages* 5(4). DOI: 10.3390/languages5040072.
- Rathcke, Tamara. 2006. A perceptual study on Russian questions and statements. In Jonathan Harrington, Christine Mooshammer & Felicitas Kleber (eds.), *Aipuk* 37, 51–62. Kiel: Universität Kiel.
- Robles-Puente, Sergio. 2014. *Prosody in contact: Spanish in Los Angeles*. University of Southern California. (Doctoral dissertation).
- Rocker, Maike H. 2022. *Variation in finite verb placement in heritage Iowa Low German: The role of prosodic integration and information structure*. State College: Pennsylvania State University. (PhD Thesis).
- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2). 155–163. DOI: 10.1177/1367006909339814.
- Schalowski, Sören. 2017. From adverbial to discourse connective. Multiple pre-fields in spoken German and the use of *dann* ‘then’ and *danach* ‘afterwards’. In Heike Wiese, Heiko F. Marten, Philip Bracker & Oliver Bunk (eds.), *Arbeitspapiere Sprache, Variation und Migration: Studentische Arbeiten*, 1–45. Potsdam: Online Publication, Universität Potsdam. [https://www.uni-potsdam.de/fileadmin/projects/svm/Arbeitspapiere/No6\\_Schalowski\\_2017.pdf](https://www.uni-potsdam.de/fileadmin/projects/svm/Arbeitspapiere/No6_Schalowski_2017.pdf).
- Schroeder, Christoph, Kateryna Iefremenko & Mehmet Öncü. Forthcoming. The post-verbal position in heritage Turkish. A comparative approach with a focus on non-clausal elements. In Zeynep Kalkavan-Aydin & Yazgül Şimşek (eds.), *Zweisprachigkeit Deutsch-Türkisch: Studien in Deutschland und in den Nachbarländern*. Münster: Waxmann.
- Sekerina, Irina A. & John C. Trueswell. 2011. Processing of contrastiveness by heritage Russian bilinguals. *Bilingualism: Language and Cognition* 14(3). 280–300. DOI: 10.1017/s1366728910000337.
- Sorace, Antonella. 2011. Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism* 1(1). 1–33. DOI: 10.1075/lab.1.1.01sor.
- Staley, Larssyn & Andreas H. Jucker. 2021. “The *uh* deconstructed pumpkin pie”: The use of *uh* and *um* in Los Angeles restaurant server talk. *Journal of Pragmatics* 172. 21–34. DOI: 10.1016/j.pragma.2020.11.004.
- Statistisches Bundesamt. 2022. *Bevölkerung mit Migrationshintergrund: Ergebnisse des Mikrozensus 2021 [Population with a migration background: results of the 2021 microcensus]*. [https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/\\_publikationen-innen-migrationshintergrund.html](https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/_publikationen-innen-migrationshintergrund.html). Last accessed 2024-01-10.

- Svetozarova, Natalia. 1998. Intonation in Russian. In Daniel Hirst & Albert Di Cristo (eds.), *Intonation systems: A survey of twenty languages*, 264–277. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. 2016. *Teen talk: The language of adolescents*. Cambridge: Cambridge University Press.
- te Velde, John R. 2017a. Temporal adverbs in the Kiezdeutsch left periphery: Combining late merge with deaccentuation for V3. *Studia Linguistica* 71(3). 301–336. DOI: 10.1111/stul.12055.
- te Velde, John R. 2017b. Towards an account of adverbials in the Vor-Vorfeld: Beyond V2-syntax. *Florida Linguistics Papers* 4(1). 62–86.
- Tottie, Gunnell. 2014. On the use of *uh* and *um* in American English. *Functions of Language* 21 (1). 6–29. DOI: 10.1075/fol.21.1.02tot.
- Tracy, Rosemarie & Elsa Lattey. 2010. “It wasn’t easy but irgendwie äh da hat sich’s rentiert, net?”: A linguistic profile. In Michaela Albl-Mikasa (ed.), *Dimensionen der Zweitsprachenforschung*, 53–73. Tübingen: Narr.
- Traugott, Elizabeth C. 1995. *The role of the development of discourse markers in a theory of grammaticalization*. Paper presented at the International Conference of Historical Linguistics XII.
- United States Census Bureau. 2009. *Language spoken at home by ability to speak English for the population 5 years and over*. <https://data.census.gov/cedsci/table?q=B16001&tid=ACSDT1Y2019.B16001>, Last accessed on 2024-01-10.
- van Rijswijk, Remy, Antje Muntendam & Ton Dijkstra. 2017. Focus marking in Dutch by heritage speakers of Turkish and Dutch L1 speakers. *Journal of Phonetics* 61. 48–70. DOI: 10.1016/j.wocn.2017.01.003.
- Wandrei, Alina. 2019. *Herkunftssprache Russisch in Deutschland: Eine Betrachtung mit Fokus auf phonologischen Aspekten*. University of Stuttgart. (MA thesis).
- Wichmann, Anne. 2011. Prosody and pragmatic effects. *Pragmatics of Society*. 181–214. DOI: 10.1515/9783110214420.181.
- Wichmann, Anne, Anne-Marie Simon-Vandenbergen & Karin Aijmer. 2010. How prosody reflects semantic change: A synchronic case study of *of course*. In Hubert Cuyckens, Kristin Davids & Lieven Vandeanotte (eds.), *Subjectification, intersubjectification and grammaticalization*, 103–154. Berlin: De Gruyter Mouton.
- Wiese, Heike. 2009. Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. *Lingua* 119(5). 782–806. DOI: 10.1016/j.lingua.2008.11.002.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers’ repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology*

- XVI (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Anke Lüdeling, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Tracy Rosemarie, Christoph Schroeder, Luka Szucsich, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2021. *RUEG Corpus*. Version 0.4.0. Zenodo. DOI: 10.5281/zenodo.3236068.
- Wiese, Heike, Oliver Bunk, Fynn Dobler, Ulrike Freywald, Sophie Hamm, Banu Hueck, Anne Junghans, Jana Kiobassa, Julia Kostka, Marlen Leisner, Nadine Lestmann, Katharina Mayr, Tiner Özcelik, Charlotte Pauli, Gergana Popova, Ines Rehbein, Nadja Reinhold, Franziska Rohland, Sören Schalowski, Kathleen Schumann, Kristina T. Sommer & Emiel Visser. 2010. *KiDKo: Ein Korpus spontaner Unterhaltungen unter Jugendlichen im multiethnischen und monoethnischen urbanen Raum*. <http://kiezdeutschkorpus.de/en/>.
- Wiese, Heike & Hans Georg Müller. 2018. The hidden life of V3: An overlooked word order variant on verb-second. In Mailin Antomo & Sonja Müller (eds.), *Non-canonical verb positioning in main clauses* (Linguistische Berichte Sonderheft 25), 202–223. Hamburg: Buske.
- Wiese, Heike & Ines Rehbein. 2016. Coherence in new urban dialects: A case study. *Lingua* 172–173. 45–61. DOI: 10.1016/j.lingua.2015.10.016.
- Zerbian, Sabine & Kristina Barabashova. 2023. Perceived naturalness of accents in noun phrases by mono- and bilingual listeners of Russian. *Proceedings of ICPhS 2023, Prague*. 1270–1274.
- Zerbian, Sabine, Kristina Barabashova & Yulia Zuban. 2025. *Left dislocations in monolingual and heritage Russian in Germany and the U.S.* DOI: <https://doi.org/10.1163/15507076-bja10035>.
- Zerbian, Sabine, Marlene Böttcher & Yulia Zuban. 2022. Prosody of contrastive adjectives in mono-and bilingual speakers of English and Russian: A corpus study. In *Proceedings 11th international conference on speech prosody. International speech communication association (ISCA) online archive*, 812–816. DOI: 10.21437/SpeechProsody.2022-165.

- Zerbian, Sabine, Yulia Zuban & Martin Klotz. 2024. Intonational features of spontaneous narrations in monolingual and heritage Russian in the U.S.: An exploration of the RUEG corpus. *Languages* 9(1). DOI: 10.3390/languages9010002.
- Zuban, Yulia. 2024. *Word order choice, expression of information status, and global intonation patterns in heritage Russian*. Stuttgart: Universität Stuttgart. (Doctoral dissertation).
- Zuban, Yulia, Maria Martynova, Sabine Zerbian, Luka Szucsich & Natalia Gagrina. 2021. Word order in heritage Russian: Clause type and majority language matter. *Russian Linguistics* 45(3). 253–281. DOI: 10.1007/s11185-021-09246-1.
- Zuban, Yulia, Tamara Rathcke & Sabine Zerbian. 2023. Do different majority languages lead to different intonational grammars? A case study of yes/no-questions in heritage Russian. *Heritage Language Journal* 20(1). 1–43. DOI: 10.1163/15507076-bja10016.

# Chapter 13

## Majority English of heritage speakers

Tatiana Pashkova<sup>a</sup>, Marlene Böttcher<sup>b</sup>, Kalliopi Katsika<sup>a</sup>,  
Sabine Zerbian<sup>c</sup> & Shanley E. M. Allen<sup>a</sup>

<sup>a</sup>University of Kaiserslautern-Landau <sup>b</sup>Kiel University <sup>c</sup>University of Stuttgart

Research on heritage speakers to date has focused largely on their heritage languages. In contrast, while their majority languages have been often examined in young children, less is known about adolescents and adults. Filling this gap, the current chapter presents our research on majority English of adolescent and adult heritage speakers of German, Greek, Russian, and Turkish compared to monolingually-raised English speakers. Using the RUEG corpus, we investigated four external interface areas prone to dynamicity in language contact – prosody, article use, reference, and clausal syntax in relation to information and discourse structure. In addition, we present findings on discourse organization. Based on our results, we discuss implications for majority language research.

### 1 Introduction

Most research on heritage speakers (HSs) to date focuses on their heritage language. In this chapter we focus on their other language – the majority language in the society – and ask whether the bilingualism resulting from heritage language (HL) maintenance leads to differences in the way heritage and monolingually-raised speakers use the majority language (ML).

HSs are bilinguals who typically speak one language within the family and other personal spheres (e.g., cultural or religious settings) – the HL – and another language that is the language of the larger society in which they live – the ML (Pascual y Cabo & Rothman 2012, Montrul 2016, Montrul & Polinsky 2021). They typically begin acquiring the HL at birth, and begin learning the ML sometime between birth and the onset of schooling. Through formal schooling they often



shift in dominance from the HL to the ML, such that they become dominant in the ML by adulthood. There is typically a wide variation across speakers in adult proficiency in the HL: they may retain full fluency, may gradually reduce dominance in HL so that they only comprehend but do not speak the HL, or anything in between.

The linguistic study of HSs has blossomed over the last two decades. However, as already noted, most research on HSs to date focuses on their HL (Montrul 2008, Kupisch 2013, Montrul 2016, Polinsky 2018, Montrul & Polinsky 2021). This is understandable from the cultural standpoint since there is considerable interest in supporting HL retention, and also from the linguistic research standpoint since it is both theoretically and practically relevant to explore what the HL looks like after dominance is lost.

However, research on the ML is also valuable for at least two reasons. First, it contributes to our understanding of cross-linguistic influence and effects of bilingualism. Many studies on HSs have focused on the influence of the ML on the HL – in other words, the more dominant language on the less dominant one. However, research on L2 learners has highlighted that the influence also occurs in the opposite direction, from the less dominant language to a more dominant one (see, for example, Schoonbaert et al. 2007, Chen et al. 2013 for late L2-to-L1 structural priming; Pavlenko & Jarvis 2002, Hohenstein et al. 2006, Gorba 2019 for late L2-to-L1 cross-linguistic influence).

Thus, we expect that research on the ML of HSs can provide insights into whether this type of cross-linguistic influence can also be attested in HL-to-ML situations. Moreover, we might observe general effects in the ML that are not traceable to any particular HL, which would add evidence that HL bilingualism in itself can lead to dynamic language patterns.

Second, ML research on adolescent and adult HSs is crucial to inform language education policies, which often argue against HL maintenance due to its harmful effects on ML ultimate attainment (e.g., McCardle 2015, Tracy & Gawlitzeck 2023: 30–33). By comparing the ML of adult and adolescent HSs to their MS counterparts, we aim to provide practical evidence that HL retention does not negatively affect the ML in the long-term perspective.

Most research on the ML focuses on children given the goal that children become competent in the ML to facilitate school achievement, with primary focus on the lexicon and morphology (e.g., Bialystok et al. 2010, Marinis & Chondrogianni 2010, Paradis & Jia 2017). Research on the ML in adolescents and adults is also evident in the literature, but it is relatively less common. Most of it shows some differences between the ML as produced or comprehended by HSs compared to MSs. In the area of *semantics*, for example, both Scontras et al. (2017)

and Lee et al. (2011) showed differences in scope marking patterns in majority English between English MSs and Korean and Mandarin HSs, which the authors attribute to cross-linguistic influence from the scope patterns in Korean and Mandarin. With regard to *phonology*, HSs are typically found to not have a detectably different accent from MSs of the ML (e.g., Kupisch et al. 2014, Lloyd-Smith et al. 2020), although there is some evidence that they articulate more explicitly (Polinsky 2018: 142–144). Findings for perception are mixed: Chang (2016) finds that Korean HSs perceive unreleased stops more accurately in majority English than English MSs due to cross-linguistic influence from Korean, while Lee-Ellis (2012) shows that HSs with the same language pair perform like English MSs in perceiving an English-specific contrast in nonce words. For prosodic aspects of the ML of HSs, cross-linguistic influence is more consistently evidenced. This was shown for phonetic and phonological aspects of rises in ML German for Turkish HSs (Queen 2012), for broader prosodic features in the intonation of ML Dutch by Turkish HSs (van Rijswijk et al. 2017), and for intonational aspects of polar questions in ML English by Icelandic HSs (Dehé 2018).

In the domain of *morphosyntax*, most but not all studies have found effects of bilingualism in the ML of HSs. For example, three studies on referent tracking found differences between MSs and HSs in the ML. Azar et al. (2020) found that Turkish HSs produced more pronouns as compared to null forms for referent maintenance in majority Dutch compared to Dutch MSs, which can be viewed as higher explicitness of HSs than MSs in the ML. The authors attribute this explicitness to differences in likely input to MSs vs. HSs in the ML: HSs may have many more interlocutors who are L2 speakers of Dutch and who themselves are more explicit in their pronoun use. Contemori and colleagues (Contemori & Ivanova 2021, Contemori et al. 2023) found a somewhat opposite pattern – that Spanish HSs often produced pronouns for referent tracking in majority English in contexts where English MSs would typically produce full NPs. They attributed this to cross-linguistic influence from Spanish, since pronouns would have been appropriate in these contexts in Spanish. Other differences between HSs and MSs were found in the use of definite articles in ML German by Turkish HSs (Felser & Arslan 2019), acceptability judgements of definite and indefinite articles in ML English by Spanish HSs (Montrul & Ionin 2010), and use of subject-verb agreement in ML English by HSs with various HIs (Paradis 2019). However, Montrul & Ionin (2010) found that Spanish HSs patterned like English MSs in some aspects of interpretation of definiteness and possession in majority English, while Lee-Ellis (2012) found that Korean HSs patterned mostly like English MSs in their interpretation of locative alternation structures in majority English, and Kupisch et al. (2017) found that Turkish HSs performed like German MSs in their compreh-

hension of subtle definiteness effects in majority German. Finally, two studies on a variety of morphosyntactic phenomena found that Spanish HSs patterned more closely to Swedish MSs in majority Swedish the higher their language aptitude (Bylund et al. 2012) and the earlier their age of onset of Swedish (Bylund et al. 2021).

In sum, most of the existing research on the ML use of HSs across linguistic domains shows some cross-linguistic influence from their HL or from language contact in general, although some research also shows no differences between HSs and MSs in ML use. Possible reasons for divergent results of various studies are the different language areas and different languages examined, different methods employed and different speakers tested. Our work reduces this methodological variation to different language areas only, while keeping the languages (majority English in contact with heritage German, Greek, Russian, and Turkish), the method (examination of elicited narratives)<sup>1</sup> and the speakers the same. This approach will show us whether there exists a more consistent pattern of difference/similarity between HSs and MSs in the ML.

Furthermore, our research addresses several open theoretical questions in the majority language literature. First, do we see consistent influence on majority English from typologically different HLs if speakers of these HLs are tested in the same methodological paradigm? Only a few studies investigate more than one HS group (Kupisch et al. 2014, Polinsky 2018, Paradis 2019, Labrenz 2023). This is important because it is only through comparing across HS groups that we can determine if a pattern really derives from cross-linguistic influence (different effect across groups depending on properties of HL) or from effects of bilingualism or language contact in general (similar effect across groups regardless of HL).

Second, is the effect of HS bilingualism equally evident in different registers of the ML and do HSs separate the registers in the ML the same way as MSs? The role of register has been examined in the HL, with some indications of register leveling (e.g., Alexiadou et al. 2022, Alexiadou & Rizou 2023, Tsehayé 2023). However, few if any studies have investigated register differentiation in the ML (see Labrenz 2023 as an example of a recent RUEG-based study), so it is unclear if registers are dynamic not only in the HL but also in the ML.

Third, what differences between HSs and MSs can we discover outside of a strict experimental setting and in a more ecologically valid set up? Many of the previous ML studies that found a difference between HSs and MSs used carefully designed experiments to test specific language phenomena (e.g., Lee et al.

---

<sup>1</sup>Except for one study on article use reported in Section 5, which used a written fill-in-the-blank test.

2011 and Scontras et al. 2017 on scope; Contemori & Ivanova 2021 on referring expressions; Felser & Arslan 2019 on definiteness; Paradis 2019 on subject-verb agreement; Bylund et al. 2012, 2021 on various phenomena; Montrul & Ionin 2010 on article choice; van Rijswijk et al. 2017 on prosody). Some authors highlighted that the experiments were designed to be very demanding even for MSs (Bylund et al. 2012, Paradis 2019). In contrast, we are aware of only a few studies that discovered a difference between HSs and MSs in the ML in an elicited narrative set up (Azar et al. 2020, Contemori et al. 2023 on referring expressions). Other studies that used naturalistic data from tasks not specifically targeting a certain linguistic phenomenon (elicited narratives, interviews, picture descriptions) either did not find a difference between HSs and MSs in the ML (Kupisch et al. 2014 and Lloyd-Smith et al. 2020 on perceived foreign accent) or did not provide a conclusive comparison of HSs and MSs (Kupisch 2014 on adjective placement; Lein et al. 2016 on VOT production; Queen 2012 on intonation).

In the present chapter, we address these gaps by reviewing several studies examining the production patterns in majority English of German, Greek, Russian and Turkish HSs living in the USA. These studies were completed in the context of the six-year research project Research Unit *Emerging Grammars* (RUEG) funded by the German Research Foundation (FOR2357). Each of these studies examines a structure at the interface between a core area of language (e.g., phonology, morphology, syntax) and a non-core area of language (e.g., pragmatics, discourse). This follows the Interface Hypothesis (Sorace 2011, Tsimpli 2014), which states that phenomena at the interface between core and non-core areas of language are particularly open to the development of dynamic patterns in bilinguals. Interface structures crucially involve the integration of knowledge between different areas of language, and thus engender a higher cognitive load for processing than structures involving only one area of language. Each study looks for dynamic patterns in the ML productions of HSs and examines them in the light of potential sources of the pattern, including cross-linguistic influence or transfer from the HL to the ML as well as the presence of a language contact situation regardless of the particular languages involved (e.g., increased explicitness compared to MSs). Crucially, the data were collected via elicited narrative in four different communicative situations leading to production of different registers, thus allowing us to see productions in a more ecologically valid context than the experimental setups used in many of the studies to date. Finally, data from the same participants were analyzed in each of the studies on different linguistic phenomena, thus reducing inter-individual variation between studies.

The remainder of the chapter is structured as follows. Section 2 describes the RUEG corpus (Lüdeling et al. 2024) – the source of most of the ML data used in

the studies presented in this chapter. Section 3 looks at HSs' prosody, Section 4 examines reference, Section 5 covers article use, Section 6 explores syntax, and Section 7 turns to openings and closings in discourse. Section 8 critically discusses the reported findings and draws conclusions for heritage language and bilingualism research.

## 2 Corpus

The studies reported in this chapter are based on the data from the English subcorpus of the RUEG corpus (Wiese et al. 2021). The subcorpus contains the English data of 223 HSs with English as their majority language and 64 English MSs, all of whom were raised in the USA and resided there at the time of testing. The HS group includes 34 German, 65 Greek, 65 Russian, and 59 Turkish HSs. A total of 91% of HSs were first exposed to English at the age of 5 or earlier, with 46% having the first contact with English from birth. All speakers – HSs and MSs – comprise two age groups, adolescents (13–18 years old) and adults (20–37 years old); see Table 1.

Speakers were interviewed in large cities in the Eastern United States, or in nearby areas. These included New York City; Washington, D.C.; Chicago, Illinois; Boston, Massachusetts; Madison, Wisconsin; St. Paul and Minneapolis, Minnesota; Long Island, New York; Fort Lee and Bloomfield, New Jersey; and New Haven, Connecticut.

Table 1: Characteristics of speakers from the English subcorpus of the RUEG corpus

| Group       | Adolescents     |            |           | Adults          |            |           |
|-------------|-----------------|------------|-----------|-----------------|------------|-----------|
|             | <i>N</i> (Male) | Age        | Eng.      | <i>N</i> (Male) | Age        | Eng.      |
|             |                 | Mean       | AoO       |                 | Mean       | AoO       |
|             |                 | (sd)       | Mean      |                 | (sd)       | Mean      |
| English MSs | 32 (13)         | 16.1 (1.4) | -         | 32 (13)         | 28.5 (3.9) | -         |
| German HSs  | 27 (15)         | 5.5 (1.5)  | 0.3 (0.7) | 7 (2)           | 25.3 (4.1) | 0.9 (1.5) |
| Greek HSs   | 33 (16)         | 16.3 (1.4) | 1.1 (1.8) | 32 (13)         | 29.1 (3.4) | 1.0 (1.8) |
| Russian HSs | 32 (13)         | 15.8 (1.4) | 2.5 (2.0) | 33 (11)         | 27.5 (3.3) | 3.7 (2.0) |
| Turkish HSs | 32 (10)         | 16.0 (1.6) | 2.6 (2.1) | 27 (9)          | 26.2 (4.1) | 2.2 (2.2) |
| Total       | 156 (67)        | 15.9 (1.5) | -         | 131 (48)        | 27.7 (3.8) | -         |

The data were collected using the Language Situations methodology (Wiese et al. 2025 [this volume], Wiese 2020), which allows eliciting comparable and naturalistic productions across communicative situations. Participants were shown a brief non-verbal video of a minor car accident and were asked to recount what they saw. The procedure consisted of two formality settings. In the formal setting, the participants were met by a formally dressed elicitor in an office-like room, whereas in the informal setting, the elicitor dressed casually and met the participant in a more relaxed environment, offering snacks and drinks. Prior to the informal session, the participant and the elicitor engaged in 10–15 minutes of conversation to create an easy-going atmosphere. The participant watched the video three times in total, twice in the first setting, once in the second setting, and recounted the event in spoken and written modes.

In the formal setting, participants left a voice message to a police hotline (spoken mode) and provided a written witness report to the police (written mode). In the informal setting, participants recorded a WhatsApp voice message to a friend (spoken mode) and typed a WhatsApp text message to a friend (written mode). The order of settings (formal and informal) and modes (spoken and written) was balanced among participants. English MSs accomplished all the tasks during one session. HSs completed the tasks in two sessions, one in English, their ML, and one in their HL, with a 3–5 day interval between sessions to reduce priming effects. The order of the language sessions was counterbalanced among HSs. After completing the narrative tasks, all participants filled out a language background questionnaire, which included a self-assessment of their language skills. The self-assessments showed that HSs and English MSs rated their English skills comparably high. In addition, HSs assessed their skills higher in their majority English than in their HL (Table 2).

The spoken data were transcribed in Praat (Boersma 2001) or EXMARaLDA (Schmidt & Wörner 2014). Subsequently, spoken and written data were annotated in EXMARaLDA for various phenomena of interest. The annotated data were accessed through the corpus tool ANNIS (Krause & Zeldes 2014; see Shadrova et al. 2025 [this volume] and Klotz et al. (2024) for further detail on the RUEG corpus).

To assess speakers' fluency and proficiency in English, we calculated their speech rate and lexical diversity (Table 3). Speech rate (in syllables per second) was calculated using a Praat script (de Jong et al. 2021) based on all spoken narratives in the English subcorpus. Lexical diversity was evaluated with two measures – the moving-average type-token ratio (MATTR) and measure of textual lexical diversity (MTLD), demonstrated to produce stable results for short texts

(Zenker & Kyle 2021). MATTR and MTLD were calculated based on all narratives in the English subcorpus using the lexical diversity package in Python (Kyle 2020)

Table 2: Language skill self-ratings of speakers from the English sub-corpus of the RUEG corpus

| Group                        | Spoken<br>understanding<br>Mean (sd) | Spoken<br>production<br>Mean (sd) | Written<br>understanding<br>Mean (sd) | Written<br>production<br>Mean (sd) |
|------------------------------|--------------------------------------|-----------------------------------|---------------------------------------|------------------------------------|
| <b>In majority English</b>   |                                      |                                   |                                       |                                    |
| English MSs                  | 4.92 (0.27)                          | 4.81 (0.47)                       | 4.80 (0.48)                           | 4.77 (0.46)                        |
| German HSs                   | 5.00 (0.00)                          | 5.00 (0.00)                       | 5.00 (0.00)                           | 4.97 (0.17)                        |
| Greek HSs                    | 4.97 (0.18)                          | 4.95 (0.21)                       | 4.88 (0.33)                           | 4.89 (0.31)                        |
| Russian HSs                  | 4.92 (0.32)                          | 4.92 (0.27)                       | 4.91 (0.29)                           | 4.89 (0.36)                        |
| Turkish HSs                  | 4.97 (0.18)                          | 4.93 (0.25)                       | 4.92 (0.38)                           | 4.92 (0.34)                        |
| <b>In heritage languages</b> |                                      |                                   |                                       |                                    |
| German HSs                   | 4.47 (0.56)                          | 3.65 (0.85)                       | 3.71 (1.06)                           | 2.88 (1.27)                        |
| Greek HSs                    | 4.16 (0.78)                          | 3.72 (1.05)                       | 3.46 (1.20)                           | 3.13 (1.29)                        |
| Russian HSs                  | 4.38 (0.75)                          | 3.87 (0.85)                       | 3.25 (1.05)                           | 2.83 (1.22)                        |
| Turkish HSs                  | 4.28 (0.83)                          | 3.90 (0.95)                       | 3.37 (1.05)                           | 3.30 (1.16)                        |

Table 3: Speech rate (syll/sec), MATTR, and MTLD of speakers from the English subcorpus of the RUEG corpus (in majority English)

| Group       | Speech rate Mean<br>(sd) | MATTR Mean (sd) | MTLD Mean (sd) |
|-------------|--------------------------|-----------------|----------------|
| English MSs | 3.26 (0.52)              | 0.678 (0.056)   | 38.80 (11.19)  |
| German HSs  | 3.34 (0.46)              | 0.678 (0.049)   | 40.43 (10.45)  |
| Greek HSs   | 3.23 (0.44)              | 0.670 (0.053)   | 37.66 (9.85)   |
| Russian HSs | 3.20 (0.47)              | 0.672 (0.049)   | 38.54 (10.06)  |
| Turkish HSs | 3.22 (0.45)              | 0.659 (0.050)   | 36.10 (10.63)  |

Most studies in this chapter examined the data of various subsets of speakers from the English subcorpus. The exact subset is specified in the description of each study. Furthermore, the studies used different corpus versions (e.g., 0.3.0 or 0.4.0), which is detailed in the cited papers. If a study did not employ the

English subcorpus, the characteristics of the data set are provided in the study's description.

### 3 Prosody

As noted earlier, previous studies have found that HSs tend to pattern with MSs in rating studies assessing their perceived foreign accent in the ML (e.g., Kupisch et al. 2014, Lloyd-Smith et al. 2020, but see Polinsky 2018: 138–153 for a review of segmental and suprasegmental differences). However, they often show effects of cross-linguistic influence in ML prosody (Queen 2012, van Rijswyk et al. 2017, Dehé 2018). In the RUEG project we contributed to research on prosodic interference in the ML of HSs (here majority English), investigating two phenomena – stressed pronouns and contrastive adjectives – at the interface of prosody and information structure.

One line of research looked at the prosodic realization of stressed pronouns in English MSs and HSs speaking majority English (for details, see Böttcher 2021, Böttcher & Zerbian 2020). In English, pronouns as discourse-given constituents are generally unaccented; thus, prominence on pronouns (stress) indicates focus or contrast (Krifka 2008, Selkirk 1996), as illustrated in (1) (small caps indicate prosodic prominence in the form of a pitch accent).

- (1) There was this couple walking down the sidewalk.
  - a. They<sub>GIVEN</sub> were about to cross the street.
  - b. [HE]<sub>CONTRAST</sub> had a ball while [SHE]<sub>CONTRAST</sub> was pushing a baby carriage.

In a pilot study, Böttcher & Zerbian (2020) investigated the realization of pronouns in majority English by six English MSs and 12 HSs (three Turkish, four Russian, and five Greek). The focus of this investigation was to analyze and explain the occurrence of stressed pronouns. Pronouns are not expected to be frequently stressed as given constituents (Krifka 2008, Selkirk 1996). However, the prosodic realization of given constituents is reportedly different in spontaneous compared to read speech (de Ruiter 2015). The spontaneous narrations from the RUEG corpus (two by each speaker, formal and informal) were, therefore, analyzed for the occurrence of stressed pronouns, prosodic phrasing, and information structure (following Krifka 2008, Baumann & Riester 2013, Himmelmann et al. 2018). The analysis revealed that stressed pronouns were produced frequently in spontaneous narrations by HSs and English MSs. Contrast was the most frequent factor for stressed pronoun realization ( $n = 52$ , 38%), while prosodic phrasing was also

important (separate or small phrase:  $n = 30$ , 22%; phrase finally:  $n = 16$ , 12%; and phrase initially:  $n = 33$ , 25%).

In a subsequent study, Böttcher (2021) explored whether HSs and English MSs differed in their prosodic realization of contrastively focused and not focused pronouns in majority English. This was done based on 40 narrations by 20 speakers from the RUEG corpus 0.3.0. – four English MSs and 16 HSs, with four speakers each of heritage Greek, Turkish, Russian, and German. Pronouns were analyzed for their information structure (Götze et al. 2007) and their prosodic realization (Beckman et al. 2006, Kügler & Baumann 2019). Overall, speakers show similar tendencies of realizing pronouns: unstressed if unfocused and stressed if contrastively focused (Figure 1). Overall, there is a tendency for fewer stressed pronouns in the narrations of HSs, independently of contrastive focus (see Böttcher 2021 for further detail). In the case of contrastive focus, only one pronoun in a monolingual production was unstressed, while heritage speakers leave these pronouns unstressed in about half of the cases. Since especially Turkish HSs produce contrastively focused pronouns frequently without stress, it is a topic for further research whether there is a relation between the absence of prosodic prominence and the pro-drop features of the languages in question (see Özsoy et al. 2025 [this volume]).

In sum, English MSs and HSs speaking majority English do not show qualitatively different patterns in the realization of stressed pronouns: both speaker groups stress pronouns in spontaneous speech, dependent on both information structure and prosodic phrasing (for majority German, see Zerbian & Böttcher 2019, Böttcher & Schubö 2021). However, there is a difference in distribution of unstressed pronouns: HSs do not always stress contrastively-focused pronouns in majority English and thus do not follow the pragmatic constraints that hold in English in all cases.

The second line of research investigated the relationship between prosody and contrastive focus of adjectives. The prosodic realization of modified NPs, such as *a blue car*, is dependent on focus and contrast, as shown in examples (2a-d). The placement of a pitch accent on the noun (indicated by small caps) allows for various focus interpretations while a pitch accent on the adjective (as in 2c) can only indicate narrow or contrastive focus on the adjective (Ladd 2008).

- (2) Pitch accent placement in modified noun phrases and its interpretation
  - a. wide/broad focus: Along came [a white CAR]<sub>FOCUS</sub>.
  - b. contrastive focus on noun: There was a white [VAN]<sub>FOCUS</sub> and a white [CAR]<sub>FOCUS</sub>.

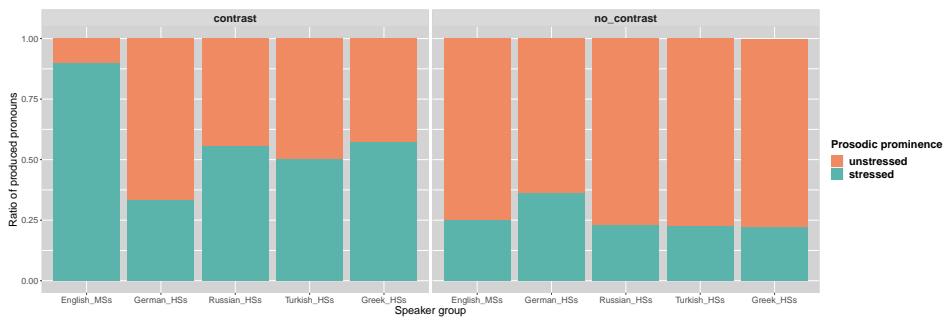


Figure 1: The occurrence of unstressed and stressed pronouns (ratio and in numbers) in the case of presence or absence of contrastive focus in English across different speaker groups in RUEG corpus 0.3.0

- c. contrastive focus on adjective: The [BLUE]<sub>FOCUS</sub> car crashed into the [WHITE]<sub>FOCUS</sub> car.
- d. contrastive focus on adjective and noun: The [BLUE]<sub>FOCUS</sub> [CAR]<sub>FOCUS</sub> rammed the [WHITE]<sub>FOCUS</sub> [VAN]<sub>FOCUS</sub>.

Böttcher (2022) investigated the accent placement in noun phrases containing adjectives in English. Data analysis included 144 narrations from the RUEG corpus (0.4.0) by adolescent English MSs and Greek and Russian HSs speaking majority English. Each group comprised 24 speakers, each producing two narratives. Narratives were searched for modified noun phrases including the lexeme *car* and were analyzed for semantic contrast (Götze et al. 2007) and for the location of pitch accents and pitch accent type (ToBI guidelines, Beckman et al. 2006). As expected, the contrastive adjectives were accentuated in most cases (Figure 2). Surprisingly, adjectives were also accented when not explicitly expressing a contrast, which might still reflect the speaker's contrasting perspective (Kaland et al. 2014). Additionally, contrastively accented adjectives were also frequently followed by an accented noun. Such a double accent pattern was found more frequently in formal speech and in the speech of Russian HSs. The double accent also occurred frequently in the Russian heritage language of these HSs indicating a possible transfer (see Zerbian et al. 2025 [this volume], 2022). The analysis of pitch accent types as high (H\*) or rising tone accents (LH\*) did not show differences across speaker groups (see Böttcher 2022 for further detail).

The results of this study reveal that English MSs and HSs speaking majority English do not show qualitatively different patterns in the prosodic realization of contrastively focused adjectives: the adjectives generally carry a pitch accent, although at the same time double accent patterns are more frequently produced

by Russian HSs. In the light of the higher use of double accents in formal narrations by MSs and Greek HSs, this can possibly be interpreted as an effort towards being linguistically – in this case prosodically – more explicit in the context of formal situations (see also next section, for HSs see Polinsky 2018: 144).

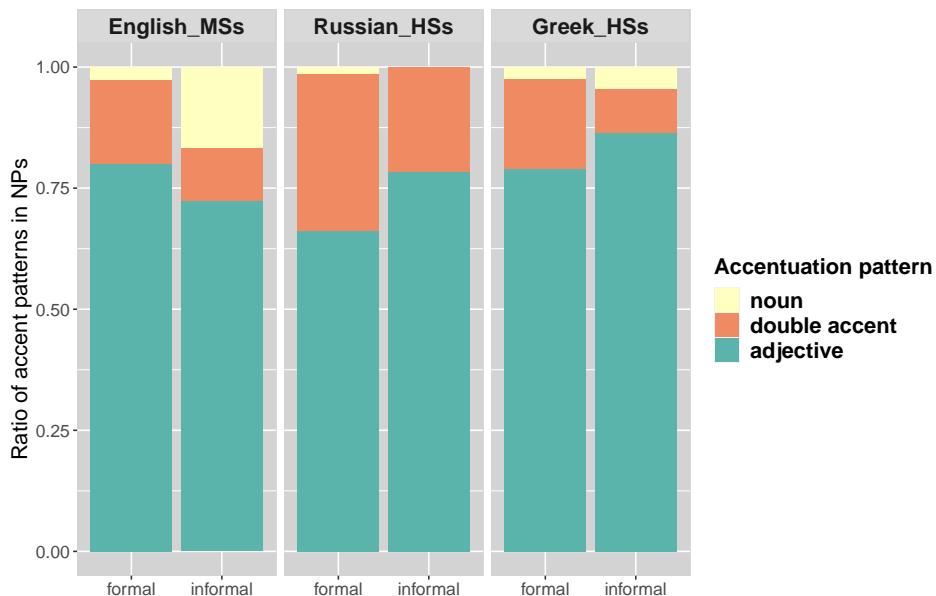


Figure 2: Distribution of accent patterns within modified noun phrases containing a contrastively focused adjective in English across speaker groups and communicative situations in the RUEG corpus 0.4.0

Overall, our results suggest that English MSs and HSs speaking majority English use similar prosodic strategies in realizing given pronouns and contrastive focus on adjectives, thus using available linguistic features along a continuum (cf. Wiese et al. 2022). Given that both structures are at the interface of prosody and semantics/pragmatics, the Interface Hypothesis (Sorace 2011) predicts these areas to be vulnerable and thus prone to change in language contact. However, our results do not lend strong support for this, at least in the ML.

## 4 Reference

Previous research has indicated that bilinguals might be more explicit than monolinguals in the domain of reference. In referent introduction, Barbosa et al. (2017)

have found that simultaneous bilingual children lexicalize more concepts in their narratives than monolinguals. In referent tracking, adult late L2 speakers tend to use full NPs instead of pronouns, compared to their L1 (Hendriks 2003, Gullberg 2006) and compared to L1 speakers of their L2 (Hendriks 2003, Yoshioka 2008). In addition, adult late L2 speakers and simultaneous bilingual children tend to overproduce and over-accept overt pronouns instead of null forms in pro-drop languages, compared to monolinguals (Serratrice et al. 2004, Sorace & Filiaci 2006, Sorace et al. 2009). Finally, adult Turkish HSs use more overt pronouns in their majority Dutch (a non-pro-drop language) compared to Dutch MSs (Azar et al. 2020). However, not all previous findings support this trend. Contemori et al. (2023) and Contemori & Ivanova (2021) found that Spanish HSs produced fewer nouns and more pronouns in their majority English compared to English MSs, and hence were less explicit.

Despite the disagreement in the literature, Pashkova and colleagues hypothesized that HSs might be more explicit either due to their frequent communication with L2 English speakers, who might benefit from extra detail (Polinsky 2018: 144) or due to more frequent input from L2 English speakers, who themselves might be more explicit (for results in majority Dutch see Azar et al. 2020). We conducted two studies to test this hypothesis: on referent introduction (Pashkova et al. 2020) and on referring expressions (Pashkova & Allen 2025).

In the referent introduction study, Pashkova et al. (2020) compared the number of introduced referents in a subset of the English subcorpus: 32 adolescent English MSs and 120 HSs (27 German, 31 Greek, 32 Russian, 30 Turkish). We identified 19 frequent referents, and coded each referent as introduced (lexicalized) or not in each narrative. Even though referent introduction is affected by various extra-linguistic factors (e.g., attention and memory play a role in which characters/objects will be mentioned in a narrative), it is an important first step of establishing reference (Vogels et al. 2019: 1). A higher proportion of introduced referents would be a sign of higher explicitness since the narrative would provide more detail about the events. Contrary to the hypothesis, a linear mixed model analysis showed no significant difference between HSs and English MSs in the number of introduced referents in majority English. In addition, we observed that HSs and MSs had the same effect of formality: both speaker groups introduced more referents in the formal setting than in the informal one.

The study of referring expressions (Pashkova & Allen 2025) consisted of two parts: modification of referring expressions and their form. In the modification part, we were interested whether HSs speaking majority English would modify referring expressions more often than English MSs, for instance, with a prepositional phrase (*the woman with the dog*) or a relative clause (*the woman who had a*

*dog*). A higher proportion of modified referring expressions would be an indication of higher explicitness because it would provide more information about the referent of the NP. We conducted a linear mixed effects model analysis based on the full English subcorpus. Contrary to our expectations, there was no significant difference between HSs and English MSs in the proportion of modified referring expressions. In addition, we observed that HSs and MSs had the same effect of formality: both speaker groups modified referring expressions more often in the formal setting than in the informal one.

In the part concerning the form of referring expressions, we hypothesized that HSs speaking majority English would be more explicit than English MSs in their use of expressions for referents that have been previously introduced; new referents are typically introduced by noun-headed NPs, so they do not exhibit much variation in referent form. We examined three types of referring expressions: noun-headed NPs (*the woman had a dog*), pronouns (*she also had some groceries*), and null anaphora (*she got scared and Ø dropped her groceries*). Each referent from the list of 19 frequent referents was annotated for its type of referring expression, irrespective of its syntactic position (i.e., subject, object, object of a preposition, etc.). Following Azar et al. (2020), we conducted two comparisons: (1) noun-headed NPs vs. pronouns and (2) pronouns vs. null anaphora. We expected that HSs would use more noun-headed NPs than MSs in the first comparison and more pronouns in the second comparison.

The first comparison was performed on the full English subcorpus, where each referent was annotated for its type of referring expression. The second comparison was done on a subset of the English subcorpus that included 40 English MSs, 42 Turkish and 40 Russian HSs (each speaker group equally split between adolescents and adults). The subsetting was done due to the necessity to annotate each referent for the type of clause it was used in – main or subordinate – since null anaphora can occur only in coordinate main clauses.

The confirmatory analyses with linear mixed effects models did not indicate a significant difference between MSs and HSs combined, neither in the noun-headed NP vs. pronoun comparison nor in the pronoun vs. null anaphora one. This shows that HSs in our sample are not more explicit than MSs across the board.

However, an exploratory analysis of noun-headed NP vs. pronoun referring expressions showed an interesting pattern that we had not previously hypothesized. This pattern has to do with formality and with a distinction that speakers make between two types of referent tracking – maintenance and reintroduction, which in our study are defined based on the syntactic roles of the referent in the current and previous finite clauses (following Hickmann & Hendriks 1999,

Serratrice 2007, Perniss & Özyürek 2015). The definitions of maintenance and reintroduction are presented in Table 4, along with examples.

Table 4: Definitions and examples of maintenance and reintroduction

| Previous clause                               | Current clause                                  | Referent tracking type   |
|---|---|--|
| Subject<br><i>The dog ran.</i>                | Any syntactic role<br><i>The driver saw it.</i> | Maintenance:<br>Referent <i>dog</i> in current clause is maintained      |
| Non-subject<br><i>The driver saw the dog.</i> | Non-subject<br><i>He didn't want to hit it.</i> | Maintenance  |
| Non-subject<br><i>The driver saw the dog.</i> | Subject<br><i>The dog ran.</i>                  | Reintroduction:<br>Referent <i>dog</i> in current clause is reintroduced |
| Absent<br><i>The driver turned right.</i>     | Any syntactic role<br><i>He saw the dog.</i>    | Reintroduction   |

The exploratory analysis individually compared German, Greek, Russian, and Turkish HSs to English MSs, as opposed to comparing all HSs together to MSs in the confirmatory analysis. Although there was considerable variability across speakers, the results (Figure 3) indicated two trends: first, in the formal setting, there was a larger difference between the proportions of full NPs in maintenance and reintroduction in narratives by German, Russian and Turkish HSs compared to English MSs (difference of 0.54 for German HSs, 0.48 for Russian HSs, 0.53 for Turkish HSs vs. 0.41 for English MSs). Second, in the informal setting, Russian and Turkish HSs produced significantly more NPs both in maintenance and reintroduction compared to English MSs. German and Greek HSs followed the same pattern but the difference did not reach significance.

These findings suggest that HSs taken together as one group are not more explicit than MSs across all contexts. The situation is more nuanced: when speaking or writing formally, some HS groups tend to separate maintenance and reintroduction more sharply than MSs. This is similar to a trend that we will present in Section 6 – there, heritage speakers show a tendency of a sharper distinction between formal and informal settings in their use of subordinate clauses and left dislocations, in comparison to MSs.

However, HSs also show a different tendency when speaking or writing informally: some HSs are more explicit in that they use more full NPs than MSs both

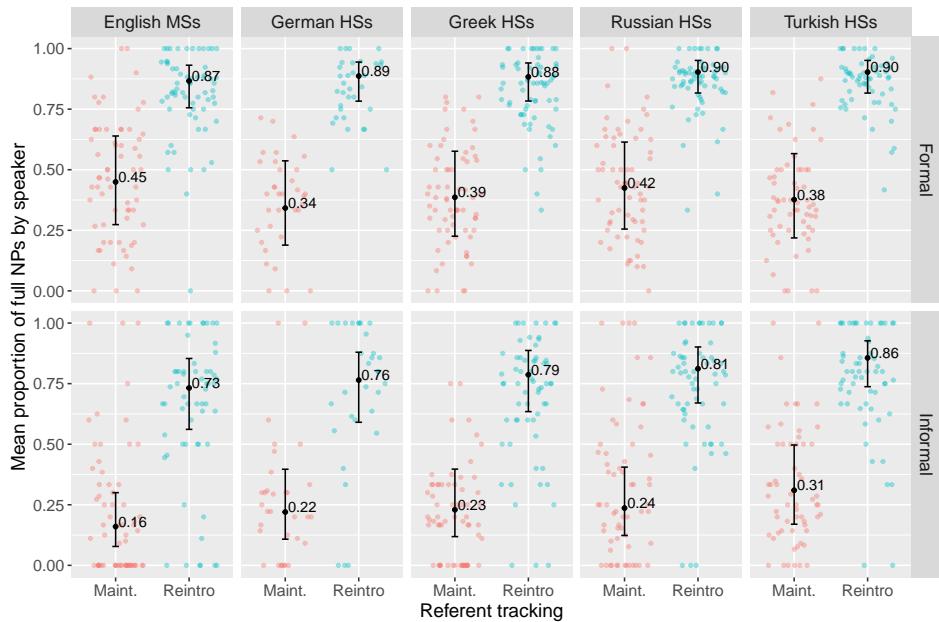


Figure 3: Predicted probabilities and individual proportions of noun-headed NPs by speaker group, formality and referent tracking. Predicted means and CIs are based on a linear model, individual proportions on raw data.

in maintenance and reintroduction. Overall, we see that HSs employ different strategies in the formal and informal settings: in a formal situation, they highlight the difference between maintenance and reintroduction, and in an informal situation they prefer to be more explicit regardless of the referent tracking status. We will come back to this trend in the general discussion.

## 5 Article use

We investigated article use by HSs speaking majority English and English MSs in narrative data from the English subcorpus and in an additional experimental study focusing specifically on article choice. The study on article use in narratives compared the proportions of unexpected articles in new and given referents based on the full English subcorpus (Wiese et al. 2022: 10–14). We extracted all strictly new referents, that is, referents that were mentioned for the first time and did not have any related anchor in the previous discourse (such as *the car* would be for *the driver*). We also extracted all given referents that were already

mentioned in the previous discourse. Next, we identified the referents that featured unexpected article choice, that is, “*the + new*” and “*a + given*” referents. All other referents, including the most expected (“*a + new*” and “*the + given*”), were marked as “other.” Comparing HSs to MSs, we observed no effect of bilingualism: the two groups produced similar proportions of unexpected “*the + new*” referents out of all new referents and of unexpected “*a + given*” referents out of all given referents. In addition, we found that HSs and MSs were influenced by spoken/written mode in a similar fashion: irrespective of the speaker group, there were significantly more unexpected “*the + new*” and “*a + given*” referents in the spoken mode than in the written mode. Overall, this study indicated no evidence of differences between HSs and English MSs in majority English, with both groups producing a small proportion of unexpected articles with new and given referents, especially in the spoken mode.

The second study on article use (Pashkova & Allen 2022) featured more constrained contexts and is the only study mentioned in this chapter that did not use the RUEG corpus data. Speakers were given a fill-in-the-blank task eliciting determiners, where each item was a short dialogue between two people (adopted from Ionin et al. 2008). This study tested the Fluctuation Hypothesis (originally developed for late L2 speakers; Ionin 2006) on adult and adolescent HSs speaking English as their ML. The Fluctuation Hypothesis posits that L2 English speakers whose L1 does not have definiteness-based articles (e.g., Korean) fluctuate between definiteness and specificity in their article choice in English: sometimes they attend to definiteness of the NP, and sometimes to specificity, resulting in productions that are different from those of L1 English speakers. On the other hand, L2 English speakers whose L1 has definiteness-based articles (e.g., Spanish) are expected to perform similarly to L1 English speakers and not exhibit any fluctuation.

HSs are similar to late L2 speakers because both speaker groups learn English while having experience of using an additional language (HL for HSs, and the only L1 for L2 speakers), so similar transfer effects could be observed in HSs as well. We hypothesized that Russian HSs in our sample ( $n = 63$ ) would exhibit the most fluctuation since their HL does not have articles, followed by Turkish HSs ( $n = 58$ ), whose HL has a complex definiteness- and specificity-based differential object marking system. Finally, we predicted that Greek HSs ( $n = 78$ ) would be the most similar to English MSs ( $n = 46$ ) since Greek has definiteness-based articles.

The results did not support the Fluctuation Hypothesis: none of the HS groups exhibited the fluctuation pattern attested in late L2 speakers. As a sole result, this would be quite unsurprising since exposure to English by the HS begins much

earlier than that of typical late L2 speakers. However, we also observed two unpredicted patterns: first, all speaker groups, including MSs, produced more unexpected articles (*the*) in the indefinite contexts (slightly above or under 1% in indefinite contexts, close to zero in definite ones). Second, Turkish HSs were the only bilingual group significantly different from English MSs: they produced more unexpected articles in both definite and indefinite contexts (*the* in definite and *a* in indefinite).

The first unpredicted pattern is similar to the finding from the narrative study described above: there the percentage of unexpected articles was also numerically higher for the new referents (= indefinite contexts) than for given ones (= definite contexts), even though this comparison was not statistically tested. However, it is hard to attribute these two results to the same cause since the narrative task and the fill-in-the-blank task were quite different. In the narrative task, it is plausible that speakers fell back to their own perspective of being able to uniquely identify the referents (since they were familiar with the characters and objects in the story after watching the video) and did not consider the addressee's perspective. This is especially likely given the fact that unexpected articles were produced more often in the spoken than in the written mode because the spoken mode does not offer a possibility to plan ahead and revise. In the fill-in-the-blank task, this explanation does not seem to apply: first, the task was written and un-timed, removing any performance pressure, and second, the speakers did not have a perspective of being familiar with the referents: they were reading the items for the first time. In this case, a possible explanation would be that speakers found the task too easy and engaged in creating background contexts that would justify the use of definite articles in indefinite contexts (e.g., prior shared knowledge or experience between the interlocutors in the dialogues). This would explain why indefinite articles were rare in definite contexts: once the uniqueness of the referent has been established, it seems almost impossible to create a context where it would disappear.

The second unpredicted pattern, the difference between Turkish HSs and English MSs, is not straightforward to explain. One possibility is that Turkish HSs engaged in creating the background contexts that justify definite articles more often than any other group. Another potential account is that acquiring a definiteness-based system in English along with a definiteness- and specificity-based system in Turkish led to the formation of a unique system that is different from both. This finding requires further investigation to see if the pattern persists in other speaker samples, and if so, to understand its underlying cause.

Summing up, in the area of article use we observed no difference between the HSs and English MSs in majority English when examining the narrative data,

and more unexpected articles produced by Turkish HSs than English MSs in an experimental task (with no difference between other HS groups and MSs). At the same time, all speakers, including English MSs, produced a small proportion of unexpected articles in indefinite contexts, which might point to task effects.

## 6 Syntax

In the domain of syntax, we compared HSs speaking majority English and English MSs for three phenomena – the use of various clause types across registers, the use of left dislocations across registers, and word order variation caused by information status of referents.

We conducted two studies on clause types across registers. Pashkova et al. (2022) compared the use of independent main clauses (*I was walking down the street. I saw a couple.*) and coordinate main clauses (*I was walking down the street, and I saw a couple.*) in a subset of the English subcorpus: 20 adolescent English MSs and 20 German HSs. Since previous research has focused on clause-type variation mainly in the HL of HSs (Keskin 2025 [this volume], Özsoy et al. 2022, Schleppegrell & Colombi 1997, Sanchez Abchi & de Mier 2017), we aimed to create a more comprehensive picture of HSs' repertoires and examine their ML as well.

We did not observe a significant difference between German HSs and English MSs: both groups used independent main clauses and coordinate main clauses with similar frequencies. In addition, HSs and MSs alike used more independent main clauses in written narratives than in spoken ones, and more coordinate main clauses in spoken narratives than in written ones. Finally, both speaker groups showed the same effect of formality: they used more coordinate main clauses in informal settings than in formal ones.

The second study on the use of clause types (Tsehayé et al. 2021) looked into the use of subordinate clauses in a larger subset of the English subcorpus: 32 adolescent English MSs and 27 adolescent German HSs. We found that bilinguals and monolinguals used subordinate clauses differently depending on the formality and mode (Figure 4). German HSs made a distinction between the formal and informal setting in both spoken and written modes, while English MSs made a distinction between the settings only in the written mode.

Our second syntactic structure of interest is left dislocation (LD), in which a noun phrase precedes its core clause and the argument position within the core clause is filled with a pronoun that is co-referential with the dislocated noun phrase (e.g., *The woman<sub>NP</sub>, she PRONOUN spilled her groceries*). LDs are indicative of an informal register (Keenan 1977, Geluykens 1992; see Bunk et al. 2025 [this volume] for a comparative review of LD in English, German, and Russian).

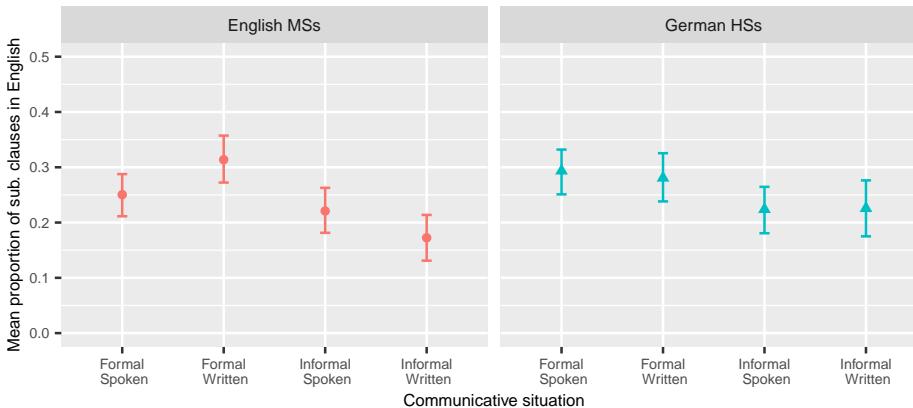


Figure 4: Mean proportions of subordinate clauses in English by speaker group and communicative situation. Means and bootstrapped CIs are based on raw data.

Pashkova et al. (submitted) hypothesized that the use of LD in the majority English of HSs might be different than that of English MSs because previous studies have found differences in the frequency of LD use by French-English bilinguals and English MSs (Hervé et al. 2016, Tagliamonte & Jankowski 2019). We also hypothesized that HSs may approach the formality constraint on LD usage differently from English MSs, given that divergent approaches to formality have been observed in other register-dependent phenomena, such as subordinate clauses (Tsehay et al. 2021, see above).

We investigated the use of LD in the full English subcorpus. The findings reveal that in the formal setting, Greek and Turkish HSs use fewer LDs than English MSs, while German and Russian HSs do not show any significant difference from the MSs. In the informal setting, Turkish HSs use more LDs than English MSs, while Greek, German, and Russian HSs do not demonstrate any significant difference from the MSs.

As to the difference between the formal and informal settings, Greek and Turkish HSs have a dissimilar approach to the formality distinction in comparison to English MSs. These HSs use slightly more LDs in informal than formal settings, whereas English monolinguals use slightly fewer LDs in informal than formal settings. That is, English monolinguals have a reverse formality effect compared to Greek and Turkish HSs. In fact, Greek and Turkish HSs adhere more strictly to the expected pattern of using LDs predominantly in the informal context. German and Russian HSs do not differ from English MSs statistically, even though

numerically they pattern with Greek and Turkish HSs because they have slightly more LDs in the informal setting than in the formal one (Figure 5). Also noticeable in Figure 5 is that Turkish HSs differentiate the formalities quite strictly, while English MSs and other HS groups do not show such a strict differentiation.

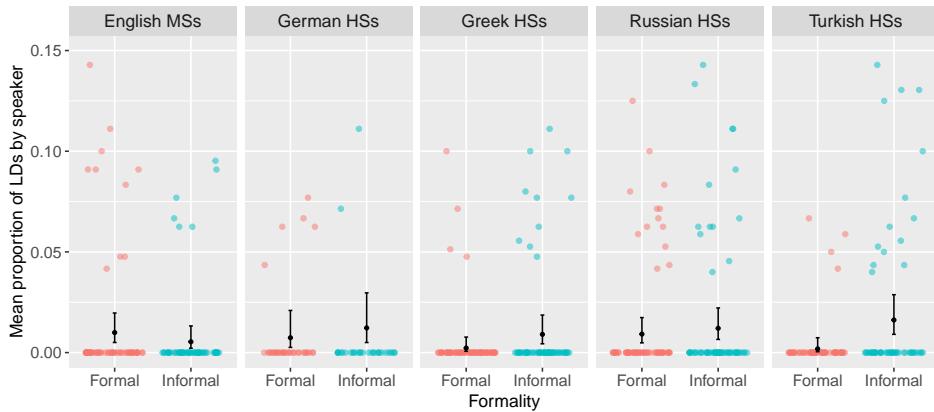


Figure 5: Predicted probabilities and individual mean proportions of LDs by speaker group and formality. Predicted means and CIs are based on a linear model, individual proportions on raw data. The y-axis is zoomed from the original size of 0–30% to the size of 0–15% to make the model predictions more visible. When zooming, 10 data points that were above the 15% mark were removed: two in English MSs, three in German HSs, two in Russian HSs, and three in Turkish HSs.

In summary, the LD study shows that some HS groups use LDs differently in their majority English compared to English MSs, while other groups do not. The difference does not lie in the overall frequency of LDs, but in the way speakers distribute LDs in formal and informal settings: some HSs are actually more in line with the expected patterns reported in the literature than English MSs. It is a matter of future research to determine the source of this difference: while we cannot exclude cross-linguistic influence from the Greek and Turkish HLs, a general HS pattern seems more likely for two reasons. First, it is quite surprising that cross-linguistic influence from two HLs would make HSs' productions more aligned with the expected LD pattern in majority English. Second, all four HS groups trend in the same direction numerically, which is also somewhat unlikely if the HL influence was the source of difference.

The third phenomenon in the domain of syntax that we examined is the variability in word order caused by the information status of referents (see Bunk

et al. (2025 [this volume]) for more on word order dynamics in language contact and Tsehayé et al. (2025 [this volume]) for word order phenomena in heritage German). English follows a fixed SVO word order, which can be changed to accommodate pragmatic factors, leading to various information packaging constructions (Huddleston 2002: 67–68). One such factor is the given-before-new principle, which suggests that a clause should begin with given/background information and end with new information (Biber et al. 2021: 888–889). However, this principle creates a conflict when it comes to new subjects in English. While the SVO word order demands that a new subject be placed at the beginning of a clause, it would be more helpful to position it closer to the end because the referent is new.

To address this conflict, English speakers may use different constructions that deviate from the SVO order, such as existential *there* (e.g., *There was a woman unloading her groceries on the other side of the street*), presentational *there* (e.g., *There stood a woman unloading her groceries on the other side of the street*), locative inversion (e.g., *On the other side of the street was a woman unloading her groceries*), or passivization (e.g., *Some groceries were being unloaded by a woman*).

Research suggests that bilingual English speakers may use information-packaging constructions differently than English MSs, potentially due to cross-linguistic influence. For instance, speakers of Singapore or Jamaican English use fewer existential *there* constructions than speakers of monolingual varieties (Winkle 2015).

Building on these previous findings, Pashkova & Allen (in preparation) investigated the syntactic structures employed for the introduction of new subjects in majority English. We asked whether HSs of Russian and Turkish, languages with more flexible word orders, utilize more non-SVO structures than English MSs, potentially due to cross-linguistic influence from their heritage languages. To test this hypothesis, we compared the syntactic patterns employed for the introduction of new subjects, based on a subset of the English subcorpus. It included 82 HSs (40 Russian and 42 Turkish) and 40 English MSs. Each subject was annotated for its information status (new or given) as well as the syntactic structure it appeared in. The syntactic structures were classified into two categories: SVO-type structures (SVO and non-inverted copular clauses), and non-SVO-type structures, encompassing existential and presentational *there*, locative and non-locative inversions, right and left dislocations, passives, questions, and *it-* and pseudo-clefts.

The results revealed no significant difference in the frequency of non-SVO structures employed by Russian and Turkish HSs and English MSs. However, a distinction emerged in the frequency of non-SVO structures in new and given

subjects: new subjects appeared in non-SVO structures more often compared to given subjects. There was no interaction between the subject information status and the speaker group (HSs vs. MSs). These findings indicate that new subjects are more commonly associated with information packaging constructions (i.e., non-SVO structures) than given subjects, and both HS and MSs in our sample follow this trend in majority English in a similar way.

Summing up, in the domain of syntax we have not observed qualitatively different patterns in HSs' and MSs' productions. If there are dissimilarities between bilinguals and monolinguals, they are connected to different use of syntactic structures in formal and informal contexts (subordinate clauses in German HSs, LDs in Greek and Turkish HSs).

## 7 Discourse

Discourse organization refers to the way in which speakers organize their ideas in a coherent way to convey a meaningful message. This is typically achieved through the use of various linguistic devices such as discourse markers, as well as through macro-level textual components such as openings and closings. Discourse organization can vary across countries and cultures, social groups, age groups, genre, speech communities, and registers, with sometimes very subtle differences (e.g., Fox Tree 2010). Previous studies have shown that HSs who are highly proficient in both their HL and ML exhibit similar patterns of discourse organization as MSs (Silva-Corvalán 2014), while those who are less proficient in one or both languages often display different patterns (Montrul 2010).

To examine this further, we investigated whether the Greek, Turkish, German and Russian HSs use similar organization strategies as English MSs in their majority English, and whether any differences are a result of cross-linguistic influence. We focus on both openings and closings and discourse markers in the English subcorpus of the RUEG corpus (see Katsika et al. (2025 [this volume]) for an analysis of openings and closings in various languages and Labrenz et al. (2025 [this volume]) for an analysis of discourse markers).

Discourse openings and closings are defined as any material preceding or following the core phase of the discourse (for English, see Schegloff & Sacks 1973, Schegloff 1968, 2007; for other languages, see Pavlidou 2014, Luke & Pavlidou 2002). By core phase, we mean the part of the text that constitutes the narrative, which in our case is the description of the events presented in the video. They are used to define textual boundaries that set the main text apart from the framing, introduce or close off a new topic, orient the addressee to what is coming next,

make the speaker's stance clear, and interpret and evaluate linguistic material (Labov 1972, Tolchinsky et al. 2002, Berman & Katzenberger 2004).

We identified the functions of openings and closings within three main categories: intersubjective (oriented towards the imaginary interlocutor), subjective (oriented towards the speaker), and textual (focusing on the organization of the text). Intersubjective, subjective and textual functions and their respective examples are presented in Table 5.

Table 5: List of intersubjective, subjective and textual functions with examples

| Functions              | Examples  |
|------------------------|---|
| <i>Intersubjective</i> |   |
| Greeting               | <i>hello</i>  |
| Summoning              | <i>bro</i>  |
| Initial inquiry        | <i>How are you?</i>                                 |
| Identification         | <i>My name is...</i>                                |
| Justification          | <i>I am calling to report an accident</i>           |
| Attention getter       | <i>You wouldn't believe what happened!</i>          |
| Reaction seeking       | <i>Can you call me back?</i>                        |
| Inquiry response       | <i>I'm fine too!</i>                                |
| Give advice            | <i>Be careful when you're driving</i>               |
| Common ground          | <i>Remember that day...</i>                         |
| Valediction            | <i>bye</i>  |
| <i>Subjective</i>      |   |
| Reaction               | <i>whoa</i>   |
| Evaluation             | <i>Totally crazy</i>                                |
| Personal statement     | <i>I was walking as I always do in the mornings</i> |
| <i>Textual</i>         |   |
| Initialize narrative   | <i>Here's what I saw:</i>                           |
| Contextual             | <i>I just saw an accident</i>                       |
| Episodic               | <i>It happened at 12pm today</i>                    |
| Resolution             | <i>Nobody got hurt</i>                              |
| Encapsulation          | <i>It was all an accident</i>                       |
| Coda                   | <i>That's the end</i>                               |

A descriptive frequency analysis revealed that all groups of speakers in our

sample used more openings (~80%) than closings (~50%) across all communicative situations in the majority English texts. The overwhelming use of an opening in speakers' narrations indicates that openings were used as a means to establish the communicative situation (e.g., formal vs. informal). In contrast, the relatively low percentage of closings shows that closings have a more optional character in speakers' narratives.

German and Russian HSs showed the highest frequency of openings and closings, while Turkish HSs showed the lowest frequency including many texts in which they only described the video events with no introductory or closing material (Figure 6). This indicates that although English is the dominant language for all speaker groups, there are differences in macro discourse strategies across groups which may originate from contact with the HL (see Katsika et al. (2025 [this volume]) for analyses of openings and closings in Turkish MSs and HSs) or from other sources that may be uncovered in further research. In terms of functions, we find that intersubjective functions are mostly used in informal settings whereas textual functions are primarily used in formal settings.

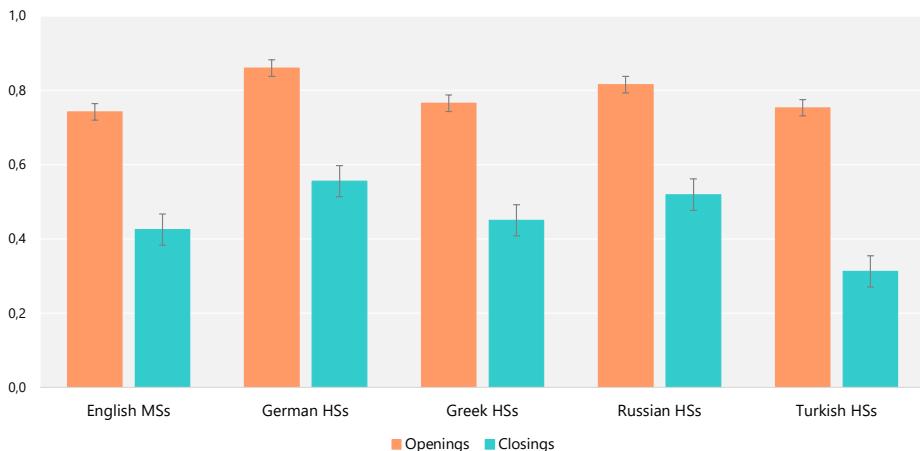


Figure 6: Frequency distribution of openings and closings in the English texts. Error bars represent standard errors. Means and error bars are based on raw data.

Discourse markers also play a crucial role in the organization of discourse. We analyzed the frequency and discourse functions of two discourse markers in the full English corpus: *so* and *yeah*. The literature reports that *so* as a discourse marker is mostly used with an inferential function to indicate "result" or "inference" (Schiffrin 1987). The MSs in our corpus used *so* with a higher frequency than all groups of HSs in majority English, although the functions of *so*

are distributed similarly across all speaker groups (see Labrenz et al. (2025 [this volume]) for a more detailed analysis of the German *so* and other discourse markers). In terms of function, *so* was primarily used across groups as a connector (3a), then to initialize the narrative (3b), and third most frequently to switch the theme (3c).

- (3) a. it was cause this guy like lost control of his soccer ball and I know you love soccer *so* thought you'd be interested (USmo11FE\_isE)  
b. *so* i just witnessed an accident it seemed like a minor accident there was a woman on the right side of the street opening her trunk (USmo05FE\_isE)  
c. and then a dog starts barking and he runs in the street *so* this guy is driving down the street and he sees a dog (USmo10ME\_isE)

The analysis of the discourse marker *yeah* showed that its frequency is much more restricted than that of *so*. Interestingly, *yeah* is used most frequently by German HSs, followed by MSs, and then Greek and Turkish HSs. The elevated use of *yeah* by German HSs may stem from the high frequency of *ja* in the German language, as well as the functional similarity between *yeah* and the German *ja* ('yeah') (Trotzke et al. 2020).<sup>2</sup> The most frequent discourse functions of both *ja* and *yeah* across groups are initializing and ending the narrative, evaluation, and hesitation. (4) shows an example of *yeah* ending a narrative.

- (4) the car stop and hit the other car *so* it's really all that happened um *yeah* (USmo69ME\_isE)

In sum, we see that the patterns of use of discourse openings and closings in majority English are similar across HS and MS groups, although the frequency of use differs somewhat. This extends to the discourse markers *so* and *yeah* as well; both are used with similar functions across groups but with somewhat different frequencies. In the case of *yeah*, this appears to be driven by cross-linguistic influence.

## 8 Discussion

We began this chapter by asking whether the majority English of adolescent and adult HSs showed any dynamic patterns due to bilingualism or whether it looked

<sup>2</sup>Although there is phonological similarity between the English *yeah* and the Turkish discourse marker *ya*, we did not find inflated use of *yeah* in Turkish HSs. This possibly stems from the fact that the English *yeah* and the Turkish *ya* do not have full functional overlap (as is the case between the German *ja* and the English *yeah*) (Şimşek 2012).

similar to patterns found in English MSs. In short, we found that the four groups of HSs patterned remarkably similarly to each other and to the English MSs in most respects, with some quantitative differences. We elucidate these similarities and differences in the following paragraphs.

### 8.1 Similarities

We begin with the similarities between HSs and MSs in majority English. In the area of phonology, HSs performed like MSs in their use of stressed pronouns based on information structure and phrasing, and in their use of pitch accents marking contrastive adjectives prosodically.

In terms of reference, they introduced and modified new referents with the same frequency as MSs, and, similarly to MSs, introduced and modified more referents in formal than informal settings. They also patterned like MSs in the proportion of NPs vs. pronouns and pronouns vs. null forms that they used to realize already-introduced referents, as well as in using more NPs for referent re-introduction than for referent maintenance.

In the domain of article use, HSs produced unexpected structures at the same rate as MSs (regarding use of both *a* for definite and *the* for indefinite), and produced more unexpected articles in indefinite than definite contexts and in spoken than written modes (both with similar frequency to MSs). For syntax, they produced independent vs. coordinate main clauses at similar frequencies to MSs, and patterned like MSs in preferring independent main clauses for the spoken mode vs. coordinated main clauses for the written mode, and in producing more coordinated main clauses in informal than formal settings. Additionally, HSs used a similar frequency of non-SVO structures to introduce new subjects compared to MSs. Finally, HSs structured their discourse similarly to MSs: all of the speaker groups produced more openings than closings, and preferred inter-subjective functions for openings and closings in the informal setting vs. textual functions in the formal setting. The functions of the discourse markers *so* and *yeah* were also distributed similarly across the HS and MS groups. In sum, we found no differences between our four HS groups and English MSs in many patterns of majority English.

It may seem surprising that our results show so many similarities in majority English between HSs and MSs, given a large number of differences found in the literature reviewed in the Introduction. We offer several possible reasons for this. One potential reason is that our studies focus on a particular type of heritage speaker for whom many demographic factors favor a positive language learning situation. They were all fluent and literate in the HL (which is not the case in

several studies), typically from mid to high SES families (50–89% of mothers in each speaker group had a postsecondary degree), and without indication of traumatic immigration history in their recent background (e.g., refugee from a war zone). In addition, 91% of them started acquiring English from the age of 5 or earlier, and nearly half from birth, so they had very early exposure to the ML in a society where the ML is highly dominant.

Another main factor relates to the methods of our study. Most of our data derive from naturalistic narrative productions, for which the cognitive demands are relatively low. The speaker is also in control of the language they produce, so they can easily keep to language that is comfortable for them. It may be that experiments or other tasks with higher cognitive demands, or focusing more narrowly on a particular structure, would provide more possibility to see dynamic patterns resulting from bilingualism. For example, Bylund et al. (2012) used materials designed to be particularly challenging for the HS participants, and Lee et al. (2011) and Scontras et al. (2017) focused narrowly on differences in scope phenomena in Korean/Mandarin vs. English that appear relatively rarely in narrative contexts such as those we elicited. Further, we examined only a subset of possible structures; examining other structures may lead to different results.

In general, however, we have added evidence to the literature that several structures in bilinguals' more dominant language (here, ML) are not influenced by cross-linguistic influence or general bilingualism effects. In addition, we have shown that for at least our sample of HSs, bilingualism is not an impediment to performance in the majority language that is consistent with that of MSs (Flege 1999, Ventureyra et al. 2004, MacWhinney 2005). This suggests that there is no reason to prohibit heritage speakers from maintaining their HL, and supports the view that heritage speakers are indeed native speakers of the ML rather than "deficient versions" of speakers of the ML (as already expressed in Rothman & Treffers-Daller 2014, Montrul 2016, Kupisch & Rothman 2018, Tsehayé et al. 2021, Wiese et al. 2022).

## 8.2 Differences

However, the productions of HSs in majority English also differed significantly in several respects from those of English MSs. Many of the differences are consistent with the idea put forth in the literature that HSs might be more explicit than MSs. In the domain of phonology, Russian HSs used double-accented structures in contrastive accent situations – accenting both the adjective and the noun – where MSs and Greek HSs typically accented only the adjective. The occasional use of double accent by MSs and HSs of Greek in formal situations gives an

indication that this phenomenon could be related to an effort to be prosodically more explicit by making the structure more salient (e.g., producing an accent on more lexical constituents than expected). Both patterns suggest that they were trying to make the structures more salient than MSs did. In referent use, Russian and Turkish HSs used more NPs in the informal setting than MSs – again a sign of greater explicitness. In one area, however, we found that HSs were less explicit than MSs: they stressed fewer contrastively-focused pronouns than MSs.

In other cases, the HS bilinguals seemed to make a stricter distinction between different linguistic and extra-linguistic contexts than MSs did. In the area of reference, for example, German, Russian, and Turkish HSs had a larger difference in the proportion of NPs between maintenance and reintroduction of referents in the formal setting than MSs did, signaling a stricter distinction between these functions. We also found an exaggerated distinction in HSs vs. MSs in their use of subordinate clauses depending on the formality and mode: German HSs distinguished between the formal and informal setting in both spoken and written modes, while MSs only made a distinction between the settings in the written mode. Finally, when producing left dislocation structures, Turkish HSs differentiated between the formal and informal settings more than MSs did.

Related to the stricter distinction between extra-linguistic contexts, in some situations HSs appeared to follow the differentiation between the contexts that is expected from the literature more strictly than MSs. In the case of left dislocations, all HS groups used more LDs in the informal setting than in the formal one (in line with the literature), while MSs had the reverse pattern – they produced slightly more LDs in the formal setting than the informal one. While the reversal of the pattern is visible in all HSs, it reached significance only in the Greek and Turkish HSs (see similar results for left dislocations in majority German in Bunk et al. (2025 [this volume])).

The finding in our data that HSs are more explicit in some contexts and tend to differentiate contexts more strictly is striking, and also consistent with other reports in the literature noted earlier (e.g., Serratrice et al. 2004, Barbosa et al. 2017, Polinsky 2018, Azar et al. 2020). It suggests that the dominant language of bilinguals can show signs of general bilingualism effects, in addition to potential cross-linguistic influence from the less dominant language.

Why might this be the case for our data? It may well be that HSs want to show that they can speak the ML “properly”, especially in a test-like situation. It could also be an effect of their typical environment – that HSs are more exposed to input containing explicit forms from L2 speakers than MSs are, as suggested by Azar et al. (2020), or that they are more explicit to make their speech clearer to their L2 interlocutors, as suggested by Polinsky (2018). Finally, it may be that

HSs choose to be explicit in a more nuanced way, such as signaling different contextual factors to the fullest degree, rather than in the straightforward way of simply being more explicit at all costs. This may explain why we see effects of explicitness or greater differentiation in some contexts but not in others where we might have expected them.

Cross-linguistic influence does not seem to be a main driver of dynamic patterns in our data, but may account for two of our findings. First, Turkish HSs produced more unexpected articles in our fill-in-the-blank study than did MSs, which may be related to the fact that Turkish has a very different article system than English. Second, German HSs used the discourse marker *yeah* more frequently than MSs, which we hypothesize results from its functional similarity with the German *ja* "yeah". However, the reduced use of the discourse marker *so* by HSs compared to MSs does not seem to stem from cross-linguistic influence.

The fact that we did not see more cross-linguistic influence in our data overall may be for at least two reasons. First, while we selected structures that differ broadly across the languages in our study (e.g., Russian and Turkish have more free word order than English so we expected more use of non-SVO utterances to introduce new subjects), we did not specifically target a narrow point on which there were clear differences. Second, the type of data we collected – elicited narratives – allows speakers to avoid structures they are uncomfortable with. While elicited narratives provide ecologically valid data in various registers, they are likely not the ideal type of data to see effects of cross-linguistic influence that are necessarily subtle due to the high proficiency of the speakers.

Another perplexing point is that we sometimes found dynamic patterns in one or two of the four groups of HSs but not all of them, and the groups that differed changed from study to study. For example, the German and Russian HSs patterned like the MSs in production of left dislocations while the Greek and Turkish HSs did not. Similarly, the Greek HSs patterned like the MSs in their use of noun phrases vs. pronouns to refer to given referents, but the other three groups of HSs patterned differently. On the surface, it does not seem like cross-linguistic influence is a likely explanation because there are no apparent salient differences between the languages. However, future research should carefully examine the relevant structures in data from monolingually-raised speakers of those languages. Another possible explanation for differences across groups is non-linguistic factors such as subtle SES differences or psychological factors. Nonetheless, we highlight the importance of comparison between different groups of HSs using the same task as the performance across groups allows us to see more clearly if a particular phenomenon is related to bilingualism in general or to some feature of the language(s) in question.

## 9 Conclusion

In sum, we examined the production of majority English by heritage speakers of German, Greek, Russian and Turkish in a variety of studies focusing on different interface phenomena covering several areas of language. The studies revealed many phenomena where heritage speakers and monolingually-raised speakers showed identical patterns, underlining the growing perception in the field that heritage speakers can reasonably be considered native speakers of their majority language (as well as their heritage language). We also found several differences between the productions of (certain groups of) heritage speakers and monolingually-raised speakers, most of which could be attributed to either cross-linguistic influence or some form of explicitness or greater differentiation between contexts than is present in monolingually-raised speakers. Overall, the studies reported here reveal the richness and complexity of patterns in the majority language of heritage speakers.

## Acknowledgements

We are very grateful to all the research assistants who contributed to data collection, annotation, and analysis of the majority English data reported in this chapter, as well as the preparation of this chapter: Iliuza Akhmezianova, Amy Amoakuh, Aban Alehojat, Chris Allison, Ricarda Bothe, Mert Can, Ryan Carroll, Franziska Cavar, Leah Doroski, Mary Elliott, Harshaa Gopalakrishnan, Gajaneh Hartz, Luc Henriquez, Abigail Hodge, Yuliia Ivashchyk, Janie Fink, Hanna Kim, Hannah Lee, Mark Murphy, Mariia Naumovets, Gizem Öskürçi, Sadaf Rezai, Simge Sargin Kisacik, Jasmine Segarra, Golshan Shakeebae, Selena Song, Shreya Srivastava, Rhiannon Stewart, Mateo Vargas-Nunez, Fiona Wong and Barbara Zeyer. We further thank Lea Coy for help with pre-publication formatting.

We also thank the members on the RUEG project and Mercator Fellows for their many contributions to the RUEG method and for very helpful discussions of our studies over the years of the project. We would also like to thank two anonymous reviewers and three reviewers from the RUEG project (Annika Labrenz, Onur Özsoy, and Wintai Tsehayé) for their constructive feedback and insightful comments.

The studies in this chapter were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation); the grant was awarded to the Research Unit *Emerging Grammars in Language Contact Situations* (FOR 2537), grant numbers 394837597 (P2), 394995401 (P5), 394844953 (P7), 313607803 (P8 and P9).

## References

- Alexiadou, Artemis & Vasiliki Rizou. 2023. The use of periphrasis for the expression of aspect by Greek heritage speakers: A case study of register variation narrowing. *Register Studies* 5(1). 82–110. DOI: 10.1075/rs.20022.ale.
- Alexiadou, Artemis, Vasiliki Rizou & Foteini Karkaletsou. 2022. A plural indefinite article in heritage Greek: The role of register. *Languages* 7(2). 115. DOI: 10.3390/languages7020115.
- Azar, Zeynep, Aslı Özyürek & Ad Backus. 2020. Turkish-Dutch bilinguals maintain language-specific reference tracking strategies in elicited narratives. *International Journal of Bilingualism* 24(2). 376–409. DOI: 10.1177/1367006919838.
- Barbosa, Poliana, Elena Nicoladis & Margaux Keith. 2017. Bilingual children's lexical strategies in a narrative task. *Journal of Child Language* 44(4). 829–849. DOI: 10.1017/S030500091600026X.
- Baumann, Stefan & Arndt Riester. 2013. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and meaning*, 119–162. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110261790.119.
- Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2006. The original ToBi system and the evolution of the ToBi framework. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.003.0002.
- Berman, Ruth A. & Irit Katzenberger. 2004. Form and function in introducing narrative and expository texts: A developmental perspective. *Discourse Processes* 38(1). 57–94. DOI: 10.1207/s15326950dp3801\_3.
- Bialystok, Ellen, Gigi Luk, Kathleen F. Peets & Sujin Yang. 2010. Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition* 13(4). 525–531. DOI: 10.1017/S1366728909990423.
- Biber, Douglas, Stig Johansson, Geoffrey N. Leech, Susan Conrad & Edward Finegan. 2021. *Grammar of spoken and written English*. Amsterdam & Philadelphia: John Benjamins. DOI: 10.1075/z.232.
- Boersma, Paul. 2001. Praat: A system for doing phonetics by computer. *Glot International* 5(9/10). 341–345.
- Böttcher, Marlene. 2021. Semantic vs. prosodic prominence: Pronouns realisation in spontaneous mono- and bilingual English. In *Proc. 1st International Conference on Tone and Intonation (TAI)*, 127–131. DOI: 10.21437/TAI.2021-26.
- Böttcher, Marlene. 2022. A comparison of pitch accent patterns in contrastive adjective+noun structures in bilingual Englishes. *Proceedings of Speech Prosody, Lisbon*. 807–811. DOI: 10.21437/SpeechProsody.2022-164.

- Böttcher, Marlene & Fabian Schubö. 2021. Pronoun vs. preposition: Where is stress assigned in German prepositional phrases? In *Proc. 1st International Conference on Tone and Intonation (TAI)*, 88–91. DOI: 10.21437/TAI.2021-18.
- Böttcher, Marlene & Sabine Zerbian. 2020. Stressed pronouns in spontaneous English. In *Proc. Speech Prosody 2020*, 131–135. DOI: 10.21437/SpeechProsody.2020-27.
- Bunk, Oliver, Shanley E. M. Allen, Sabine Zerbian, Tatiana Pashkova, Yulia Zuban & Erica Conti. 2025. Information packaging and word order dynamics in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 379–412. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775181.
- Bylund, Emanuel, Niclas Abrahamsson & Kenneth Hyltenstam. 2012. Does first language maintenance hamper nativelikeness in a second language? *Studies in Second Language Acquisition* 34(2). 215–241. DOI: 10.1017/S0272263112000034.
- Bylund, Emanuel, Kenneth Hyltenstam & Niclas Abrahamsson. 2021. Age of acquisition – not bilingualism – is the primary determinant of less than native-like L2 ultimate attainment. *Bilingualism: Language and Cognition* 24(1). 18–30. DOI: 10.1017/S1366728920000188.
- Chang, Charles B. 2016. Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition* 19(4). 791–809. DOI: 10.1017/s1366728914000261.
- Chen, Baoguo, Yuefang Jia, Zhu Wang, Susan Dunlap & Jeong-Ah Shin. 2013. Is word-order similarity necessary for cross-linguistic structural priming? *Second Language Research* 29(4). 375–389. DOI: 10.1177/0267658313491962.
- Contemori, Carla & Iva Ivanova. 2021. Bilingual referential choice in cognitively demanding situations. *Bilingualism: Language and Cognition* 24(1). 83–95. DOI: 10.1017/S1366728920000176.
- Contemori, Carla, Naoko Tsuboi & Alma L. Armendariz Galaviz. 2023. Referential choice in two languages: The role of language dominance. *International Journal of Bilingualism*. DOI: 10.1177/13670069231170545.
- de Jong, Nivja H., Jos Pacilly & Willemijn Heeren. 2021. PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice* 28(4). 456–476. DOI: 10.1080/0969594X.2021.1951162.
- de Ruiter, Laura E. 2015. Information status marking in spontaneous vs. read speech in story-telling tasks: Evidence from intonation analysis using GToBI. *Journal of Phonetics* 48: The impact of stylistic diversity on phonetic and phono-

- logical evidence and modeling*. Petra Wagner, Jürgen Trouvain & Frank Zimemer (eds.). 29–44. DOI: 10.1016/j.wocn.2014.10.008.
- Dehé, Nicole. 2018. The intonation of polar questions in North American (“heritage”) Icelandic. *Journal of Germanic Linguistics* 30(3). 213–259. DOI: 10.1017/s1470542717000125.
- Felser, Claudia & Seçkin Arslan. 2019. Inappropriate choice of definites in Turkish heritage speakers of German. *Heritage Language Journal* 16(1). 22–43. DOI: 10.46538/hlj.16.1.2.
- Flege, James Emil. 1999. Age of learning and second language speech. In David Birdsong (ed.), *Second language acquisition and the critical period hypothesis*, 101–131. Mahwah: Lawrence Erlbaum Associates Publishers.
- Fox Tree, Jean E. 2010. Discourse markers across speakers and settings. *Language and Linguistics Compass* 4(5). 269–281. DOI: 10.1111/j.1749-818X.2010.00195.x.
- Geluykens, Ronald. 1992. *From discourse process to grammatical construction: On left-dislocation in English* (Studies in Discourse and Grammar 1). Amsterdam & Philadelphia: John Benjamins. DOI: 10.1075/sidag.1.
- Gorba, Celia. 2019. Bidirectional influence on L1 Spanish and L2 English stop perception: The role of L2 experience. *The Journal of the Acoustical Society of America* 145(6). 587–592. DOI: 10.1121/1.5113808.
- Götze, Michael, Thomas Weskott, Cornelia Endriss, Ines Fiedler, Stefan Hinterwimmer, Svetlana Petrova, Anne Schwarz, Stavros Skopeteas & Ruben Stoel. 2007. Information structure. *Interdisciplinary Studies on Information Structure: ISIS* (7). 147–187.
- Gullberg, Marianne. 2006. Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning* 56(1). 155–196. DOI: 10.1111/j.0023-8333.2006.00344.x.
- Hendriks, Henriet. 2003. Using nouns for reference maintenance: A seeming contradiction in L2 discourse. In Anna Giacalone Ramat (ed.), *Typology and second language acquisition*, 291–326. Berlin: De Gruyter Mouton. DOI: 10.1515/978310891249.291.
- Hervé, Coralie, Ludovica Serratrice & Martin Corley. 2016. Dislocations in French–English bilingual children: An elicitation study. *Bilingualism: Language and Cognition* 19(5). 987–1000. DOI: 10.1017/S1366728915000401.
- Hickmann, Maya. & Henriet Hendriks. 1999. Cohesion and anaphora in children’s narratives: A comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language* 26(2). 419–452. DOI: 10.1017/S030500099003785.

- Himmelmann, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology* 35(2). 207–245. DOI: 10.1017/s0952675718000039.
- Hohenstein, Jill, Ann Einsenberg & Letitia Naigles. 2006. Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish–English bilingual adults. *Bilingualism: Language and Cognition* 9(3). 249–261. DOI: 10.1017/S1366728906002616.
- Huddleston, Rodney. 2002. Syntactic overview. In Rodney Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 43–70. Cambridge: Cambridge University Press. DOI: 10.1017/9781316423530.003.
- Ionin, Tania. 2006. This is definitely specific: Specificity and definiteness in article systems. *Natural Language Semantics* 14(2). 175–234. DOI: 10.1007/s11050-005-5255-9.
- Ionin, Tania, Maria Luisa Zubizarreta & Salvador Bautista Maldonado. 2008. Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua* 118(4). 554–576. DOI: 10.1016/j.lingua.2006.11.012.
- Kaland, Constantijn, Emiel Krahmer & Marc Swerts. 2014. White bear effects in language production: Evidence from the prosodic realization of adjectives. *Language and Speech* 57(4). 470–486. DOI: 10.1177/0023830913513710.
- Katsika, Kalliopi, Annika Labrenz, Kateryna Iefremenko & Shanley E.M. Allen. 2025. Discourse openings and closings across languages in contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 527–576. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775189.
- Keenan, Elinor O. 1977. Why look at planned and unplanned discourse? In Elinor Ochs & Tina Bennet (eds.), *Discourse across time and space*, 1–42. University of Southern California, Los Angeles.
- Keskin, Cem. 2025. Balkan Turkic as a model for understanding contact-induced change in Turkish. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 287–323. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775175.
- Klotz, Martin, Rahel Gajaneh Hartz, Annika Labrenz, Anke Lüdeling & Anna Shadrova. 2024. Die RUEG-Korpora: Ein Blick auf Design, Aufbau, Infrastruktur und Nachnutzung multilingualer Forschungsdaten. *Zeitschrift für germanistische Linguistik* 52(3). 578–592. DOI: 10.1515/zgl-2024-2026.
- Krause, Thomas & Amir Zeldes. 2014. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1). 118–139. DOI: 10.1093/llc/fqu057.

- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3-4). 243–276.
- Kügler, Frank & Stefan Baumann. 2019. *Annotationsrichtlinien DIMA*. 4th. <http://dima.uni-koeln.de/>.
- Kupisch, Tanja. 2013. A new term for a better distinction? A view from the higher end of the proficiency scale. *Theoretical Linguistics* 39(3-4). 203–214. DOI: 10.1515/tl-2013-0012.
- Kupisch, Tanja. 2014. Adjective placement in simultaneous bilinguals (German–Italian) and the concept of cross-linguistic overcorrection. *Bilingualism: Language and Cognition* 17(1). 222–233. DOI: 10.1017/S1366728913000382.
- Kupisch, Tanja, Dagmar Barton, Katja Hailer, Ewgenia Klaschik, Ilse Stangen, Tatjana Lein & Joost van de Weijer. 2014. Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal* 11(2). 123–150. DOI: 10.46538/hlj.11.2.2.
- Kupisch, Tanja, Alyona Belikova, Öner Özçelik, Ilse Stangen & Lydia White. 2017. Restrictions on definiteness in the grammars of German-Turkish heritage speakers. *Linguistic Approaches to Bilingualism* 7(1). 1–32. DOI: 10.1075/lab.13031.kup.
- Kupisch, Tanja & Jason Rothman. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5). 564–582. DOI: 10.1177/1367006916654355. (27 April, 2023).
- Kyle, Kristopher. 2020. *Lexical-diversity*. Version 0.1.1. <https://pypi.org/project/lexical-diversity/>.
- Labov, William. 1972. The transformation of experience in narrative syntax. In William Labov (ed.), *Language in the inner city: Studies in the black English vernacular* (Conduct and communication 3), 355–396. Philadelphia: Univ. of Pennsylvania Press.
- Labrenz, Annika. 2023. Functional variation of German *also* across registers and speaker groups. *Contrastive Pragmatics* 4(2). 289–320. DOI: 10.1163/26660393-bja10077.
- Labrenz, Annika, Kateryna Iefremenko, Kalliopi Katsika, Shanley E.M. Allen, Christoph Schroeder & Heike Wiese. 2025. Dynamics of discourse markers in language contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 493–526. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775187.
- Ladd, D. Robert. 2008. *Intonational phonology*. 2nd edn. (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press.

- Lee, Miseon, Hye-Young Kwak, Sunyoung Lee & William O’Grady. 2011. Processing, pragmatics, and scope in Korean and English. In *Proceedings of 19th Japanese-Korean linguistic conference*. Stanford: Center for the Study of Language & Information.
- Lee-Ellis, Sunyoung. 2012. *Looking into bilingualism through the heritage speaker’s mind*. University of Maryland (College Park). (Doctoral dissertation).
- Lein, Tatjana, Tanja Kupisch & Joost van de Weijer. 2016. Voice Onset Time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism* 20(6). 732–749. DOI: 10.1177/1367006915589424.
- Lloyd-Smith, Anika, Marieke Einfeldt & Tanja Kupisch. 2020. Italian-German bilinguals: The effects of heritage language use on accent in early-acquired languages. *International Journal of Bilingualism* 24(2). 289–304. DOI: 10.1177/1367006919826867.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Anika Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehayé, Heike Wiese, Sabine Zerbian, Yulia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Luke, Kang-kwong & Theodossia-Soula Pavlidou (eds.). 2002. *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (Pragmatics & Beyond New Series 101). Amsterdam: John Benjamins. DOI: 10.1075/pbns.101.
- MacWhinney, Brian. 2005. A unified model of language acquisition. In *Handbook of bilingualism*, 50–70. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195151770.003.0004.
- Marinis, Theodoros & Vasiliki Chondrogianni. 2010. Production of tense marking in successive bilingual children: When do they converge with their monolingual peers? *International Journal of Speech-Language Pathology* 12(1). 19–28. DOI: 10.3109/17549500903434125.
- McCardle, Peggy. 2015. Bilingualism: Research and policy. *New Directions for Child and Adolescent Development* (147). 41–48. DOI: 10.1002/cad.20088.
- Montrul, Silvina. 2008. *Incomplete Acquisition in Bilingualism* (Studies in Bilingualism). Amsterdam: John Benjamins. DOI: 10.1075/sibil.39.

- Montrul, Silvina. 2010. How similar are adult second language learners and Spanish heritage speakers? Spanish clitics and word order. *Applied Psycholinguistics* 31(1). 167–207. DOI: 10.1017/S014271640999021X.
- Montrul, Silvina. 2016. The language of heritage speakers. In *The acquisition of heritage languages*, 41–89. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139030502.003.
- Montrul, Silvina & Tania Ionin. 2010. Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism: Language and Cognition* 13(4). 449–473. DOI: 10.1017/S1366728910000040.
- Montrul, Silvina & Maria Polinsky (eds.). 2021. *The Cambridge handbook of heritage languages and linguistics*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108766340.
- Özsoy, Onur, Kateryna Iefremenko & Christoph Schroeder. 2022. Shifting and expanding clause combining strategies in heritage Turkish varieties. *Languages* 7(3). 242. DOI: 10.3390/languages7030242.
- Özsoy, Onur, Vasiliki Rizou, Maria Martynova, Natalia Gagarina, Luka Szucsich & Artemis Alexiadou. 2025. Null subjects in heritage Greek, Russian and Turkish. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 179–217. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775167.
- Paradis, Johanne. 2019. English second language acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors. In Megan M. Brown & Brady Dailey (eds.), *Proceedings of the 43rd Boston University Conference on Language Development*, 11–26. Somerville, MA: Cascadilla Press.
- Paradis, Johanne & Ruiting Jia. 2017. Bilingual children's long-term outcomes in English as a second language: Language environment factors shape individual differences in catching up with monolinguals. *Developmental science* 20(1). DOI: 10.1111/DESC.12433.
- Pascual y Cabo, Diego & Jason Rothman. 2012. The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics* 33(4). 450–455. DOI: 10.1093/applin/ams037.
- Pashkova, Tatiana & Shanley E. M. Allen. 2022. Specificity and definiteness in article choice in heritage speakers' majority English. *Talk presented at Heritage Languages and Variation, September 23-25 in Limassol, Cyprus*.
- Pashkova, Tatiana & Shanley E. M. Allen. 2025. Explicitness of referring expressions in heritage speakers' majority English. *Lingua* 314. DOI: 10.1016/j.lingua.2024.103854.

- Pashkova, Tatiana & Shanley E. M. Allen. In preparation. *Syntactic structures in the introduction of new subjects in h-speakers' majority English in the US.*
- Pashkova, Tatiana, Hannah Lee, Mark Murphy & Shanley E. M. Allen. Submitted. *Left dislocations in heritage speakers' majority English in the US.*
- Pashkova, Tatiana, Mark Murphy & Shanley E. M. Allen. 2020. *Explicitness in referent introduction in heritage speakers' majority English.* Poster presented at Boston University Conference on Language Development (BUCLD 45). Boston, USA.
- Pashkova, Tatiana, Wintai Tsehayé, Shanley E. M. Allen & Rosemarie Tracy. 2022. Syntactic optionality in heritage language use: Clause type preferences of German heritage speakers in a majority English context. *Heritage Language Journal* 19(1). 1–41. DOI: 10.1163/15507076-12340022.
- Pavlenko, Aneta & Scott Jarvis. 2002. Bidirectional transfer. *Applied Linguistics* 23(2). 190–214. DOI: 10.1093/applin/23.2.190.
- Pavlidou, Theodossia-Soula. 2014. Phases in discourse. In Klaus P. Schneider & Anne Barron (eds.), *Pragmatics of discourse*, 353–384. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110214406-014.
- Perniss, Pamela & Aslı Özyürek. 2015. Visible cohesion: A comparison of reference tracking in sign, speech, and co-speech gesture. *Topics in Cognitive Science* 7(1). 36–60. DOI: 10.1111/tops.12122.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.
- Queen, Robin. 2012. Turkish-German bilinguals and their intonation: Triangulating evidence about contact-induced language change. *Language* 88(4). 791–816.
- Rothman, Jason & Jeanine Treffers-Daller. 2014. A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics* 35(1). 93–98. DOI: 10.1093/applin/amt049.
- Sanchez Abchi, Veronica Soledad & Mariela Vanesa de Mier. 2017. Syntactic complexity in narratives written by Spanish heritage speakers. *Vial: Vigo International Journal of Applied Linguistics* 14. 125–148.
- Schegloff, Emanuel A. 1968. Sequencing in conversational openings. *American Anthropologist* 70(6). 1075–1095. DOI: 10.1525/aa.1968.70.6.02a00030.
- Schegloff, Emanuel A. 2007. *Sequence organization in interaction: A primer in conversation analysis*. 1st edn. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511791208.
- Schegloff, Emanuel A. & Harvey Sacks. 1973. Opening up closings. *Semiotica* 8(4). DOI: 10.1515/semi.1973.8.4.289.

- Schiffrin, Deborah. 1987. *Discourse markers* (Studies in Interactional Sociolinguistics 5). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511611841.
- Schleppegrell, Mary J. & M. Cecelia Colombi. 1997. Text organization by bilingual writers: Clause structure as a reflection of discourse structure. *Written Communication* 14(4). 481–503. DOI: 10.1177/0741088397014004003.
- Schmidt, Thomas & Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *Handbook on corpus phonology*, 402–419. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199571932.013.030.
- Schoonbaert, Sofie, Robert J. Hartsuiker & Martin J. Pickering. 2007. The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language* 56(2). 153–171. DOI: 10.1016/j.jml.2006.10.002.
- Scontras, Gregory, Maria Polinsky, Cheng-Yu Edwin Tsai & Kenneth Mai. 2017. Cross-linguistic scope ambiguity: When two systems meet. *Glossa: A Journal of General Linguistics* 2(1). 36. DOI: 10.5334/gjgl.198.
- Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–215. Mahwah: Lawrence Erlbaum Associates, Inc.
- Serratrice, Ludovica. 2007. Referential cohesion in the narratives of bilingual English-Italian children and monolingual peers. *Journal of Pragmatics* 39(6). 1058–1087. DOI: 10.1016/J.PRAGMA.2006.10.001.
- Serratrice, Ludovica, Antonella Sorace & Sandra Paoli. 2004. Crosslinguistic influence at the syntax-pragmatics interface: Subjects and objects in English-Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition* 7(3). 183–205. DOI: 10.1017/S1366728904001610.
- Shadrova, Anna, Martin Klotz, Rahel Gajaneh Hartz & Anke Lüdeling. 2025. Mapping the mappings and then containing them all: Quality assurance, interface modeling, and epistemology in complex corpus projects. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 69–110. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775161.
- Silva-Corvalán, Carmen. 2014. *Bilingual language acquisition: Spanish and English in the first six years* (Cambridge Approaches to Language Contact). Cambridge University Press.
- Şimşek, Yazgül. 2012. *Sequenzielle und prosodische Aspekte der Sprecher-Hörer-Interaktion im Türkendeutschen*. Münster: Waxmann. 328.

- Sorace, Antonella. 2011. Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism* 1(1). 1–33. DOI: 10.1075/lab.1.1.01sor.
- Sorace, Antonella & Francesca Filiaci. 2006. Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22(3). 339–368. DOI: 10.1191/0267658306sr271oa.
- Sorace, Antonella, Ludovica Serratrice, Francesca Filiaci & Michela Baldo. 2009. Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua* 119(3). 460–477. DOI: 10.1016/j.lingua.2008.09.008.
- Tagliamonte, Sali A. & Bridget L. Jankowski. 2019. Grammatical convergence or microvariation? Subject doubling in English in a French dominant town. *Proceedings of the Linguistic Society of America* 4(1). 1–15. DOI: 10.3765/plsa.v4i1.4514.
- Tolchinsky, Liliana, Victoria Johansson & Anita Zamora. 2002. Text openings and closings in writing and speech: Autonomy and differentiation. *Written Language & Literacy* 5(2). 219–253. DOI: 10.1075/wll.5.2.05tol.
- Tracy, Rosemarie & Ira Gawlitzek. 2023. *Mehrsprachigkeit und Spracherwerb*. Tübingen: Narr.
- Trotzke, Andreas, Ermenegildo Bidese & Manuela Caterina Moroni. 2020. German discourse particles in the second language classroom: Teasing apart learning problems at the syntax-pragmatics interface. *Pedagogical Linguistics* 1(2). 184–210. DOI: 10.1075/pl.20008.tro.
- Tsehaye, Wintai. 2023. Light-weights placed right: Post-field constituents in heritage German. *Frontiers in Psychology* 14. DOI: 10.3389/fpsyg.2023.1122129.
- Tsehaye, Wintai, Tatiana Pashkova, Rosemarie Tracy & Shanley E. M. Allen. 2021. Deconstructing the native speaker: Further evidence from heritage speakers for why this horse should be dead! *Frontiers in Political Science* 12. 1–14. DOI: 10.3389/fpsyg.2021.717352.
- Tsehaye, Wintai, Rosemarie Tracy & Johanna Tausch. 2025. Inter- and intra-individual variation: How it materializes in Heritage German and why it matters. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 141–177. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775165.
- Tsimpli, Ianthi-Maria. 2014. Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313. DOI: 10.1075/lab.4.3.01tsi.
- van Rijswyk, Remy, Antje Muntendam & Ton Dijkstra. 2017. Focus marking in Dutch by heritage speakers of Turkish and Dutch L1 speakers. *Journal of Phonetics* 61. 48–70. DOI: 10.1016/j.wocn.2017.01.003.

- Ventureyra, Valérie A.G, Christophe Pallier & Hi-Yon Yoo. 2004. The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics* 17(1). Attrition, 79–91. DOI: 10.1016/S0911-6044(03)00053-8.
- Vogels, Jorrig, Emiel Krahmer & Alfons Maes. 2019. Accessibility and reference production. In Jeanette Gundel & Barbara Abbott (eds.), *The Oxford handbook of reference*, 336–364. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199687305.013.16.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Anke Lüdeling, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Tracy Rosemarie, Christoph Schroeder, Luka Szucsich, Wintai Tsehay, Sabine Zerbian & Yulia Zuban. 2021. *RUEG Corpus*. Version 0.4.0. Zenodo. DOI: 10.5281/zenodo.3236068.
- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 33–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775159.
- Winkle, Claudia. 2015. *Non-canonical structures, they use them differently: Information packaging in spoken varieties of English*. Freiburg: Albert-Ludwigs-Universität Freiburg. (Dissertation). <https://freidok.uni-freiburg.de/data/10600>.
- Yoshioka, Keiko. 2008. Gesture and information structure in first and second language. *Gesture* 8(2). 236–255. DOI: 10.1075/gest.8.2.07yos.
- Zenker, Fred & Kristopher Kyle. 2021. Investigating minimum text lengths for lexical diversity indices. *Assessing Writing* 47. DOI: 10.1016/j.asw.2020.100505.
- Zerbian, Sabine & Marlene Böttcher. 2019. Stressed pronouns in mono-and bilingual German. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren

- (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 2640–2644. Australasian Speech Science & Technology Association Inc.
- Zerbian, Sabine, Marlene Böttcher & Yulia Zuban. 2022. Prosody of contrastive adjectives in mono-and bilingual speakers of English and Russian: A corpus study. In *Proceedings 11th international conference on speech prosody. International speech communication association (ISCA) online archive*, 812–816. DOI: 10.21437/SpeechProsody.2022-165.
- Zerbian, Sabine, Yulia Zuban, Marlene Böttcher & Oliver Bunk. 2025. Intonation in heritage languages. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 413–448. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775183.



# Chapter 14

## Dynamics of discourse markers in language contact

Annika Labrenz<sup>a</sup>, Kateryna Iefremenko<sup>b,c</sup>, Kalliopi Katsika<sup>d</sup>,  
Shanley E.M. Allen<sup>d</sup>, Christoph Schroeder<sup>e</sup> & Heike Wiese<sup>a</sup>

<sup>a</sup>Humboldt-Universität zu Berlin <sup>b</sup>University of Potsdam <sup>c</sup>Leibniz-Centre  
General Linguistics <sup>d</sup>University of Kaiserslautern-Landau  
<sup>e</sup>Humboldt-Universität zu Berlin, University of Potsdam

This chapter addresses the dynamics of discourse markers (DMs) in language contact. DMs are functionally variable elements that are grammatically not integrated and hence easily transferable – in terms of matter as well as in terms of pattern. This makes them especially dynamic in language contact and, in particular, in heritage languages. We present findings from the RUEG corpus on verbal DMs and on the three-dot sign, which we approach as a graphic DM, in different heritage languages (German, Greek, Turkish, Russian) and majority languages (English, German). Qualitative findings indicate an impact of societal status differences on frequencies, functional extension/restriction, and (inverted) pragmatisation processes.

### 1 Introduction

Discourse markers (DMs) are linguistic devices that speakers use to signal a range of pragmatic and communicative functions in their speech and writing. These markers are functionally variable elements, and their polyfunctional nature (e.g. Moosegaard Hansen 1998, Schiffrin 1987, Schourup 1999) makes them a highly versatile lexical domain in general (Auer & Günthner 2005). This inherent characteristic of DMs may be conducive to innovation, especially in informal communicative situations where there is less normative pressure compared



to more formal situations. In language contact settings, the use of DMs can be particularly interesting, as speakers' repertoires contain DMs from different languages whose functions might overlap. Research suggests that bilingual speakers profit from shared mental planning processes while using different resources in their repertoire (cf. Matras 2010: 69). As a consequence, bilingual speakers might have access to a shared functional pool of such DMs. Previous studies discuss two main reasons for the dynamics of DMs in language contact: 1) their detachability and 2) their tendency for functional convergence.

- (1) Detachability: According to Matras, DMs are highly detachable not only in a structural sense but also from the bulk of the lexicon (Matras 1998, 2020; see also Fuller 2001). Because of their "pragmatic role [...] as highly automatic conversational routines" (Matras 2020: 209) and their gesture-like quality, Matras suggests the involvement of distinct control mechanisms compared to those governing lexical items or grammatical inflections (Matras 2020: 209). In bilinguals, this easily leads to the tendency to integrate DMs from one language into the other, as various studies on borrowings have shown (e.g. Salmons n.d., Goss & Salmons 2000, Fuller 2001, Matras 1998, see also Wiese et al. 2025 [this volume]). This phenomenon is also called matter replication (see Matras & Sakel 2007, Matras 2010).
- (2) Functional convergence: DMs are not only susceptible to matter replication (or borrowing) but also to pattern replication (see Matras & Sakel 2007, Heine et al. 2021: 212). For the discourse domain and specifically DMs, this means that the functional spectra of specific DMs have the potential to converge. This can lead to functional extensions, restrictions and/or an increase in frequency.

While previous studies on DMs in language contact settings mainly focused on borrowings (e.g. Salmons n.d.; contributions in Maschler 2000), this chapter focuses on functional convergence in the discourse domain. We examine the dynamics and variability of DMs across different languages and contact settings where languages have a different societal status: They have either the status of a minoritised heritage language that is mainly spoken in families and in certain communities but is not the language of the larger society or as a majority language, i.e. the language spoken by the larger society in which administrative processes and school education usually take place. A typical scenario for bilingual heritage speakers (HSs) usually results in them being dominant in their majority language, at least in certain communicative situations, at some point after

the start of school (cf. Rothman 2009, Flores et al. 2019). As a consequence, one might expect a stronger influence of the majority on the heritage language. In this chapter, we will look at how this influence manifests itself in the case of DMs. More specifically, we will investigate how overlapping functional spectra of DMs in one language influence the use of DMs in a contact language. This leads us to our first research question: What are the patterns of functional variation, extension, and restriction of DMs in language contact?

We also introduce a new kind of DM, namely graphic DMs, such as the three-dot sign, from instant messaging (Labrenz et al. 2022; see also Wiese & Labrenz 2021 on emoji as graphic DMs), thus including the (informal) written domain into DM research. By so doing, we follow Imo's call to broaden DM studies (Imo 2017). Graphic DMs are an interesting case for studies in language contact because they are not confined to any specific language, but rather keep their form or matter across languages, which makes them translinguistic elements. In view of this fact, it is particularly interesting to see if language-specific uses or functions still occur or if the same matter also leads to cross-linguistic universal patterns. This leads us to our second research question: Are there language-specific uses of graphical DMs even if they have the same form across languages? If so, what are the dynamics in bilinguals? In pursuit of these questions, we conduct a cross-linguistic study that includes a range of contact settings which allows us to disentangle the effects of societal language status and bilingualism in general. The chapter is structured as follows: Section 2 discusses conceptual aspects of DMs, followed by an overview of the functions of the verbal and graphic DMs explored in this chapter based on prior literature. Section 3 presents the database and outlines the procedure of the corpus study. In Section 4, we present the findings regarding specific uses in language contact. Following a brief discussion and the synthesis of the findings in Section 5, we conclude in Section 6.

## 2 Verbal and graphic DMs

DMs typically come from a variety of part-of-speech categories, such as adverbs and conjunctions (cf. Schourup 1999, Crible 2018: 69 for an extended enumeration). According to the literature, such polyfunctional lexical items carry a core meaning that has been partly lost through a process of semantic bleaching and syntactic disintegration (cf. Hopper & Traugott 2003: 94), a process also referred to as pragmatisation or grammaticalisation (see Brinton 2017 and Diewald 2011 for a discussion of these terms). A consequence of this process is that DMs contribute solely to the discourse-functional level. And this can be said for verbal

as well as graphic DMs alike (cf. Wiese & Labrenz 2021 on emoji, Labrenz et al. 2022 on the three-dot sign). We define DMs

[...] as elements that are not fully syntactically integrated and do not directly contribute to the propositional meaning and truth value of an utterance but rather operate on the level of discourse (cf. Blakemore 2004, Fraser 2006, Blühdorn et al. 2017). (Wiese & Labrenz 2021: 3)

Throughout the literature there are different approaches to what should be included in the category of DMs. This terminological inconsistency in the field (e.g. Andersen 2001, Fraser 1999) makes it particularly challenging to work cross-linguistically on discourse-pragmatic markers. In this paper, we address this challenge by taking on a functional perspective on DMs (cf. Maschler 2009). For this purpose, we consider a range of functions that are prototypical for the class of DMs (see also Pons Bordería & Fischer 2021 on DM features).

For instance, connectors that mark semantic relations are more lexicalised and, in our view, less prototypical as DMs than markers that solely structure discourse. This is in line with Ariel (1994), who distinguishes between “semantically transparent markers as opposed to relatively opaque ones” (Ariel 1994: 3251). Pragmatic detachability (Matras 1998) in terms of degree of lexicalisation and semantical transparency can thus serve as indicators for prototypicality of DMs. Following this logic, English *so* as a semantically transparent marker, is less typical for the class of DMs than English *well* (Ariel 1994, Müller 2005: 62).

Below we give an overview of the canonical functional spectrum of the items under investigation – that are German *so*, Turkish *yani* and *iste*, Greek *etsi* and *lipon*, and the three-dot sign based on previous descriptions in the literature, and list those functions that are relevant for later analysis. Since most of the verbal DMs overlap functionally with either English *so* or with German *so* or *also*, we expect convergent developments between these markers. As they are important counterparts of the items under investigation, we first give a brief overview of their functional spectrum.

English *so* is widely being acknowledged as an adverb of degree and manner (like German *so*) and as a conjunction (e.g. Müller 2005: 62). As a conjunction, it indicates an inference or a consequence. In this function, it is seen as a DM by most researchers, although it might not be the most prototypical representative of its class (see Müller 2005: 62 and discussion in the Introduction). Another function of *so* in peripheral position is that of initiating a new narrative sequence (cf. Bolden 2009). The functional spectrum of English *so* encompasses initialising

a narrative, indicating the end of a narrative, thematic switch, and [indicating a consequence]<sup>1</sup>.

German *also* is a highly polyfunctional lexical item (e.g. Alm 2007, Fernández-Villanueva 2007, Labrenz 2023). A central use of *also* is as a consequence indicating adverbial connector and as marker for indicating inferences (Konerding 2004). It is also commonly used as a repair marker, including the indication of elaborations/specifications, corrections, restarts and hesitations (see Pfeiffer 2017). Similar to English *so*, it can be used to initiate a new narrative sequence. In addition, it can indicate the speaker's stance (evaluation) (Labrenz 2023). The functional spectrum of *also* includes initialising a narrative; indicating evaluations/inferences, elaborations, or corrections/restarts; and [indicating a consequence (adverbial connector)]<sup>2</sup>.

German *so* is a highly polyfunctional lexical item and its classification has been widely discussed in a range of grammatical descriptions and studies (see Schumann 2021 for an overview). A central function of *so* is its deictic use (e.g. Thurmail 2001, Ehlich 2007). As a modal indexical marker, it can indicate manner, quality, or intensity (Wiese 2011, Schumann 2021). Additional uses are as a comparison particle, as a quotative marker, as a progressive marker, and to indicate approximations and hedging (cf. Hennig 2006, Wiese 2011, 2012, Schumann 2021). In addition, Hennig (2006) mentions the use of *so* in the function of initialising a new narrative sequence in left peripheral position (cf. Hennig 2006). A non-canonical use of *so* as a focus marker has been described by Wiese (2011) and explored in detail by Schumann (2021). In this use, *so* is an optional element that is semantically bleached such that it does not contribute any meaning and solely contributes on the level of discourse (Wiese 2011). As a focus marker and in hedges (in *und so* 'and so' or *oder so* 'or so'), *so* can be located in utterance final position (Wiese 2012, Schumann 2021: 48). Hence, in these uses it fits our criteria for DMs. In this paper, the focus function of *so* in the right peripheral position will be treated as an assertion because, in contrast to medial positions, the scope is extended from a phrase to the entire preceding clause. German *so* functionally overlaps with English *so*, but the only overlap in discourse functions is the initialising function. The discourse-functional spectrum of German *so* includes initialising a narrative, assertion, and hedging. Examples are provided in Section 4.1.

---

<sup>1</sup>Functions in square brackets indicate that they are not the most prototypical cases of the DM category.

<sup>2</sup>German *also* is not a DM in this case: When it is used for indicating a consequence on the propositional level it is syntactically integrated.

Turkish *iste* is one of the most frequent discourse markers in Turkish (Yılmaz 2004). Its main functions are marking shared knowledge between discourse participants, claiming the turn in a conversation, marking topic boundary by indicating the end of a discourse unit and emphasizing the speaker's points (Özbek 2000, Yılmaz 2004). In the latter function which we call *assertion*, it overlaps with German *so*. In addition, *iste* has functional overlap with English *so* and *well*. The discourse-functional spectrum of *iste* encompasses indicating a thematic switch, ending a narrative, and assertion. Examples are provided in Section 4.2.

Turkish *yani* is the most frequent discourse marker in conversational Turkish (Yılmaz 1994, 2004, Özbek 1995). Turkish *yani* is originally a loanword from Arabic (*ya'nī*, “[he/she/it] means”) and entered into Turkish directly as a DM with predominant linking functions, such as indicating elaborations. However, in modern Turkish *yani* is multifunctional, and in addition to the function of elaboration, it is used to summarize ideas, to emphasize the speaker's stance on something (indicate an evaluation), to initiate a turn, and to hold the floor (Yılmaz 2004). It functionally overlaps with German *also*. Depending on its functional contribution in discourse, it overlaps functionally with English *well* or *so*. The discourse-functional spectrum of *yani* includes indicating elaboration, evaluation, the end of a narrative, and hesitation. Examples are provided in Section 4.3.

Greek *lipon* is used most frequently as a DM according to Georgakopoulou & Goutsos (1998). In formal grammars and Greek-English dictionaries, *lipon* is characterised as a deductive conjunction (Triantafyllidis 2019) which can be translated as “so”, “then”, “therefore”, “hence” or as an interjection meaning “well”, “so”, “now” expressing surprise, relief, query, or decision (Stavropoulos 1988: 119). In spoken narratives, *lipon* functions as a hypotactic marker (Redeker 1990) linking secondary parts of speech (such as comments, corrections, etc.) to the main narrative part (Antoniou 2008). On the basis of their analysis of spoken and written narratives and non-narratives, Georgakopoulou & Goutsos (1998) claim that *lipon* is used for the transition from a secondary to the main line of narration but never the other way around. Such uses of *lipon* can also be found in classroom environments where *lipon* is used by teachers in turn-initial position to return to previously interrupted narration or discussion (Christodoulidou 2014). There is thus functional overlap with German *also* and with English *so* or *well*. The discourse-functional spectrum of *lipon* includes initialising a narrative, and indicating a thematic switch. Examples are provided in Section 4.4.

Greek *etsi*'s main function in Modern Greek is that of an adverb of manner or quantity (Tzartzanos 2002). Regarding its function as a DM, Georgakopoulou & Goutsos (1998) did not find evidence that *etsi* functions as a DM in spoken modality and only found limited use of *etsi* as a DM in written modality. After

conducting a corpus analysis of spoken Greek texts, however, Tsampoukas (2015) found a variety of DM functions of *etsi*. More specifically, Tsampoukas (2015) provides evidence and argumentation that *etsi* is currently in the beginning of a grammaticalisation process. The main DM function that Tsampoukas (2015) identified for *etsi* is that it indicates a result or a consequence of a previously mentioned event or series of events in the discourse. Thus, similarly to German *also* and English *so*, *etsi* may have the function of a consecutive connector – a function which we do not see as the most prototypical for the category of DMs, but which could be a step with potential for further pragmatalisation. The discourse-functional spectrum of *etsi* includes [indicating a consequence (emerging)]. Examples are provided in Section 4.5.

The three-dot sign's cross-linguistic canonical function is its use as a placeholder for omitted material (see Raclaw 2006: 301, Baron & Ling 2011: 60 on English; Bredel 2008: 125, Meibauer 2019: 68, Androutsopoulos 2020: 154 on Greek, Rosenthal 2012: 90 on Russian; Turan 2014: 44 ff. on Turkish). A number of additional functions come into play, especially when writing digitally. These functions closely resemble those of verbal DMs. Consequently, we approached the three-dot sign as a graphical DM. Like verbal DMs the three-dot sign is positioned at the periphery and does not contribute directly to the propositional meaning, but rather at the level of discourse. This perspective allowed us to subsume its functions under textual, subjective and intersubjective discourse functions. Examples are given in Section 3.2.

### 3 Method

#### 3.1 Database

The informal and formal-spoken productions in the English, German, Greek, and Turkish subcorpora snapshot versions 1.0<sup>3</sup> of the RUEG corpus constitute the database for the study of verbal DMs. The informal-written productions (instant messages) of the RUEG corpus version 0.4.0 (Wiese et al. 2021) are the database for the study of the three-dot sign, and also include Russian.

The data comprise naturalistic and comparable productions that were elicited using the Language Situations setup (Wiese 2020). As described in more detail in Wiese et al. (2025 [this volume]), in this setup, participants watched a video about a car accident and were asked to imagine themselves as eyewitnesses, and to tell different interlocutors about it in four different communicative situations. We

---

<sup>3</sup><https://hu.berlin/rueg-corpus> (last accessed January 10th, 2024)

focus on the spoken mode in formal and informal communicative situations for the verbal DMs (Table 1). In these conditions, participants were asked to act out leaving a voicemail for the police and a voicemail via WhatsApp for a friend. For the three-dot sign, we focus on the informal-written communicative situation in which participants were asked to write an instant message to a friend (Table 2). Tables 1 and 2 give an overview of the word tokens and speakers included in the respective analyses. We included heritage German (Table 1), Turkish, Greek (Tables 1 and 2) and Russian (Table 2) in contact with majority German and English and additionally looked at the respective languages spoken in a majority context (e.g. German in Germany, Turkish in Turkey). Each subcorpus contains bilingually and monolingually raised adolescents (age 14–18 years) and adults (22–35 years). The monolingually raised speakers grew up with English, German, Greek, Russian, or Turkish as their only language in the family. These languages have the societal status of a majority language in the respective countries of elicitation. The bilinguals were HSs of Greek, Russian, and Turkish, raised in Germany or the US and HSs of German in the US. Thus, they grew up with either German or English as a majority language, in addition to their respective heritage languages.

Table 1: Database for the analysis of verbal DMs, RUEG corpus Snapshot 1.0: Number of overall word tokens in informal (is) and formal-spoken (fs) data and speakers across languages and language status (excluding heritage Russian speakers of the English and German subcorpus)

| Language | Societal status | Tokens (is, fs) | Speakers |
|----------|-----------------|-----------------|----------|
| English  | majority        | 63,813          | 222      |
| German   | majority        | 59,300          | 175      |
|          | heritage        | 9,187           | 35       |
| Greek    | majority        | 15,560          | 64       |
|          | heritage        | 23,505          | 108      |
| Turkish  | majority        | 11,538          | 66       |
|          | heritage        | 25,370          | 126      |

### 3.2 Procedure

For the verbal DMs, we selected specific polyfunctional lexical items (Pichler 2016: 1) from each language that serve discourse-pragmatic functions that partly

Table 2: Database for the three-dot sign, RUEG corpus version 0.4.0:  
Word tokens of informal-written (iw) data and speaker numbers across  
languages, societal language status, and countries

| Language | Societal status | Tokens (iw) | Speakers |
|----------|-----------------|-------------|----------|
| English  | majority        | 16,363      | 187      |
| German   | majority        | 25,094      | 165      |
| Greek    | majority        | 5,488       | 64       |
|          | heritage-DE     | 3,819       | 47       |
|          | heritage-US     | 3,391       | 64       |
| Russian  | majority        | 4,398       | 67       |
|          | heritage-DE     | 4,927       | 58       |
|          | heritage-US     | 3,889       | 66       |
| Turkish  | majority        | 4,222       | 64       |
|          | heritage-DE     | 4,502       | 65       |
|          | heritage-US     | 3,575       | 56       |

overlap with their counterpart in the respective contact language. In particular, we look at German *so*, Turkish *yani* and *iste*, and Greek *etsi* and *lipon*. We chose these markers because they are among the most frequent DMs in the respective languages, and our corpus data indicated interesting patterns in heritage language use. We extracted all occurrences of these items along with their verbal context. Two annotators independently annotated positions and functions. The annotated positions include left (1) and right (2) peripheral and medial (3) positions. To guide our functional annotation process, we referred to an established annotation catalogue derived from functions of these DMs as described in the existing literature (see Section 2). We regularly discussed cases of doubt. Special emphasis was placed on instances that did not align with the established functions of the respective DMs. In such cases, we determined functions by analysing the preceding and subsequent verbal content and the role of the discourse marker in that specific context.

- (1) left periphery  
**so i:** just witnessed an accident [USmo05FE\_isE]<sup>4</sup>

<sup>4</sup>The speaker codes are constructed as follows: The first two letters refer to the country of elicitation (DE=Germany, TU=Turkey, GR=Greece, US=USA RU=Russia, TU=Turkey); the next two

(2) right periphery

weil äh die müssen Rücksicht auf uns nehmen so  
because uh they have to respect of us give DM  
'because they have to respect us so'  
[DEbi26FT\_isD]

(3) medial

ama iki tane araba *işte* kaza yap-tı  
but two piece car DM accident do-3SG.PST  
'but (-) two cars *işte* had an accident'  
[DEbi26FT\_isD]

For the three-dot sign the procedure was similar, but additional factors were taken into account: We extracted all occurrences of the three-dot sign together with their preceding and following context and annotated the data manually. In our analysis, we considered the position (lone, message-final, CU-internal, between discourse units, see (4–10)) and the discourse context on the macro- and micro-levels. On the level of macro-structure, the entire message was taken into account and divided into opening, main narration and closing ((4), see also Katsika et al. 2025 [this volume]). At the level of micro-structure, preceding and following discourse units of the three-dot sign were classified according to their functional contribution as real-world-referring (mainly referring to the retelling of the accident as in 'witnessed a rear end collision' in (4)), subjective (convey attitudes, stance, evaluations as in 'weird day today' in (9)), and intersubjective (concerned with speaker-hearer relationship for example through greeting in (4) (Labrenz et al. 2022: 255)). Subsequently, we derived textual, subjective, and intersubjective discourse functions of the three-dot sign considering its position and contextual environment.

(4) macro-structure

[hey ...sorry im running a bit late ...witnessed a rear end  
collision.]OPENING  
[someone last their ball and it went onto oncoming traffic!]  
MAIN NARRATION  
[need to stay and give testimony to the officers about what i saw! see you  
soon]CLOSING  
[USbi19FG\_iwE]

---

to the speaker's language background (bi=bilingual, mo=monolingual; followed by age group (1–49=adults, >50=adolescents); then gender self-identification (F=female, M=male), heritage language (G=Greek, D=German, T=Turkish, R=Russian) or majority language in case of German (D) and English (E) monolinguals; communicative situation (iw = informal-written, is = informal-spoken, fw=formal-written, fs = formal-spoken); language of elicitation (D=German, E=English, G=Greek, R=Russian, T=Turkish).

Based on previous literature (e.g. Androutsopoulos 2020, Busch 2021, Meibauer 2019) and in view of position and context, we identified the following DM functions of this graphic device in our data and subsumed them under three broader categories as follows.

1. textual: discourse organisation (5) and segmentation (6), creating dramatic effect (7)
2. subjective: indicating speechlessness and/or emphasis (8)
3. intersubjective: general openness of the communication inviting the interlocutor to react (9), (10)

In (5) and (6), the three-dot sign has a textual function. While in (2) it indicates a major shift at the macro- and micro-structure levels, in (6) it is simply segmenting the informational content in the main narration (see Labrenz et al. 2022: 252–58 for details of the method).

- (5) Between discourse units and between main narration and closing  
[они плсвонили полицие]MAIN NARRATION ... [как ты думаешь: мне  
тоже надо в полицию поехать и им как свидетел всё  
рассказать?]CLOSING  
oni plsvonili polizie ... kak ty dumaesh: mne tozhe nado v poliziyu  
poekhat' I im kak svidetel vsjo rasskazat'?  
‘they called the police ... Do you think I should go to the police and tell  
them everything as a witness?’ [DEbi03FR\_iwR]
- (6) Between discourse units: preceding and following real-world referring  
DUs  
this 1 guy was playing with a ball ... it slipped from his hand  
[USbi57FG\_iwE]
- (7) Internal with preceding interjection  
Whoa...just witnessed a chain reaction accident. [USmo32ME\_iwE]
- (8) Between discourse units with preceding subjective DU  
bonitaaaa, tam auffahrunfall, aber sooo unnötig ... siehste man kann nicht  
kontrollieren habibi  
‘bonitaaaa, tam rear-end collision, but sooo unnecessary ... you see you  
cannot control habibi’ [DEbi03FT\_iwD]

- (9) Message-final with preceding subjective DU  
and crashed into each other, wierd day today ... [USbi55MR\_iwE]
- (10) Message-final with preceding real-world referring DU  
ich bleib mal hier, falls ich aussagen muss ...  
'I'm staying here in case I have to give a witness statement...' [DEbi37FR\_iwD]

The following section includes an analysis of each of the DMs that we identified and analysed in the RUEG corpus.

## 4 Findings

### 4.1 German *so*

Although German *so* and English *so* differ in their phonological representation, they are orthographically identical. As described in Section 2, the only functional overlap regarding DM uses is the initializing function. We start from the following hypothesis for German as a majority language<sup>5</sup> compared to German as a heritage language in a majority-English setting: Due to the orthographic similarity of German *so* and English *so* and their divergent functional spectrum, we expect an impact of majority English on heritage German, leading to functional differences in heritage German compared to majority German. A close look at the positions and specific functions of *so* reveals interesting trends. As shown in Table 3, most of the uses of *so* in all groups are in the integrated position. If we focus on peripheral uses, i.e. those cases that correspond to our DM definition, we see that while HSs prefer left peripheral positions, majority speakers (MSs) show a more frequent use in the right peripheral position. This positional difference between MSs and HSs is related to differences in functional use. Note that we included mono- and bilingual MSs in the analysis in order to disentangle bilingualism from language status effects. In the following, we first present specific functions and then show the proportional use of these functions per speaker group and how this is related to position.

In HSs of German, we observe the use of *so* with a novel function in left peripheral position, namely as a connector indicating a consequence (11a). This is one of the functions of English *so* (11b). In German it is an innovative use of *so*, since this function would canonically be covered by German *also*.

---

<sup>5</sup>MSs of German integrated in this analysis are either monolingually or bilingually raised with Turkish as their (other) family language.

Table 3: Raw numbers (*n*) and percentages of syntactically integrated, right and left peripheral positions of ‘so’ across mono- and bilingual MSs of German and HSs of German

|         | integrated |      | right peripheral |      | left peripheral |      |
|---------|------------|------|------------------|------|-----------------|------|
|         | <i>n</i>   | %    | <i>n</i>         | %    | <i>n</i>        | %    |
| MS (mo) | 306        | 88.0 | 41               | 11.8 | 1               | 0.3  |
| MS (bi) | 282        | 87.9 | 34               | 10.6 | 5               | 1.68 |
| HS      | 55         | 70.5 | 1                | 1.3  | 22              | 28.2 |

(11) consecutive connector

- a. aber die Autos haben gekommen so sie [die Leute]  
but the cars have.AUX.3PL PTCP-come-PTCP so they [the people]  
haben gestoppt  
have.AUX.3PL PTCP-stop-PTCP  
'but the cars came, so they [the people] stopped' [USbi57FD\_fsD]
- b. *it was cause this guy like lost control of his soccer ball and I know you love soccer so thought you'd be interested* [USmo1FE\_isE]

We also observe a more frequent use of the ‘initialising narrative’ function in HSs compared to MSs (cf. Hennig 2006 on that function). This may also reflect an influence from English *so*, for which the initialisation of a new narrative (sequence) has been described as a function (cf. Bolden 2009). We also find evidence for this function in our data for English *so* (12b).

(12) initialising message/narrative<sup>6</sup>

- a. **so** F16 da war ein Mann  
so F16 there was a man  
'so F16, there was a man' [USbi75MD\_fsD]
- b. so i: just witnessed an accident it seemed like a minor accident there was a woman on the right side of the street opening her trunk [USmo05FE\_isE]

Additionally, as shown in Table 3, there is only one case of right peripheral use of *so* in HSs, while for MSs this is the most frequent peripheral position. This is related to functions. The one use in the right periphery is as a general extender

<sup>6</sup>F16 in (12a) refers to the case number of the accident that was provided by the elicitors.

or hedge<sup>7</sup> as in (13). We found no evidence though for the use as an utterance final assertion marker as in (14) which we found to be very common in majority language use (see Table 4). This might indicate that the non-overlap with the functional spectrum of English *so* might have an additional impact on the functional use of German *so* in HSs. However, the fact that these functions are (almost) absent in our data does not necessarily mean that HSs do not actually use such functions.

- (13) general extender/hedge  
 ich soll auch nochmal anrufen und **so**  
 I should also again call and so  
 'I should also call again and *so*' [DEmo17MD\_isD]
- (14) (utterance final) assertive function  
 dieser Ball ist dann einfach auf die Straße gerollt **so**  
 this ball is then simply onto the street rolled DM  
 'this ball then simply rolled onto the street *so*' [DEmo76FD\_isD]

Table 4: Percentages of functions of 'so' at the periphery across speaker groups. Columns should sum to 100%, missing values are functions that occur less than 4 times

| Function     | Peripheral position | MS (mo) | MS (bi) | HS   |
|--------------|---------------------|---------|---------|------|
| hedge        | right               | 59.5    | 58.9    | -    |
| assertion    | right               | 38.1    | 28.2    | -    |
| consequence  | left                | -       | -       | 39.1 |
| initialising | left                | -       | -       | 30.4 |

In summary, our analysis of German HSs' use of German *so* revealed functional extensions (indicating a consequence), and restrictions (hedge, assertion) when no overlapping functions were present. Furthermore, when the functional spectrum overlapped (initialising function), German HSs displayed a higher frequency of use compared to German MSs. For this item, we observe a unidirectional cross-linguistic influence from the majority into the heritage language which supports our initial hypothesis.

<sup>7</sup>Note that English *so* can be used as general extender in the construction 'and so on (and so forth)' in parallel to German 'und so weiter und so fort'. In (colloquial) German this is often reduced to 'und so'.

## 4.2 Turkish *iste*

Turkish *iste* overlaps with German *so* in its utterance final assertive function. Thus, our hypothesis is the following: Due to functional similarities of German *so* and Turkish *iste*, we expect a more frequent use of overlapping functions in heritage Turkish in Germany. In our data, addressee-oriented functions such as marking shared knowledge between discourse participants and claiming turn in a conversation are very rare, presumably because the elicited productions are monologues that do not presuppose an answer or a reaction from the interlocutor. In Table 5, we present the percentage of functions of *iste* across speaker groups. Only functions that occurred more than 2 times are considered.

Table 5: Percentages of functions of *iste* across speaker groups; only functions that occurred more than 2 times are considered

| Function               | HS      |      | MS     |
|------------------------|---------|------|--------|
|                        | Germany | US   | Turkey |
| assertion              | 46.8    | 37.0 | 41.9   |
| thematic switch        | 13.9    | 21.9 | 12.9   |
| connector              | 11.4    | 13.7 | 9.7    |
| hesitation             | 12.7    | 4.1  | 8.1    |
| initializing narrative | 5.1     | 2.7  | 12.9   |
| elaboration            | 6.3     | 15.1 | 12.9   |
| other                  | 3.2     | 5.5  | 1.6    |

Among all three speaker groups in our data, *iste* most frequently functions as an utterance-final assertion. Notably, Turkish HSs in Germany exhibit a slightly higher use of *iste* in this function (15a) compared to Turkish MSs and HSs in the US. In fact, this is the function that is shared by Turkish *iste* (15a), (15b) and German *so* (14). Hence, the higher frequencies of *iste* in this function might be driven by the functional overlap with German *so*.

(15) assertive function

- a. arka-daki arabə yetiş-e-me-di ve (-) kaza  
back-ATTR car keep-up-NEG-PST.3SG and (-) accident  
ol-du *iste*  
happen-PST.3SG DM  
'The car at the back wasn't quick enough and the accident happened  
*iste*' [DEbi59MT\_isT]

- b. karşı-dan (-) bebek araba-li iki kişi geç-iyō-du (-)  
opposite-ABL stroller-ATTR two person pass-PROG-PST.3SG  
yan-ın-da bi de çocuğ-u var-dı işte  
side-POSS-LOC also child-POSS be-PST.3SG DM  
'From the opposite side two persons with a baby were passing. They  
also had a child *işte*.' [TUmo73MT\_fsT]

In addition to the proportionally more frequent use of the assertive function in HSs in Germany, we find novel functions of *işte* as a circum-connector (16) and what could be called an emergent subordinate connector (17) in heritage and majority Turkish; although the frequency of such examples is slightly higher in HSs. Drawing a parallel to the developments in Balkan Turkish, Keskin et al. (n.d.) discuss similar examples where DMs are used in the function of connectors and argue that such innovation points to an ongoing shift of Turkish to Indo-European subordination patterns (see Keskin 2025 [this volume]).

The development of *işte* from a marker with predominant discourse-pragmatic functions to novel connecting functions corresponds to an opposite development to typical pragmatisation processes characterised by semantic bleaching (cf. Section 2). Compared to DMs as defined in this article, connectors contribute more meaning to the logical coherence of utterances and are, therefore, at least less prototypical for the DM category than those that operate purely on a meta-communicative level indicating, for instance, a new narrative sequence.

- (16) circum-connector  
*işte* (bi) ilk araba da köpeğ-e es-me-sin diye bi fi/firen-e  
DM ART first car also dog-DAT hit-NEG-OPT.3SG CON ART break  
bas-tı  
push-PST.3SG  
'*işte* the first car not to crush the dog diye braked' [DEbi25MT\_isT]

Example (16) demonstrates the use of *işte* in combination with the subordinator *diye* in the function of a circum-connector which links the finite adverbial clause introduced with *işte* to the main clause. Similar use is also observed in the combination of the DM *yani* with *diye* (see Section 4.3).

- (17) relative-connector  
sol taraf-ta da bi tane kadın *işte* alışveriş falan yap-mış  
left side-LOC also ART piece woman DM groceries stuff do-PST.3SG  
'On the left, there was a woman *işte* did groceries and stuff'  
[DEbi09FT\_isT]

Example (17) can be interpreted as a main clause which is followed by a relative clause with the relative connector *iste*. Hence, the sentence follows a subordination pattern similar to the Indo-European one: The relative clause is postpositive finite and is linked to the main clause by means of a free subordinating element. Interestingly, the novel function of *iste* is found also in majority Turkish, which suggests that Turkish in Turkey may also be undergoing syntactic change. At the same time, a higher number of *iste* used in the function of a connector in heritage Turkish might be interpreted as a convergence phenomenon, where Turkish develops patterns of clause combining which follow the model of German and English, respectively (Özsoy et al. 2022, see also Schroeder et al. 2025 [this volume]).

HSs in the US show the lowest frequency in the use of *iste* as an assertion marker – a function that is typically used in informal situations and that is not in the functional spectrum of English *so*. Instead, this speaker group has higher frequencies of *iste* as a connector and as an indication of thematic switch – functions that are in the functional spectrum of English *so*. In addition, we also find a cross-linguistic influence of the majority on the heritage language, but only in terms of the frequency of use of a specific function. The extension of the functional spectrum in terms of specific connective functions in HSs may be more indicative of a language-internal development, as it is also found in Turkish in Turkey, but this process may be accelerated by language contact. Thus, there is only partial support for our original hypothesis.

### 4.3 Turkish *yani*

Turkish *yani* functionally overlaps with German *also* in its repair marking functions, including elaborations. It also overlaps with English *well* for marking hesitations and *so*, especially for indicating the end of a narrative.

Due to functional similarities of German *also* and English *well/so* with Turkish *yani*, we expect a more frequent use of overlapping functions in heritage Turkish. In Table 6, we present percentages for use of the different functions of *yani* across speaker groups. Only functions that occurred more than two times are considered.

HSs in the US show a more frequent use of *yani* as a hesitation marker (18) – a common function of English *well* (cf. Aijmer 2011) compared to the other two groups. HSs in Germany use *yani* predominantly for indicating elaborations (19) and corrections – two common functions of German *also*. The functional overlap might favor a more frequent use of such functions in HSs compared to Turkish MSs and to HSs with English or German as a majority language respectively.

Table 6: Percentages of use of the different functions of *yani* across speaker groups; only functions that occurred more than 2 times are considered

| Function         | HS      |      | MS     |
|------------------|---------|------|--------|
|                  | Germany | US   | Turkey |
| elaboration      | 26.5    | 18.7 | 19.2   |
| evaluation       | 8.8     | 4.4  | 32.7   |
| correction       | 23.8    | 10.9 | 5.8    |
| assertion        | 8.2     | 6.6  | 13.5   |
| connector        | 10.2    | 5.5  | 5.8    |
| hesitation       | 7.5     | 41.7 | 5.8    |
| encapsulation    | 4.1     | 3.3  | 5.8    |
| ending narration | 3.4     | 3.3  | 11.5   |
| thematic switch  | 4.8     | 5.5  | -      |

## (18) hesitations marker

çünkü *yani* o a hab/aa (-) *yani* (-) aile-ye doğru git-me-ye  
 because DM he hmm DM family-DAT towards go-NMZ-DAT  
 çalış-tı  
 try-PST.3SG

‘Because *yani* hmm he *yani* tried to go towards the family’

[USbi16MT\_fsT]

## (19) elaboration

ay/ ondan: ee dolayı işte karşı taraf-a yuvarlan-ıyo köppek de  
 because of that DM opposite side-DAT roll-PRS.3SG dog also  
 on-a sa/ *yani* top-a saldır-ıyo  
 it-DAT DM ball-DAT attack-PRS.3SG

‘That’s why it rolled to the opposite side. The dog attacked it *yani* the ball.’

[DEbi16FT\_isT]

Similar to what was observed with regard to *işte*, we find a novel function of *yani* in the function of what might be called an emergent subordinate connector. Like with *işte*, this may combine with the subordinator *diye*, as in (16), and again this is found in heritage as well as majority Turkish speakers, but the number of such occurrences is higher in HSs.

## (20) circum-connector

- köpeğ-i-ni      hemen      sıkı tut-tu      *yani* bi şey  
 dog-POSS-ACC immediately tight hold-PST.3SG DM something  
 ol-ma-sin      *diye*  
 happen-NEG-OPT.3SG CON  
 ‘S/he immediately held her/his dog tight so that nothing would happen  
 (to it).’

[USbi16MT\_isT]

## (21) relative connector

- şu/ orda bi kadın vardi      *yani* araba-sı-na (-) alışveriş  
 th/ there one woman be-PST.3SG DM car-POSS-DAT (-) grocery  
 torba-lar-in-1 falan koyuyodu  
 bag-PL-POSS-ACC stuff put-ROG-PST.3SG  
 ‘There was a woman *yani* (who) was putting grocery bags and stuff into  
 her car.’

[DEbi87FT\_isT]

In line with our hypothesis, the example of *yani* shows the influence of the respective majority language: HSs show a more frequent use of exactly those functions that overlap with the functional spectrum of the respective majority language counterpart. In addition, similar to *iste*, there may be language internal developments at play, which may be even more dynamic in HSs.

#### 4.4 Greek *lipon*

Greek *lipon* overlaps mainly with German *also* in the initialising-narrative function and with English *so* in indication of thematic switch. Our hypothesis is the following: Due to the overlapping functional spectrum of English *so*, and German *also* with *lipon*, we expect a more frequent use of overlapping functions in heritage Greek.

Examining the use of *lipon* across majority and heritage Greek, we see that although all groups use *lipon* to initialise narrations, other uses of *lipon*, such as the indication of thematic switch, are only found in majority Greek speakers. Data are shown in Table 7.

Although the use of *lipon* in the two HS groups is not as high as the use of *lipon* in majority Greek speakers, all groups use *lipon* to initialise their narrative. This means that they either use it in the very beginning of the text, as in Example (22), or they use it after they provide an opening (such as greeting) and then start their main narrations. In addition, only the majority Greek speakers also use *lipon* to introduce thematic switch, such as in Example (23), in which the

speaker uses *lipon* to switch back to the main narration after having inserted a personal comment.

(22) initialise narrative

lipon otan icha scholasi ap to scholio ke pijena spiti  
 DM when have.PST.1s finish.PST.1s from the school and go.PST.1s home  
 ksafnika icha dhi pos mia mpala  
 suddenly have.PST.1s see.PST.PTCP that a ball

‘*lipon* when I finished school and was going home suddenly I saw that a ball.’ [DEbi57MG\_fsG]

(23) thematic switch

vevea tora entaksi ine dheka metra makria opote mu  
 certainly now alright be.PRS.3s ten meters away so me.GEN.1s  
 fanike paralogho e: i sinechia lipon ine oti apla  
 seem.PST.13s illogical uh the continuity DM be.PRS.3s that simply  
 chtipai to piso amaksi to mprosta amaksi  
 hit.PRS.3s the back car the front car

‘certainly now okay it’s ten meters away so it seemed crazy to me uh  
 next *lipon* is that the car at the back simply hits (-) the car in the front.’

[GRmo10MG\_fsG]

Table 7: Percentages of the most frequent functions of ‘*lipon*’ across countries/speaker groups

| Function             | MS     |         | HS  |
|----------------------|--------|---------|-----|
|                      | Greece | Germany | US  |
| initialise narrative | 75.8   | 75.0    | 100 |
| thematic switch      | 16.2   | 16.7    | -   |

On the basis of partial functional overlap between *lipon* and the English *so* and *lipon* and the German *also*, we expected to find extended functions of *lipon* in the data of HSs. However, our analysis did not confirm this hypothesis. HSs used *lipon* almost exclusively to initialise narrations. This restricted functional spectrum in the HSs’ use of *lipon* might indicate that the initialising function is the most transparent function of this marker and that there is probably no real perceived functional overlap from the perspective of the HSs.

#### 4.5 Greek *etsi*

Greek *etsi* has the emerging function of indicating a consequence. In this function it overlaps with German *also* and English *so*. We hypothesise the following: Due to functional similarities of English *so*, and German *also* with *etsi*, we expect a more frequent use of overlapping functions in heritage Greek. Across all speaker groups, we find functions of *etsi* consistent with the literature. Summary data are given in Table 8. Similar to German *also*, *etsi* indicates a consequence in Example (25). In Example (24), it is similar to the adverbial use of German *so* that indicates a way or manner. In majority Greek, we exclusively find instances of *etsi* in which *etsi* functions as adverb, and hence not as a DM (24), and as a consecutive connector (25). Tsampoukas (2015) identifies the consecutive connector function as a DM function.

- (24) adverb
- na sas apodhiksi oti ontos **etsi** sinevin/sinevisan ta  
 to you.PL prove.SBJ.3SG that indeed like happen/happen.PST.3PL the  
 pragmata  
 things  
 'to prove to you that things happened *etsi/like* this' [GRmo08FG\_fsG]
- (25) consequence
- to amaksi pu ine piso pu itan piso ap afty t=  
 the car that be.PRS.3SG behind that be.PST.3SG behind from this the  
 amaksi dhe prolave na patisi freno ki **etsi** trakaran  
 car not get.PST.3SG to push.SBJ.3SG brake and so bump.PST.3PL  
 'the car that is behind that was behind that car did not get to brake and  
*etsi/so* they bumped' [GRmo76MG\_isG]

In heritage Greek in the US and Germany, however, we observe functional extension of *etsi*. Although heritage Greek speakers in Germany use *etsi* almost as frequently as majority Greek speakers, heritage Greek speakers in Germany extend the use of *etsi* to indicate instances of corrections or restarts (26) and elaborations/specification (27). This corresponds to common functions of German *also*. Heritage Greek speakers in the US use relatively fewer instances of *etsi* compared to the two other speaker groups. Despite the low number of *etsi* occurrences in the spoken narrations of heritage Greek speakers in the US, the use of *etsi* is again extended. Thus, *etsi* is used for indicating the end of a narrative, as in Example (28), which is a common function of majority English *so*. Overall, the functional extension of *etsi* seems to be a convergence with the respective majority languages.

- (26) indicate correction/restart  
ey ghia su ti kanis dhe pistevis ti idha simera  
ey hello you how do.PRS.2SG not believe.PRS.2SG what see.PST.1SG today  
itane etsi perpatusane dhio  
be.PRT.3SG DM walk.PST.3SG two  
'hey hello how are you you won't believe what I saw today two people  
were *etsi* were walking' [DEbi07FG\_isG]

(27) specification  
ghiat i etrekse ke o antras meta pros ti meria tu skiliu  
because run.PST.3SG and the man then towards the side the dog  
ki etsi itane oli mazi kapos ston aftokinitodromo  
and DM be.PST.3PL all together somehow on.the motorway  
'because the man also ran towards the side of the dog and *etsi* they were  
all together somehow on the driveway' [DEbi03FG\_fsG]

(28) end narrative  
ke stamatise o enas o mprostinos etsi ksafnika ke o allos  
and stop.PST.3SG the one the front DM suddenly and the other  
ap=p/ apo piso ton trakarise ke etsi  
fr=f/from behind him bump.PST.3SG and DM  
'and the one in the front stopped suddenly and the other one at the back  
bumped on him and *etsi*' [GRmo76MG\_isG]

The findings on *etsi* were different than what we expected: We did not find a more frequent use of overlapping functions, but HSs of Greek in Germany and the US extend the functional spectrum of *etsi* according to common functions of the respective majority language counterpart.

Table 8: Percentages of the most frequent functions of *etsi* across countries/speaker groups

| Function    | MS     | HS      |      |
|-------------|--------|---------|------|
|             | Greece | Germany | US   |
| adverb      | 59.2   | 36.4    | 79.2 |
| consequence | 40.8   | 47.7    | 4.2  |

#### 4.6 Three-dot sign

In analyzing data containing the three-dot sign, we used the indicators of position and context to assign discourse functions. First, we report findings for individual languages. In Table 9 we present raw numbers and percentages for functions that we assigned based on the position. The main finding regarding language specifics is that textual functions are predominant across languages but with a larger variation in German, Russian and Turkish compared to English and Greek (cf. Labrenz et al. 2022: 260–61).

Table 9: Raw numbers (*n*) and percentage use (%) of functions per language relative to all uses of the three-dot sign in that language

|         | Textual  |      | Subjective |     | Intersubjective |      |
|---------|----------|------|------------|-----|-----------------|------|
|         | <i>n</i> | %    | <i>n</i>   | %   | <i>n</i>        | %    |
| English | 22       | 88.0 | 0          | 0.0 | 3               | 12.0 |
| German  | 52       | 68.4 | 4          | 5.3 | 20              | 26.3 |
| Greek   | 17       | 81.0 | 0          | 0.0 | 4               | 19.0 |
| Russian | 15       | 62.5 | 1          | 4.2 | 8               | 33.3 |
| English | 12       | 63.2 | 1          | 5.3 | 6               | 31.6 |

The second indicator (cf. Labrenz et al. 2022: 261–62) was the context which is especially relevant in message-final position and in the position between two discourse units. In these positions the three-dot sign has two sources of interpretation, and can therefore have additional subjective or intersubjective functions depending on the preceding discourse unit. Taken together, the findings from Tables 9 and 10 show a tendency for German, Russian, and Turkish to use the three-dot sign relatively more frequently compared to English and Greek in intersubjective and subjective functions and polyfunctionally.

Regarding potential dynamics in bilinguals (cf. Labrenz et al. 2022: 264–66), we found an overall tendency of bilinguals to adapt the three-dot use in their heritage language to that in the respective majority language in two areas: functional variation and frequency of use. With respect to frequencies, the following pattern emerged for HSs of Turkish and Russian: Just as MSs of German use the three-dot sign more frequently than MSs of English, so HSs of Turkish and Russian with German as their majority language use it more frequently than HSs of Turkish and Russian with English as their majority language (see Table 11). The numbers in brackets are occurrences of the three-dot sign per 100 communicative

Table 10: Raw numbers (*n*) and percentage use (%) of monofunctional (either textual, subjective, or intersubjective) and polyfunctional (additional (inter)subjective functions) use per language relative to all uses of the three-dot sign in that language

|         | Monofunctional |      | Polyfunctional |      |
|---------|----------------|------|----------------|------|
|         | <i>n</i>       | %    | <i>n</i>       | %    |
| English | 17             | 68.0 | 8              | 32.0 |
| German  | 39             | 51.3 | 37             | 48.7 |
| Greek   | 17             | 81.0 | 4              | 19.1 |
| Russian | 12             | 50   | 12             | 50   |

Table 11: First pattern for HSs of Turkish and Russian

| Germany (DE)   |        | US |                       |
|----------------|--------|----|-----------------------|
| MSs-German     | (2.71) | >  | MSs-English (0.96)    |
| HSs-Turkish/DE | (1.89) | >  | HSs-Turkish/US (0.73) |
| HSs-Russian/DE | (1.5)  | >  | HSs-Russian/US (0.52) |

units (roughly all independent sentences) in the respective group/language. MSs of German and English includes mono- as well as bilingually raised speakers). Another pattern that emerged is the tendency within bilingual speakers with Greek and Russian as heritage languages to use the three-dot sign relatively more frequently in their majority language (German or English) than in their heritage language (see Table 12).

Table 12: Second pattern for HSs of Greek and Russian

| HS of Greek      |        |   |                            |
|------------------|--------|---|----------------------------|
| majority-German  | (1.1)  | > | heritage-Greek/DE (0.73)   |
| majority-English | (1.52) | > | heritage-Greek/US (0.79)   |
| HS of Russian    |        |   |                            |
| majority-German  | (3.0)  | > | heritage-Russian/DE (1.5)  |
| majority-English | (1.16) | > | heritage-Russian/US (0.52) |

To sum up, we found similarities as well as slight differences in the use across countries and languages, as well as across bilingual speakers' two languages. Cross-linguistically, the three-dot sign is used for discourse organisation and segmentation in digital informal writing (see also Busch 2021 on German, and Androutsopoulos 2020 on Greek). In terms of language contact, our data indicate that patterns of use in the majority language influence the use in the heritage language. Additionally, we found that within bilingual speakers the use of the three-dot sign in the heritage language is less frequent and with a smaller functional spectrum compared to their majority language use, maybe pointing to an insecurity in using such salient markers of informality in their heritage language (see Labrenz et al. 2022: 259–266 for more details on results).

## 5 Discussion

Taken together, our findings indicate similarities between verbal and graphic DMs concerning the impact of societal language status, that is, the status of a contact language as a majority language or as a minoritised heritage language: We observed an influence of the majority language on the use of DMs in the heritage language in terms of frequencies, functional extension/restriction, and pragmaticalisation processes.

More specifically, we find functional extensions, restrictions and/or a more frequent use of specific functions in heritage German *so*, heritage Turkish *yani* and *iste*, heritage Greek *lipon* and *etsi*, pointing to convergence with the functional spectrum of a counterpart DM in the majority language (either *so* in English or *so* or *also* in German). This might point to a shared pool of functions in the bilingual mind with stronger activations of additional or overlapping functions of the (societally) dominant language when the heritage language is used than vice versa. Additionally, we find tendencies for language internal developments for Turkish *yani* and *iste*, and Greek *etsi* which seem to be especially dynamic in contact situations. As for graphic DMs, specifically the three-dot sign, we observed language-specific trends, which we found particularly interesting given its translinguistic status. For this graphic DM, we also observed a tendency for the majority language to influence heritage language use, even though graphic markers have no specific language affiliation and exhibit common functions across languages (such as structuring discourse in the case of the three-dot sign).

The observed influence from the majority onto the heritage language might be on one hand favoured by the dominance of the majority language in countries with a strong monolingual bias such as Germany and the US. On the other

hand, innovations in heritage languages might be generally favoured because of the lower presence of normative authorities such as school education in that language. Our findings indicate that this might be especially true for less salient phenomena, such as pattern replication. In contrast, more noticeable phenomena, like matter replication or the use of informal register markers such as the three-dot sign, might be used more carefully, at least in the context of a monolingual mode, which is a condition of the elicitation method (see also Wiese et al. 2025 [this volume]). This, in turn, might indicate that in a communicative situation characterised by a monolingual mode, HS might not only be careful in their choice of shared language resources, but also in their use of salient graphic register markers in their heritage language.

## 6 Conclusion

This chapter has shed light on the dynamics of DMs in language contact situations, both in formal and in informal-spoken discourse as well as in informal-written texts. By exploring the graphic domain through the inclusion of informal-written messages, we underline the crucial role of informality in DM usage, demonstrating its presence beyond spoken language.

Our key findings reveal similar trends in language contact for both verbal and graphic DMs. First, the societal status of a language has an impact on pattern replication and/or frequency in use of DMs. Secondly, pragmaticalisation processes can be accelerated in specific language contact scenarios.

These two findings are highly relevant for persistent issues in the field. First, by addressing functional convergence in the discourse domain, we have ventured into a less explored aspect of DMs in language contact, which complements the extensively studied area of borrowings (see also Wiese et al. 2025 [this volume] for borrowed or translanguaged DMs in the RUEG data). The fact that HS replicated relatively few actual word forms might be due to the monolingual mode deliberately induced by the RUEG setup. This suggests that in situations with monolingual interlocutors, HSs are more likely to resort to pattern replication than to matter replication, since matter replication is more salient and may not be perceived as appropriate by the speakers themselves in such contexts (cf. Matras 2010). This is also supported by the fact that translanguaging of DMs only occurred in the informal data, a communicative situation in which speakers are less exposed to normative pressures than in a more formal situation (see Wiese et al. 2025 [this volume]). Moreover, in our examination involving the majority language and the heritage language of bilingual speakers, we were able to

distinguish the effects of bilingualism in general from the societal status of the languages: While we found no effects of bilingualism in the use of the majority language, we found an influence of the majority language on bilingual speakers' use of the heritage language.

Second, we identified a link between societal status and the process of pragmatisation, observing two tendencies in heritage languages. The first tendency exemplifies a typical path of pragmatisation: From an adverbial and/or semantically transparent connective through a process of semantic bleaching into uses with mere discourse functions. This trend holds for both majority and heritage languages. Notably, Greek *etsi* is of particular interest as it appears to be in the early stages of this process, assuming a new function as a semantically transparent consecutive connector. In heritage Greek, this process seems to have progressed further, as more DM functions can already be found here. The second tendency is the transformation of relatively opaque DMs into more transparent connectors indicating an inverted pragmatisation process, a phenomenon primarily observed in language contact situations. This is in line with Matras (Matras & Sakel 2007, Matras 2010), who claims that "[...] a model of convergence must also be able to account for potential exceptions to the unidirectionality of grammaticalization" (Matras 2010: 71). In our data, Turkish in contact with German and English showcases this development, potentially influenced by typological differences between Turkic and Indo-European languages: German and English typically rely on finite clauses with pre-posed subordinating connectors, while in Turkish, non-finite subordination is preferred, without the employment of connectors with word status. Finite subordination with pre-posed or circum-connectors is possible; however, in the heritage varieties we note an increase in this pattern of subordination. DMs like *iste* and *yani* belong to the pool of words from which the language recruits new connectors following this dynamic. This leads to an inverted pragmatisation path from DMs to connectors, particularly in the contact situations.

Taken together, our findings point to three main factors important for studies on the dynamics of discourse markers: 1) contact scenarios, offering insights into their functional convergence, 2) the influence of societal language status, and 3) the impact of informality beyond spoken language. The study of both verbal and graphical DMs in different language contexts and contact scenarios contributes to a more comprehensive picture of the dynamics and evolution of these linguistic phenomena.

As a last point, we want to discuss the observed unidirectional influence of the majority language on the heritage language use of verbal and graphic DMs. This may be due to the dominant monolingual bias in Germany and the United States.

In such societal settings, heritage languages are challenged by strong hegemonic majority languages. In contrast, in multilingually oriented societies (for instance, in many Asian and African countries), crosslinguistic influences may be more fluid, allowing for bidirectional pattern replication. In future research, it would be interesting to build on our results in comparative studies that include heritage languages from such settings.

## Acknowledgements

We would like to thank our amazing and dedicated student assistants, each of whom was able to bring their own unique strengths to project P6 and P9: Ilyuza Akhmetzioanova, Christian Anders, Yeşim Bayram, Franziska Groth, Oğuzan Kuyrukçu, Alexander Lehman, Ksenia Nowak, Hannah Plückebaum, Simge Türe, Sharon Rauschenbach, Guendalina Reul, Tjona Sommer, Charlott Thomas, and Barbara Zeyer. We further thank Lea Coy for help with pre-publication formatting.

We also thank two external anonymous reviewers and three internal reviewers Oliver Bunk, Johanna Tausch, and Wintai Tsehayé for their helpful and very constructive suggestions. The research was supported through funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit *Emerging Grammars in Language Contact Situations*, projects P6 (394838878) and P9 (313607803)

## References

- Aijmer, Karin. 2011. *Well I'm not sure I think...: The use of well by non-native speakers*. *International Journal of Corpus Linguistics* 16(2). 231–254. DOI: 10.1075/ijcl.16.2.04aij.
- Alm, Maria. 2007. *Also darüber lässt sich ja streiten! Die Analyse von also in der Diskussion zu Diskurs- und Modalpartikeln*. Stockholm: Almqvist & Wiksell.
- Andersen, Gisle. 2001. *Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents*. Amsterdam: John Benjamins.
- Androutsopoulos, Jannis. 2020. Auslassungspunkte in der schriftbasierten Interaktion: Sequenziell-topologische Analysen an Daten von griechischen Jugendlichen. In *Register des Graphischen: Variation, Interaktion und Reflexion in der digitalen Schriftlichkeit*, 133–158. Berlin: De Gruyter. DOI: 10.1515 / 9783110673241-006.

- Antoniou, Kyriakos. 2008. *Discourse markers in oral discourse* [Δείκτες λόγου στον προφορικό λόγο]. Athens, Greece: National & Kapodistrian University of Athens, School of Philosophy, Philology Department. (Undergraduate dissertation.).
- Ariel, Mira. 1994. *Pragmatic operators: The encyclopedia of language and linguistics*. Oxford: Pergamon Press.
- Auer, Peter & Susanne Günthner. 2005. Die Entstehung von Diskursmarkern im Deutschen: Ein Fall von Grammatikalisierung? In Torsten Leuschner & Tanka Mortelmans (eds.), *Grammatikalisierung im Deutschen*, 335–362. Berlin: de Gruyter. DOI: 10.1515/9783110925364.335.
- Baron, Naomi S. & Rich Ling. 2011. Necessary smileys & useless periods: Redefining punctuation in electronically-mediated communication. *Visible Language* 45(1/2). 45–67.
- Blakemore, Diane. 2004. *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486456.
- Blühdorn, Hardarik, Arnulf Deppermann, Henrike Helmer & Thomas Spranz-Fogasy (eds.). 2017. *Diskursmarker im Deutschen: Reflexionen und Analysen*. Göttingen: Verlag für Gesprächsforschung.
- Bolden, Galina B. 2009. Implementing incipient actions: The discourse marker *so* in English conversation. *Journal of Pragmatics* 41(5). 974–998. DOI: 10.1016/j.j pragma.2008.10.004.
- Bredel, Ursula. 2008. Die Interpunktions des Deutschen: Ein kompositionelles System zur Online-Steuerung des Lesens. In *Die Interpunktions des Deutschen* (Linguistische Arbeiten 522). Berlin/New York: Max Niemeyer Verlag. DOI: 10.1515/9783484970502.
- Brinton, Laurel J. 2017. *The evolution of pragmatic markers in English: Pathways of change*. Cambridge: Cambridge University Press. DOI: 10.1017/9781316416013.
- Busch, Florian. 2021. *Digitale Schreibregister: Kontexte, Formen und metapragmatische Reflexionen*. Berlin & Boston: De Gruyter. DOI: 10.1515/9783110728835.
- Christodoulidou, Maria. 2014. The meaning of discourse markers in classroom interaction. In *Major Trends in Theoretical and Applied Linguistics* 2, 11–26. De Gruyter Open Poland. DOI: 10.2478/9788376560885.p11.
- Crible, Ludivine. 2018. *Discourse markers and (dis)fluency: Forms and functions across languages and registers* (Pragmatics & Beyond New Series 286). Amsterdam: John Benjamins. DOI: 10.1075/pbns.286.
- Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49(2). 365–390. DOI: 10.1515/ling.2011.011.

- Ehlich, Konrad. 2007. so: Überlegungen zum Verhältnis sprachlicher Formen und sprachlichen Handelns, allgemein und an einem widerspenstigen Beispiel. In Konrad Ehlich (ed.), *Sprache und sprachliches Handeln*, vol. 1, 141–167. Berlin, New York: De Gruyter. DOI: 10.1515/9783110922721.
- Fernández-Villanueva, Marta. 2007. Uses of *also* in oral semi-informal German. *Catalan Journal of Linguistics* 6(1). 95–115. DOI: 10.5565/rev/catjl.126.
- Flores, Cristina, Ayşe Gürel & Michael T. Putnam. 2019. Different perspectives on critical factors in heritage language development and maintenance. *Language Learning* 70(S1). 5–14. DOI: 10.1111/lang.12367.
- Fraser, Bruce. 1999. What are discourse markers? *Journal of Pragmatics* 31(7). 931–952. DOI: 10.1016/S0378-2166(98)00101-5.
- Fraser, Bruce. 2006. Towards a theory of discourse markers. In Kerstin Fischer (ed.), *Approaches to Discourse Particles* (Studies in Pragmatics 1), 189–204. Amsterdam: Elsevier. DOI: 10.1163/9780080461588\_012.
- Fuller, Janet M. 2001. The principle of pragmatic detachability in borrowing: English-origin discourse markers in Pennsylvania German. *Linguistics* 39(2). 351–369. DOI: 10.1515/ling.2001.014.
- Georgakopoulou, Alexandra & Dionysis Goutsos. 1998. Conjunctions versus discourse markers in Greek: The interaction of frequency, position, and functions in context. *Linguistics* 36(5). 887–918. DOI: 10.1515/ling.1998.36.5.887.
- Goss, Emily L. & Joseph C. Salmons. 2000. The evolution of a bilingual discourse marking system: Modal particles and English markers in German-American dialects. *International Journal of Bilingualism: Cross-disciplinary, Cross-linguistic Studies of Language Behavior* 4(4). 469–484. DOI: 10.1177/13670069000040040501.
- Heine, Bernd, Gunther Kaltenböck, Tania Kuteva & Haiping Long. 2021. *The rise of discourse markers*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108982856.
- Hennig, Mathilde. 2006. *Grammatik der gesprochenen Sprache in Theorie und Praxis*. Kassel: Kassel University Press.
- Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization*. 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139165525.
- Imo, Wolfgang. 2017. Diskursmarker im gesprochenen und geschriebenen Deutsch. In Hardarik Blühdorn, Arnulf Deppermann, Henrike Helmer & Thomas Spranz-Fogasy (eds.), *Diskursmarker im Deutschen: Reflexionen und Analysen*, 49–72. Göttingen: Verlag für Gesprächsforschung.

- Katsika, Kalliopi, Annika Labrenz, Kateryna Iefremenko & Shanley E.M. Allen. 2025. Discourse openings and closings across languages in contact. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 527–576. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775189.
- Keskin, Cem. 2025. Balkan Turkic as a model for understanding contact-induced change in Turkish. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 287–323. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775175.
- Keskin, Cem, Kateryna Iefremenko, Jaklin Kornfilt & Christoph Schroeder. N.d. *Aspects of clause combining in Turkish language contacts*. MS, University of Potsdam & Syracuse University. MS, University of Potsdam & Syracuse University.
- Konerding, Klaus-Peter. 2004. Semantische Variation, Diskurspragmatik, historische Entwicklung und Grammatikalisierung: Das Phänomenspektrum der Partikel also. In Inge Pohl & Klaus-Peter Konerding (eds.), *Stabilität und Flexibilität in der Semantik: Strukturelle, kognitive, pragmatische und historische Perspektiven*. 198–237. Frankfurt a. M.: Peter Lang.
- Labrenz, Annika. 2023. Functional variation of German *also* across registers and speaker groups. *Contrastive Pragmatics* 4(2). 289–320. DOI: 10.1163/26660393-bja10077.
- Labrenz, Annika, Heike Wiese, Tatiana Pashkova & Shanley E. M. Allen. 2022. The three-dot sign in language contact. *Pragmatics & Cognition* 29(2). 246–271. DOI: 10.1075/pc.21021.lab.
- Maschler, Yael. 2000. Discourse markers in bilingual conversation. *International Journal of Bilingualism* 4(4). 437–561.
- Maschler, Yael. 2009. *Metalanguage in interaction: Hebrew discourse markers* (Pragmatics & Beyond New Series 181). Amsterdam/Philadelphia: John Benjamins.
- Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36(2). 281–332. DOI: 10.1515/ling.1998.36.2.281.
- Matras, Yaron. 2010. Contact, convergence, and typology. In *The handbook of language contact*, 66–85. Oxford: John Wiley & Sons. DOI: 10.1002/9781444318159.ch3.
- Matras, Yaron. 2020. *Language contact*. 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Matras, Yaron & Jeanette Sakel. 2007. *Grammatical borrowing in cross-linguistic perspective*. Berlin/New York: De Gruyter Mouton. DOI: 10.1515/9783110199192.

- Meibauer, Jörg. 2019. How omission marks mark omission ...: An inquiry into the graphemics/pragmatics interface. In Claudia Claridge & Merja Kytö (eds.), *Punctuation in context: Past and present perspectives*, 67–83. Frankfurt a. M.: Peter Lang.
- Moosegard Hansen, Maj-Britt. 1998. *The function of discourse particles: A study with special reference to spoken standard French*. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/pbns.53.
- Müller, Simone. 2005. *Discourse markers in native and non-native English discourse* (Pragmatics & Beyond New Series 138). Amsterdam/Philadelphia: John Benjamins.
- Özbek, Nihal. 1995. *Discourse markers in Turkish and English: A comparative study*. Nottingham University. (Doctoral dissertation).
- Özbek, Nihal. 2000. *Yani, işte, şey, ya: Interactional markers of Turkish*. In Aslı Göksel & Celia Kerslake (eds.), *Proceedings of the 9th International Conference on Turkish Linguistics*, 393–401. Wiesbaden: Harroowitz.
- Özsoy, Onur, Kateryna Iefremenko & Christoph Schroeder. 2022. Shifting and expanding clause combining strategies in heritage Turkish varieties. *Languages* 7(3). 242. DOI: 10.3390/languages7030242.
- Pfeiffer, Martin. 2017. Über die Funktion der Reparaturmarker im Deutschen. In Hardarik Blühdorn, Arnulf Depermann, Henrike Helmer & Thomas Spranz-Fogasy (eds.), *Diskursmarker im Deutschen: Reflexionen und Analysen*, 259–283. Göttingen: Verlag für Gesprächsforschung.
- Pichler, Heike. 2016. Introduction: Discourse-pragmatic variation and change. In Heike Pichler (ed.), *Discourse-pragmatic variation and change in English: New methods and insights*, 1–18. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781107295476.001.
- Pons Bordería, Salvador & Kerstin Fischer. 2021. Using discourse segmentation to account for the polyfunctionality of discourse markers: The case of well. *Journal of Pragmatics* 173. 101–118. DOI: 10.1016/j.pragma.2020.11.021.
- Raclaw, Joshua. 2006. Punctuation as social action: The ellipsis as a discourse marker in computer-mediated communication. *Berkeley Linguistics Society* 32(1). 299–306. DOI: 10.3765/bls.v32i1.3469.
- Redeker, Gisela. 1990. Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14(3): Selected papers from The International Pragmatics Conference, Antwerp, 17-22 August, 1987. 367–381. DOI: 10.1016/0378-2166(90)90095-U.
- Rosenthal, Dietmar E. 2012. *Spravochnik po pravopisaniyu I literaturnoi pravke [Handbook of orthography and literary editing]*. Moskow: Iris Press.

- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2). 155–163. DOI: 10.1177/1367006909339814.
- Salmons, Joseph C. N.d. Bilingual discourse marking: Code switching, borrowing, and convergence in some German-American dialects. *Linguistics* 28(3) (). 453–480. DOI: 10.1515/ling.1990.28.3.453.
- Schiffrin, Deborah. 1987. *Discourse markers* (Studies in Interactional Sociolinguistics 5). Cambridge: Cambridge University Press. DOI: 10 . 1017 / CBO9780511611841.
- Schourup, Lawrence. 1999. Discourse markers. *Lingua* 107(3-4). 227–265. DOI: 10.1016/S0024-3841(96)90026-1.
- Schroeder, Christoph, Kateryna Iefremenko, Kalliopi Katsika, Annika Labrenz & Shanley E. M. Allen. 2025. Clause combining in narrative discourse: A contrastive analysis across heritage and majority languages. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 251–286. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775173.
- Schumann, Kathleen. 2021. *Der Fokusmarker so: Empirische Perspektiven auf Gebrauch und Verarbeitung eines Ausnahmeelements* (Linguistik: Impulse & Tendenzen 93). Berlin & Boston: De Gruyter. DOI: 10.1515/9783110731149.
- Stavropoulos, Dimitri N.. 1988. *Oxford Greek-English learner's dictionary*. 3rd edn. Oxford: Oxford University Press.
- Thurmair, Maria. 2001. *Vergleiche und Vergleichen: Eine Studie zu Form und Funktion der Vergleichsstrukturen im Deutschen* (Linguistische Arbeiten 433). Berlin & Boston: Max Niemeyer Verlag. DOI: 10.1515/9783110927054.
- Triantafyllidis, Manolis. 2019. *Modern Greek grammar [Νεοελληνική Γραμματική, Αναπροσαρμογή της Μικρής Νεοελληνικής Γραμματικής]*. Athens, Greece: Manolis Triantafyllidis Foundation.
- Tsampoukas, Alexandros. 2015. *Grammaticalization in process: The case of etsi [Γραμματικοποίηση σε εξέλιξη: Η περίπτωση του έτσι]*. Presentation. Athens, Greece.
- Turan, Feride. 2014. *Türk Dil Kurumu: Yazım Kilavuzu ve Kurallar*.
- Tzartzanos, Achilleas. 2002. *Greek syntax [Νεοελληνική σύνταξις (της κοινής δημοτικής)]*. 2nd edn. Thessaloniki, Greece: Kiriakidis brothers publishing.
- Wiese, Heike. 2011. So as a focus marker in German. *Linguistics* 49(5). 991–1039. DOI: 10.1515/ling.2011.028.
- Wiese, Heike. 2012. *Kiezdeutsch: Ein neuer Dialekt entsteht*. Originalausgabe, zweite, durchgesehene Ausgabe (Beck'sche Reihe 6034). München: Verlag C. H. Beck.

- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Anke Lüdeling, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Tracy Rosemarie, Christoph Schroeder, Luka Szucsich, Wintai Tsehayé, Sabine Zerbian & Yulia Zuban. 2021. *RUEG Corpus*. Version 0.4.0. Zenodo. DOI: 10.5281/zenodo.3236068.
- Wiese, Heike & Annika Labrenz. 2021. Emoji as graphic discourse markers: Functional and positional associations in German WhatsApp messages. In Daniël Van Olmen & Jolanta Šinkūnienė (eds.), *Pragmatic markers and peripheries* (Pragmatics & Beyond New Series), 277–302. Amsterdam: John Benjamins. DOI: 10.1075/pbns.325.10wie.
- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 33–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775159.
- Yılmaz, Erkan. 1994. *Descriptive and comparative study of the discourse markers well in English and şey in Turkish*. University of Essex. (MA thesis).
- Yılmaz, Erkan. 2004. *A pragmatic analysis of Turkish discourse particles: Yani, işte and şey*. Middle Eastern Technical University, Ankara, Turkey. (Doctoral dissertation).

# Chapter 15

## Discourse openings and closings across languages in contact

Kalliopi Katsika<sup>a</sup>, Annika Labrenz<sup>b</sup>, Kateryna Iefremenko<sup>c,d</sup>  
& Shanley E.M. Allen<sup>a</sup>

<sup>a</sup>University Kaiserslautern-Landau <sup>b</sup>Humboldt-Universität zu Berlin

<sup>c</sup>Leibniz-Centre General Linguistics <sup>d</sup>University of Potsdam

This chapter presents an exploratory study focused on the analysis of macro discourse strategies across various speaker groups using data from the Research Unit *Emerging Grammars* (RUEG) corpus. Specifically, the present study examines openings and closings – the parts of a narration or conversation that precede and follow the main discourse phase – in adult and adolescent heritage speakers of Greek, German, Russian and Turkish and monolingual speakers of Greek, German, Russian and Turkish. The main aim of this study is to unveil the complex dynamics of discourse organization in multilingual contexts and to investigate the role of age, language status and communicative situation in the distribution of discourse functions in openings and closings across languages in contact. Our methodology includes the creation of an annotation scheme through identification and subsequent annotation of textual, subjective and intersubjective functions in Greek, German, Russian and Turkish openings and closings in the narrations of heritage and monolingual speakers. Statistical analyses revealed similarities and differences in the distribution of openings and closings, and the functions included in them, across speaker groups. The research and analyses included in this chapter aim to contribute to a better understanding of macro discourse organization in language contact situations.



## **1 Introduction**

### **1.1 Openings and closings in monolingual and bilingual speakers**

Discourse openings and closings are defined as the linguistic material that precedes or follows the core phase of the discourse (Schegloff 1968, 2007, Schegloff & Sacks 1973, Luke 2002, Pavlidou 2014). Their typical functions include defining textual boundaries that set the core of the text apart from the framing, introducing or closing of a new topic, setting the time frame and/or space of the event, and interpreting and evaluating the event (Labov 1972, Tolchinsky et al. 2002, Berman & Katzenberger 2004). The construction of discourse borders through openings and closings (Schegloff 2007) has been shown to be an area of dynamic difference across languages (e.g. Luke & Pavlidou 2002), but has rarely been studied in situations of language contact (but see Dollnick & Pfaff 2013). Our study fills this gap in the literature by exploring discourse opening and closing patterns in monolingually raised speakers in Germany, Greece, Russia, and Turkey (speakers who were raised in a monolingual environment), and heritage speakers of German, Greek, Russian and Turkish in the USA and Germany. All heritage speakers provided narrations in their heritage and majority languages (either English or German). Although we have annotated and analyzed majority language data too, we do not present them in the current chapter for space reasons.

Heritage speakers (HSs) are speakers who grow up acquiring a heritage language in the home environment in addition to the language of the larger society which is different than the heritage language (see Wiese et al. 2025 [this volume]). In many cases, HSs gradually become dominant in the societal language. Thus, although HSs acquire their heritage language as a first language (L1), they may end up using their heritage language in only certain communicative situations, such as among family members or peers (see e.g. Rothman 2009, Montrul & Ionin 2012, Polinsky 2018, Tsehayé et al. 2021). The restricted use of a heritage language sometimes results in notable differences between the language that is spoken by the HSs and the homeland variety of the language. In contrast to speakers of the homeland variety of the language, HSs acquire their L1 in a language contact situation (heritage language and societal language). In addition, HSs may transfer from the majority language (which is also usually HSs' dominant language) to the heritage language, or they may create unique forms and structures that are not found directly or transparently in either of their languages (see Polinsky 2018, Özsoy et al. 2022). Understanding the discourse abilities of HSs is crucial for comprehending the complexities of heritage language development and the maintenance of linguistic heritage.

Openings and closings have been a central topic in the study of discourse since the early 1970s, often within the framework of Conversation Analysis (see review in Schegloff 2007) and also in the discourse analysis of narratives (e.g. Tolchinsky et al. 2002). Openings and closings have played a key role in demonstrating the formulaic nature of routines in both interpersonal interactions (e.g., telephone conversations, face-to-face communication) and monologic productions (e.g., narratives, expository texts). Since our data come from the Research Unit *Emerging Grammars* (RUEG) corpus (Wiese et al. 2021), which combines elements of both interpersonal interactions and narratives, we focus on that literature here.

In interpersonal interactions, openings typically begin with an exchange of greetings (e.g., *hi*), then initial inquiries (e.g., *how are you*), and finally an anchor position where the reason for the interaction is introduced (e.g., *I called to let you know that ...*) (Schegloff 1968). Emanuel Schegloff, one of the pioneers in the study of openings and closings in conversational environments, identified four ‘core sequences’ in openings of telephone conversations in English: (1) summons/response, (2) identification/recognition, (3) a greeting; and (4) a “how are you” sequence (Schegloff 1968, 1979). These patterns were initially observed in data from English, and although these basic sequences seem to hold across cultures and languages, many differences are also apparent (Luke 2002, Pavlidou 2014). For example, the “initial inquiries” phase of introductions is typically reduced or omitted in German and Finnish, while it is often extended in Greek and Spanish. The content of this phase is often different as well, with politeness routines playing a main role in Persian, humor in Greek, and health inquiries in Spanish. Closings in English interactions typically comprise a pre-closing (e.g., *okay*, *I have to go*) and a closing greeting (e.g., *goodbye*, *have a good day*) (Schegloff & Sacks 1973). A similar pattern holds in cross-linguistic studies so far, but again the content of the phases varies to some degree (Luke 2002, Pavlidou 2014). For example, the “pre-closing” tends to be more direct in Greek, Chinese, and Spanish, using phrases equivalent to *that's all for now* and *that's about it*. In addition, since closings in Greek, Chinese, and Spanish show a higher tendency towards repetition and a higher use of phrases of relationship affirmation or even thanking for something implying politeness/ positive attitude in communication settings (see Gkouma 2023 for Greek), closings tend to be longer in those languages than in English and German. As pointed out by Márquez Reiter & Luke (2010), Schegloff’s sequences should not be taken as a ‘must’ in every phone conversation. In fact, after analyzing twenty-five openings in English, Hopper (1989) showed that most of the interactions did not actually follow Schegloff’s sequences. What Schegloff’s work in the 70s – and also Sacks (1967)’s earlier work – has successfully achieved is to spark interest in the study of openings and closings in conversational set-

tings across different languages, e.g. Arabic (Hopper & Koleilat-Doany 1989), Chinese (Hopper & Chen 1996, Luke 2002), Dutch (Houtkoop-Steenstra 1991, 2002), French (Godard 1977, Hopper & Koleilat-Doany 1989), German (Pavlidou 1994, 1997), Japanese (Park 2002), Persian (Taleghani-Nikazm 2002), Samoan (So'o & Liddicoat 2000), Spanish (Coronel-Molina 1998), Swedish (Lindström 1994), and Vietnamese (Do et al. 2018) with researchers identifying different opening and closing functions cross-linguistically.

In narratives, a somewhat different pattern is found since there is no need for the interpersonal exchange (Berman 1997, Labov 1972). Narratives typically begin with a setting or orientation that identifies relevant aspects of the story to come (e.g., time, place, people involved, activity, situation) or provides initial interpretation and evaluation, followed by an initiating event that launches the sequence of events that constitute the plot line. Closings in narratives typically begin with some kind of summary, followed by an indication that the narrative is completed. Also relevant for narratives is the way in which the speaker establishes their stance towards the story, including what means or point of reference they select for framing their text and how they attract the attention of the interlocutor or reader (Berman 1997, Tolchinsky et al. 2002). Stance can be more general in situating the event in relation to other similar events in the world, or more specific in mentioning a concrete scene or an evaluation or resolution of the event itself.

Parallel to interpersonal interactions, the basic patterns for narrative construction seem to bear similarities across the cultures and languages that have been studied so far, but show clear differences as well. In their study of narratives in Spanish, English and Swedish, for example, Tolchinsky et al. (2002) found that openings and closings were typically longer in Spanish than in English narratives, and also longer in English than in Swedish narratives. In addition, they found that English speakers tended to introduce the topic of the narrative directly while Spanish and Swedish speakers tended to introduce it by means of referring to an external event. In the only study, to our knowledge, on openings and closings in the narratives of bilinguals, Dollnick & Pfaff (2013) showed that Turkish HSs who grew up in Germany showed different approaches in each of their languages both to stance and to the functions conveyed by their openings and closings.

As noted earlier, research on openings and closings has focused predominantly on interpersonal interactions, with some research on monologic productions. However, very little work has been published to date on openings and closings in computer-mediated communication environments. One example is Raclaw (2008)'s study on closings in instant messaging (IM) of university students. Raclaw found that the IM environment led to an extension of the typical

closing routines in spoken conversations. In particular, the pre-closing portion was typically elaborated with accounts for why the speaker initiated the closing (e.g., *so like, i love you and all, but i should probably start my homework :/*) and arrangements to talk at a later date (e.g., *I will talk to you tomorrow, jah?*). Within the accounts, it was common to include a hedge (e.g., *probably*, :) and/or a palliative or apology (e.g., *I love you and all, but ...*). Raclaw hypothesized that closings are extended in IM environments because of perceived difficulty or rudeness of removing oneself from the constant availability that IM offers.

In addition to analyzing opening and closing strategies across languages, our study explores the extent to which adolescents may have different distributional patterns in openings and closings than adults. Previous studies have shown differences in openings and closings across age groups in monolingual and bilingual populations (e.g. Tolchinsky et al. 2002, Dollnick & Pfaff 2013) although differences are mostly found between child and adolescent groups. Studies in several languages show that skills in producing openings and closings in narratives develop with age, from early childhood through and beyond adolescence (Berman & Verhoeven 2002, Berman & Katzenberger 2004, Dollnick & Pfaff 2013, Ravid & Tolchinsky 2002, Tolchinsky et al. 2002).

In sum, openings and closings in discourse constitute a highly dynamic environment that offers potential for numerous emerging patterns in cases of language contact such as those in the RUEG corpus (Wiese et al. 2021). Existing literature shows substantial differences in patterns of openings and closings across languages that could possibly lead to new patterns across the two languages of the speakers. Differences in patterns by age (adolescent vs. adult), by levels of formality (informal vs. formal), and by mode (spoken vs. written) are also evident in the literature. Given that very little research has been conducted on openings and closings in bilingual populations, our corpus study aims to add to our knowledge of discourse patterning in bilinguals. More specifically, the current study focuses on adolescent and adult HSs' discourse organization strategies with the goal of exploring the means that HSs – as well as monolingual speakers (MSs) of the respective languages – use in order to introduce and close narrations extracted in different communicative situations, that is, informal-spoken, formal-spoken, informal-written, formal-written. For example, previous research has shown that speakers may use different opening functions in formal settings (Márquez Reiter & Luke 2010, Zimmerman 1992, 1984). Our study is exploratory in nature and aims to present distributional patterns of discourse openings and closings in monolingual and heritage speakers so as to set the foundations for more detailed analyses in the future.

## 1.2 Research questions and hypotheses

Our research questions are the following:

*Research Question 1:* Do monolingual and heritage speakers of German, Greek, Russian and Turkish exhibit a similar distribution of openings and closings in their narratives?

On the basis of the research across different languages reported in Section 1, we hypothesize that we may find partial evidence for Schegloff's opening sequences (our data do not include actual telephone conversations), and that we may also find variability in opening and closing functions across speaker groups.

*Research Question 2:* Does age play a role in the distribution of openings and closings across monolingual and heritage speakers?

Based on the previous studies by Tolchinsky et al. (2002) and Dollnick & Pfaff (2013), we expect to find differences in the distribution of openings and closings between the adolescent and adult groups.

*Research Question 3:* Does the distribution of intersubjective, subjective and textual functions differ across speaker groups and communicative situations? We expect that formality (formal vs. informal communicative situations) and mode (spoken vs. written) will affect the functions used in openings and closings across language groups. One difference across communicative situations may be an increase of intersubjective functions in informal (and possibly also spoken) settings, and in contrast, a possible increase of textual functions in formal-written settings.

In sum, given the results of previous research on openings and closings in conversations and narratives, we expect to find age-, contact-, formality- and mode-related differences in heritage speakers' openings and closings.

## 2 Method

### 2.1 Corpus database and data elicitation

Data for this study come from the corpus collected for the Research Unit *Emerging Grammars* (RUEG; Wiese et al. 2021). An exciting and pioneering aspect of the RUEG corpus is that it comprises majority and heritage language data in four different languages (German, Greek, Russian, and Turkish) in two contact environments, namely Germany and the USA. These data are complemented with

data from monolingual speakers of all RUEG languages. In the present study, we analyzed heritage language data from adolescent and adult German, Greek, Russian and Turkish HSs as well as data from monolingual speakers of the aforementioned languages (Table 1). The age of the adolescents ranged from 13 to 19 years, and the age of the adults ranged from 20 to 37 years.

The data were collected using the Language Situations method (Wiese 2020). Participants were shown a short video clip of a minor car accident and were asked to describe it, imagining that they had witnessed the event. Importantly, they narrated the video in four different communicative situations: informal-spoken (voice message to a friend), informal-written (WhatsApp message to a friend), formal-spoken (voice message of a police witness report) and formal-written (witness report for the police). Participants in all sites and situations were given similar instructions, that is, to narrate the incident that occurred in a video. The data were collected by two elicitors at each site: one for the informal situations and the other for the formal situations. Heritage speakers provided descriptions of the video event in both their heritage and their majority languages in two separate sessions with at least a three-day interval between the two sessions, although only data from their heritage language is considered in this chapter.

This data collection approach allows us to easily compare openings and closings across different languages (German, Greek, Russian, and Turkish) and also different speaker groups for each (MSs and HSs). Thus, we can examine the extent to which MSs and HSs of different languages produce openings and closings in their narrations, as well as if they use similar (or distinct) functions in openings and closings across communicative situations.

The present study is based on the German, Greek, Russian and Turkish sub-corpora of the RUEG corpus (Lüdeling et al. 2024), which we enriched with additional annotations for openings and closings. Table 1 presents the total number of heritage speakers with US English and German as their majority language as well as the total number of monolingual speakers for each language. Note that every speaker produced four different narrations and thus the total number of texts is quadruple the total number of speakers.

Overall, we annotated and analyzed 2608 texts from the RUEG corpus (see Table 2). The annotations were extracted using the ANNIS corpus search tool (Krause & Zeldes 2014).

## **2.2 Openings and closings annotation scheme**

Openings and closings were annotated using the Partititur-Editor component of the annotation software Exmaralda (Schmidt & Wörner 2014). Annotations were

Table 1: Distribution of adult and adolescent speakers per speaker group

| Speakers    | Majority language | Adolescents | Adults | Total N |
|-------------|-------------------|-------------|--------|---------|
| German HSs  | English           | 29          | 7      | 36      |
| German MSs  | German            | 32          | 32     | 64      |
| Greek HSs   | English           | 32          | 32     | 64      |
| Greek HSs   | German            | 18          | 26     | 44      |
| Greek MSs   | Greek             | 32          | 32     | 64      |
| Russian HSs | English           | 33          | 33     | 66      |
| Russian HSs | German            | 28          | 29     | 57      |
| Russian MSs | Russian           | 34          | 33     | 67      |
| Turkish HSs | English           | 32          | 27     | 59      |
| Turkish HSs | German            | 32          | 33     | 65      |
| Turkish MSs | Turkish           | 34          | 32     | 66      |

Table 2: Distribution of analyzed texts per language

| Language of texts | Total N of texts |
|-------------------|------------------|
| German            | 400              |
| Greek             | 688              |
| Russian           | 760              |
| Turkish           | 760              |

completed in two rounds: annotating in the first round and proof checking annotations in the second round. Our study included three levels of annotation, each tagged on a different tier in Exmaralda. The first tier identified and spanned the complete area of openings and closings, while the second and third tiers included the functions inside each opening and closing – a detailed presentation of the functions can be found in Section 2.2.2 of this chapter.

Figure 1 shows the three tiers of annotation for an opening. The first tier [macro-dstr] spans the whole opening including punctuation and filled or silent pauses (such as (-)). In the second tier [macro-function1], general functions in openings and closings (subjective, intersubjective, textual) are tagged. In the third tier [macro-function], each tag has the same length as in the second tier, and comprises the specific functions of openings, such as contextual, evaluation,

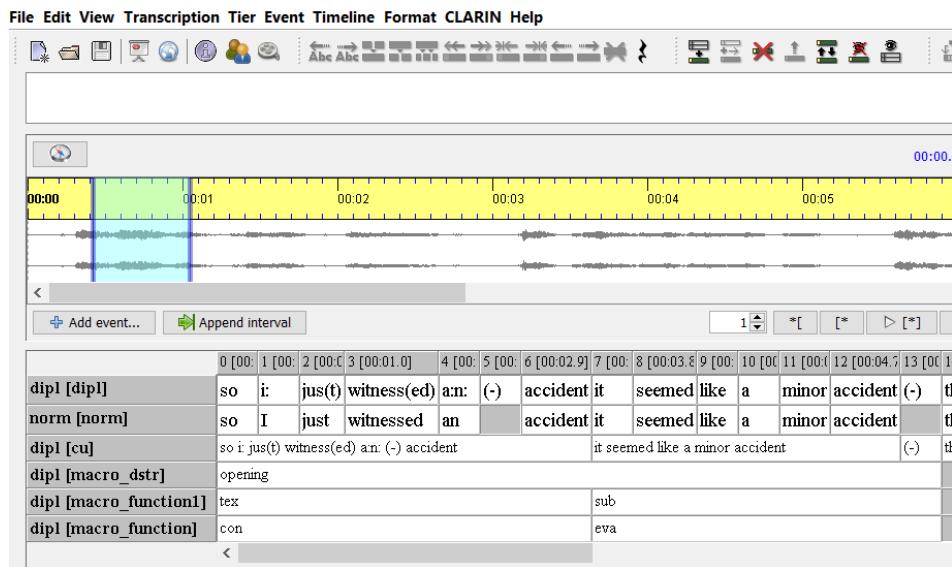


Figure 1: Screenshot example of the three layers of annotation of an opening in a spoken narration in Exmaralda

greeting, etc. In the example of an opening in Figure 1, the functions for the first phrase (*so I just witnessed an accident*) are textual and contextual, while the functions for the second part (*it seemed like a minor accident*) are subjective and evaluation.

Closing annotations are shown in Figure 2. The first tier [macro-dstr] includes the whole span of the closing, while the second and third tiers include discourse functions in closings. The example in Figure 2 has three functions all of which are textual functions. The first two parts (*no one got hurt, everything's okay*) are tagged as resolutions, while the third part (*but yeah that's what happened*) is tagged as coda.

### 2.2.1 Annotation procedure

The minimum size of an opening and/or closing is a Communication Unit (CU). As seen in the annotation layers in Figures 1 and 2, CUs were already annotated in the RUEG corpus. CUs roughly correspond to clauses, and one CU may be a main clause plus all possible embedded clauses (Topaj 2020). For long CUs containing multiple functions, we often overrode the pre-annotated CU division to assign multiple functions per CU. There were also cases of multifunctional openings that could not be divided (e.g., *I just saw a stupid accident* (contextual + evaluation

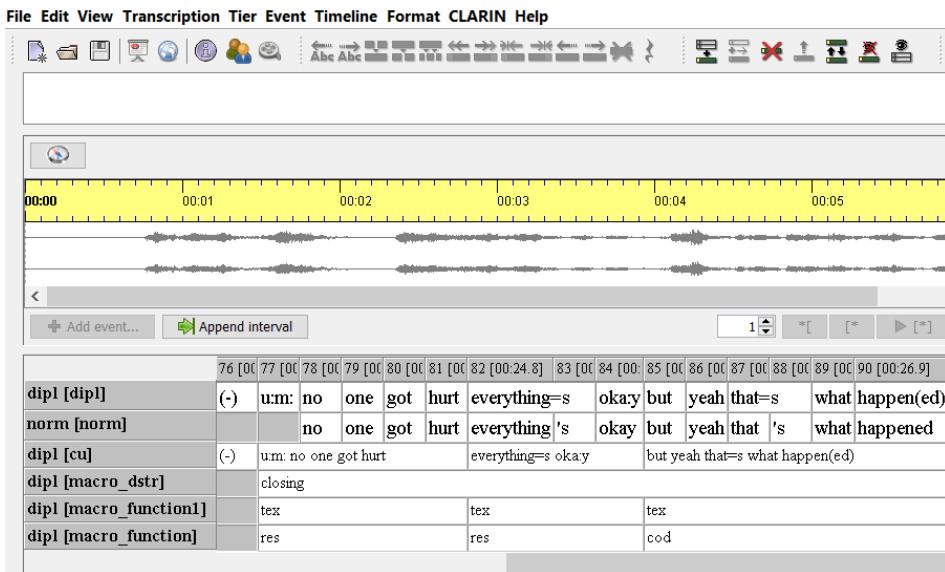


Figure 2: Screenshot example of the three layers of annotation of a closing in a spoken narration in Exmaralda

at the same time)). In such cases, we tagged as either contextual or evaluation on the basis of the amount of elements that pointed towards either one or the other function per CU. We did not insert combined tags (e.g. con-eva) so as to not end up with a very long list of tags and thus make the annotation process excessively complicated. There were at least two rounds of annotations per sub-corpus, and a sample of the annotations was checked for inter-annotator agreement. Annotation issues were discussed at weekly meetings across all languages at the same time so as to ensure similarity of annotations across sub-corpora to the highest possible extent. There were two to five annotators per sub-corpus, all of whom were native or near-native speakers of the language of the sub-corpus.

### 2.2.2 Functions in openings and closings

Analyzing the discourse functions of openings and closings allows for an understanding of how linguistic choices structure discourse, guide the reader or listener, and convey meaning. They also provide insights as to the contribution of (inter-)subjectivity in monolingual and heritage speakers' texts across communicative situations. To facilitate our analysis, we annotated the discourse functions at both a general and more specific level, as mentioned earlier. Following functional categorization of discourse markers (e.g., Brinton 1996), we identified

three main functional categories of openings and closings: intersubjective, subjective, and textual. Each of these three general categories of functions can be elaborated into further, more detailed functions.

### 2.2.2.1 Intersubjective functions

Intersubjective functions can be understood as contributions to the overall dynamics of communication and social interaction. Essentially, intersubjective functions correspond to interpersonal interactions, for which conversational analysis research has shown cross-linguistic differences (e.g., Luke & Pavlidou 2002). Within the general intersubjective functions, we annotated eleven more specific functions: greeting, summoning, initial inquiry, inquiry response, identification, justification, attention getter, common ground, reaction seeking, giving advice, and valediction. These are each elaborated below.

(a) Greeting: expression of a greeting.

- (1) hallo Max  
hello Max  
'hello Max' (*opening*, DEbi01MR\_isD<sup>1</sup>)
- (2) hallo schönen guten Tag  
hello beautiful good day  
'hello good day' (*opening*, DEbi02FT\_fsD)

(b) Summoning: calling the other person's name or using some other form of address such as *dude*, *bro*, *babe*, etc.

- (3) kanka merhaba  
dude hello  
'dude hello' (*opening*, TUmo10MT\_iwT)

---

<sup>1</sup>Explanation of participant codes: country where data was collected: "DE" for Germany, "US" for USA, "TU" for Turkey, "RU" for Russia, "GR" for Greece; "bi" for bilingual speakers, "mo" for monolingual speakers; speaker number incl. age group: 1 to 49: adults; 50 to 99: adolescents; gender: M(ale), F(emale) (there were no speakers who identified as non-binary); language of speakers: "D" for German, "E" for English, "G" for Greek, "R" for Russian, "T" for Turkish; communicative situation: "f" formal vs. "i" informal; "s" spoken vs. "w" written; language of production: "D" for German, "E" for English, "G" for Greek, "R" for Russian, "T" for Turkish. For example: "DEbi01MR\_isD" represents an adult male majority German, Russian HS speaking German in an informal-spoken communicative situation.

- (4) dicker  
dude  
'dude' (*opening*, DEbi65MG\_isD)
- (c) Initial inquiry: inquiry about how the other person is doing.
- (5) nasilsin  
how.are.you  
'how are you?' (*opening*, DEbi08FT\_isT)
- (d) Inquiry response: response to an imaginary inquiry of an imaginary interlocutor, such as "how are you?". This type of opening emerged from the communicative situations showing that participants were engaged in the imaginary scenario.
- (6) kala ki ego  
fine and I  
'I am also fine' (*opening*, DEbi59MG\_isG)
- (e) Identification: statement of identifying information, usually the participant's name (which has been replaced by their participant code for anonymization purposes). In formal-spoken and written communicative situations, participants were asked to mention that their case number was "F16". Participants sometimes used this number in an identification function, e.g. 'I am F16' or to introduce a textual function, such as 'This is case F16'.
- (7) DEbi06FT ist mein name  
DEbi06FT is my name  
'My name is DEbi06FT' (*opening*, DEbi06FT\_fsD)
- (f) Justification: mention of the reason why the participant was calling or texting.
- (8) tilefono na kano mia katathesi ja ena peristatiko pu ida  
I.call to make a statement about an incident that I.saw  
'I am calling to make a statement about an incident that I saw'  
(*opening*, GRmo78MG\_fsG)

(g) Attention getter: use of certain phrases or words to get the imaginary interlocutor's attention.

- (9) simdi ne oldu bi bildeniz  
now what happened if you.knew  
'If only you knew what happened now' (*opening*, USbi66FT\_iwT)
- (10) sana bir sey anlatmam lazim  
to.you one something to.tell I.need  
'I need to tell you something' (*opening*, DEbi69FT\_iwT)
- (11) du glaubst echt nicht was mir grad passiert ist  
you believe really not what to.me just happened  
'You really won't believe what just happened to me' (*opening*, DEbi54FR\_isD)

(h) Common ground: reference to information that the participant shares in common with the imaginary interlocutor.

- (12) du weißt was isch meine  
you know what I mean  
'You know what I mean' (*opening*, DEbi53MG\_isD)

(i) Reaction seeking: statement seeking reaction from imaginary interlocutor.

- (13) kakie nashi dal'nejshie dejstvija  
what our next actions  
'What are our next steps?' (*closing*, RUmo56FR\_fsR)

(j) Giving advice: statement offering advice to the imaginary interlocutor.

- (14) halbuki şehir içi yavaş gitsene  
however city in slow go  
'However, go slower in the city' (*closing*, TUmo32FT\_iwT)

(k) Valediction: expression of goodbye/closing with a greeting.

- (15) do svidanija  
to dates  
'goodbye' (*closing*, USbi77FR\_fsR)

### 2.2.2.2 Subjective functions

Subjective functions focus on the subjective elements of discourse, such as opinions, beliefs, emotions, evaluations, emotional markers (which we have tagged as ‘reactions’), and personal statements. Subjective functions are important in understanding how individuals express their subjective stance, shape their identities, and influence the interpretation of discourse. Within the general subjective function, we annotated three more specific functions: reaction, evaluation, and personal statement. These are each elaborated below.

- (a) Reaction: all expressions of emotion, such as happiness, anger, surprise, or sadness, as well as reactions to the event that the participants were asked to describe including expressions such as *wow*, laughter, and also emojis in written texts<sup>2</sup>.

- (16) oh mein gott  
oh my god  
'Oh my god' (*opening*, DEmo61FD\_isD)

- (b) Evaluation: Ethical, evaluative, or prescriptive statements relating to the events that are found in the situational narrative. Parts of the situational narrative may or may not be explicitly referred to inside the statements.

- (17) oder na ja so krass ist es auch nicht  
or oh yes so bad is it also not  
'or well it's not that bad either' (*closing*, DEbi03FG\_isD)

- (c) Personal statement: Personal information or personal statements not necessarily linked to the description of the incident.

- (18) ich hoffe ich muss hier nicht als Zeuge dableiben  
I hope I must here not as witness stay  
'I hope I won't have to stay here as witness' (*closing*, DEbi62MG\_isD)

---

<sup>2</sup>We tagged all emojis as “reactions” because our annotation scheme did not include a detailed analysis of discourse functions in emojis. Such an analysis, which also conveys textual and intesubjective functions of emojis, can be found in Wiese & Labrenz (2021).

### 2.2.2.3 Textual functions

Textual functions focus on the structural and cohesive aspects of texts, examining how linguistic features are used to create coherence, guide the flow of information, and shape the overall textual organization. Textual functions are crucial in understanding how discourse is constructed and how meaning is conveyed through linguistic choices. Within the general textual function, we annotated six more specific functions: initialize narrative, contextual, orientation, encapsulation, resolution, and coda. These are each elaborated below.

- (a) Initialize narrative: Phrase or statement that indicates the beginning of the situational narration.

- (19) yasanan kazanin gidisati su yönedor  
that.happened of.accident its.course this in.direction  
'the course of the accident is as follows' (*opening*, TUmo80FT\_fwT)
- (20) und es ereignete sich folgendermaßen:  
and it happened REFL as.follows:  
'and it happened as follows' (*opening*, DEbi07FG\_fsD)

- (b) Contextual: Explicit reference to the events depicted in the video as a starting point for introducing the main body of the narrative (events in the video/accident). The contextual function differs from the initialize narrative function in that the core narration starts immediately after the initialize narrative function, but this is not necessarily the case in contextual parts of the opening.

- (21) demin bi kaza oldu burada  
just.now one accident happened here  
'there was just an accident here' (*opening*, TUmo23MT\_iwt)
- (22) Zeugenaussage zum Fall F16  
witness.report for.the case F16  
'witness report for the case F16' (*opening*, DEmo85FD\_fwD)

- (c) Orientation: Expression of a particular event or circumstance located specifically in terms of time and space, thematically related to the video.

- (23) park yerinde duruyordum.  
in.the.parking place I.was.standing  
'I was standing in the parking lot' (*opening*, DEbi51MT\_fwt)

(d) Encapsulation: Statement that summarizes the events that have been previously described in the main narration including explicit mention of the main event, i.e., the accident.

- (24) und so kam es zum unfall  
and so came it to.the accident  
'and this is how the accident happened' (*closing*, DEbi15MG\_fsD)

- (25) ve kaza öyle oluştu  
and accident in.such.a.way happened  
'and that's how the accident happened' (*closing*, TUMo16FT\_fwT)

(e) Resolution: Resolution of the plot developed in the narrative.

- (26) ja aber niemandem ist was passiert alles gut  
yeah but nobody is something happened everything good  
'but yeah nothing happened to anybody, all good' (*closing*, DEbi06FG\_isD)

(f) Coda: Conclusion of the text with a formulaic or non-formulaic expression, relating the chain of events to the state of affairs at the time of telling the story. Codas differ from encapsulation instances in that the encapsulation instances include specific mention of the "accident" that the participants saw in the video.

- (27) ich bedanke mich  
I thank me  
'thank you' (*closing*, DEbi24FR\_fsD)

- (28) jeto vsjo chto sluchilos  
that's all that happened  
'that's all (that happened)' (*closing*, USbi67MR\_isR)

### 3 Results

#### 3.1 Levels of analysis

Our analysis includes the following levels of analysis:

- (a) Distribution of openings and closings across speaker groups.

- (b) Distribution of discourse functions inside openings and closings per communicative situation.
- (c) Distribution of openings and closings per age group (adolescents vs. adults).

### 3.2 Distribution of openings and closings across speaker groups

We first analyzed the distribution of openings and closings that monolingual and heritage speakers used in their narratives across the four different communicative situations in the RUEG corpus. To this end, we calculated the percentage of texts including openings and closings across the total number of texts. The results are presented per speaker group.

#### 3.2.1 Openings and closings in German texts

German MSs and HSs had the same pattern of distribution of openings and closings in their texts. For both groups, the overwhelming majority of texts included an opening, and more than half the texts included a closing (Figure 3). We conducted a Kruskal-Wallis Test in R (RStudioTeam 2020) to examine possible differences in opening and closing distribution in the two groups of speakers. No significant differences were found between MSs and HSs ( $\chi^2 = 0.33$ ,  $p = 0.56$ ,  $df = 1$ ).

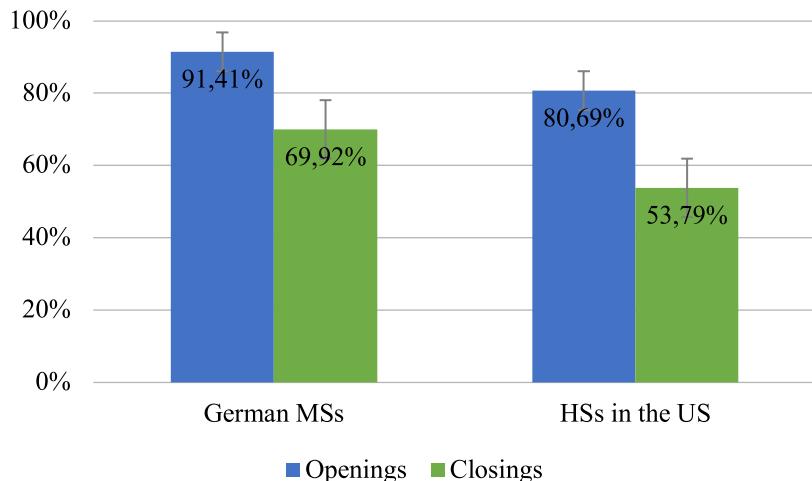


Figure 3: Distribution of openings and closings per speaker group in German narrations

### 3.2.2 Openings and closings in Greek texts

The Greek texts in the RUEG corpus were produced by three groups of Greek-speaking participants: monolingually-raised Greek speakers in Greece (Greek MSs), Greek HSs in Germany and Greek HSs in the US. As can be seen in Figure 4, all groups had a similar pattern of opening and closing distribution. This was also reflected in the statistical analysis checking for possible differences in the distribution of openings and closings, which revealed no significant differences between MSs and HSs ( $\chi^2 = 0.50, p = 0.77, \text{df} = 2$ ). All three groups produced significantly more openings than closings ( $\chi^2 = 10.38, p = 0.001, \text{df} = 3$ ).

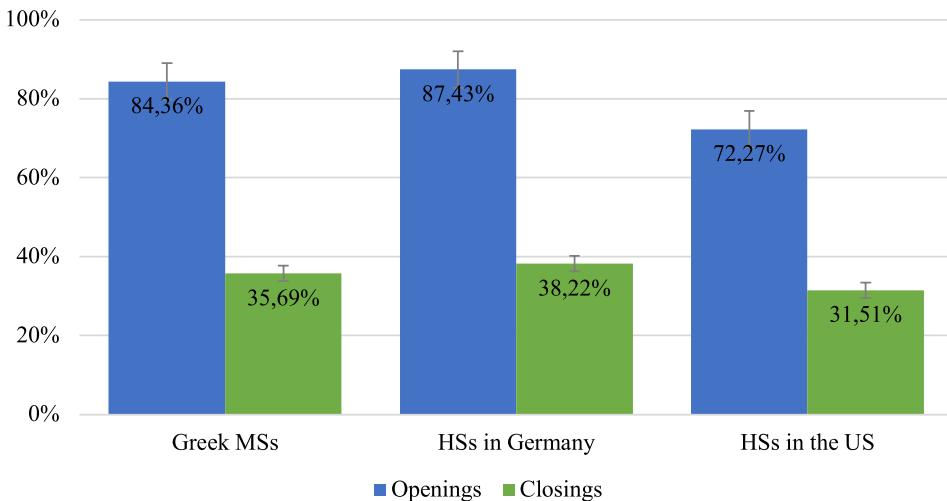


Figure 4: Distribution of openings and closings per speaker group in Greek narrations.

### 3.2.3 Openings and closings in Russian texts

We also analyzed the distribution of openings and closings in the texts of Russian MSs, HSs in Germany, and HSs in the US (Figure 5). We compared frequency distribution and the analysis did not reveal any statistical difference across the three Russian-speaking groups ( $\chi^2 = 0.26, p = 0.87, \text{df} = 2$ ). All three Russian-speaking groups produced significantly more openings than closings ( $\chi^2 = 9.66, p = 0.02, \text{df} = 3$ ).

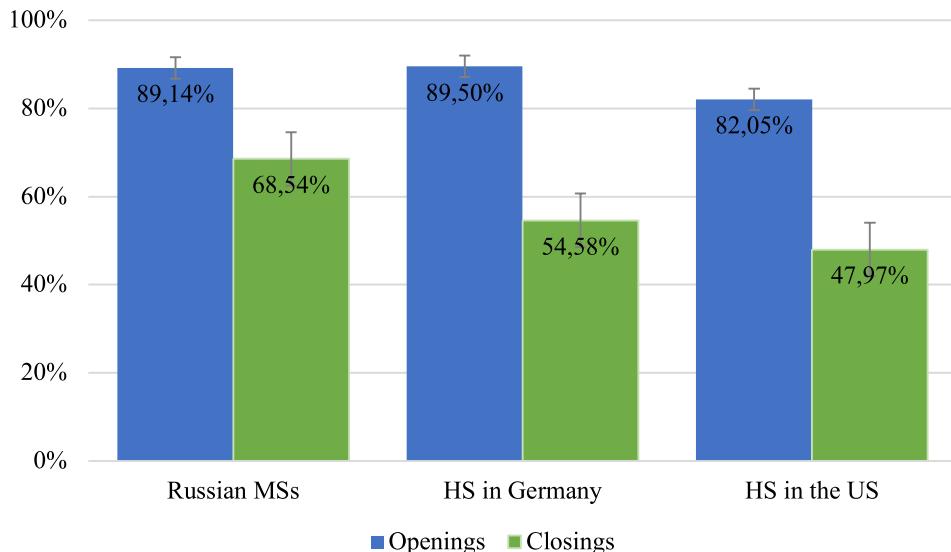


Figure 5: Distribution of openings and closings per speaker group in Russian narrations

### 3.2.4 Openings and closings in Turkish texts

In addition, we statistically compared the distribution of openings and closings in Turkish MSs, HSs in Germany and HSs in the US, and found no statistically significant differences in the distribution of openings and closings in the Turkish texts ( $\chi^2 = 0.03, p = 0.98, \text{df} = 2$ ). As can also be seen in Figure 6, all three Turkish-speaking groups produced significantly more openings than closings ( $\chi^2 = 9.66, p = 0.02, \text{df} = 3$ ).

### 3.2.5 Summary

In sum, the analysis of openings and closings across MSs and HSs in the RUEG corpus texts reveals that HSs of German, Greek, Turkish and Russian did not differ significantly in the distribution of openings and closings from the MSs of the respective languages. Speakers produced far more openings than closings cross-linguistically. We believe that this pattern of results may relate to the fact that participants were required to produce narrations in four different communicative situations, and may have thus used their openings to establish each of the four different communicative situations. Overall, the percentages of openings are more or less similar cross-linguistically, but the percentages of closings seem

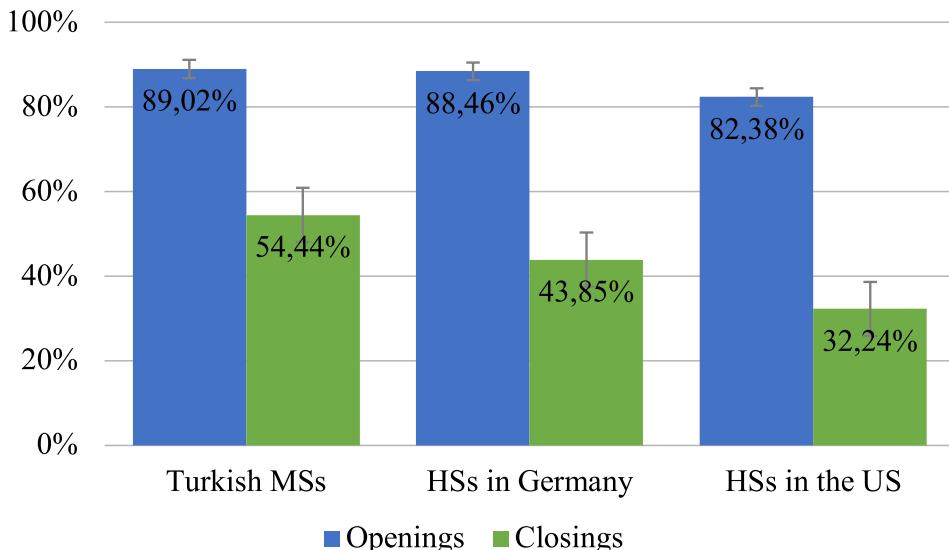


Figure 6: Distribution of openings and closings per speaker group in Turkish narrations

to be lower in HSs compared to MSs in German speakers. In addition, the distribution of closings follows the pattern Russian/Turkish MSs > HSs in Germany > HSs in the US in the Russian and Turkish-speaking groups. We discuss this point further in Section 4.

### 3.3 Distribution of openings and closings across age groups

The RUEG corpus includes data from adolescents and adults. Based on previous studies in openings and closings including different age groups (Tolchinsky et al. 2002, Dollnick & Pfaff 2013), we analyzed the different age groups per language in order to examine possible differences in the distribution of openings and closings across adolescent and adult HSs and MSs for each language.

We statistically compared adolescents and adults' openings and closings per speaker group using the Kruskal-Wallis Test. None of the analyses revealed statistically significant results between adolescents and adults neither in the distribution of openings nor in the distribution of closings. There were small differences across groups, which however did not reach statistical significance. For example, adult German MSs produced more openings than adolescents whereas adolescent HSs produced more openings than adult HSs. We can also see in Table 3 that the distributional difference between MSs and HSs in Russian and Turkish

Table 3: Distribution of openings and closings per speaker and age group in each majority and heritage language

| Speakers            | Openings |             | Closings |             |
|---------------------|----------|-------------|----------|-------------|
|                     | Adults   | Adolescents | Adults   | Adolescents |
| German HSs          | 76%      | 82%         | 55%      | 53%         |
| German MSs          | 97%      | 86%         | 70%      | 70%         |
| Greek HSs_US        | 69%      | 75%         | 34%      | 29%         |
| Greek HSs_Germany   | 86%      | 85%         | 39%      | 37%         |
| Greek MSs           | 81%      | 88%         | 39%      | 37%         |
| Russian HSs_US      | 79%      | 85%         | 55%      | 42%         |
| Russian HSs_Germany | 87%      | 92%         | 64%      | 43%         |
| Russian MSs         | 86%      | 92%         | 66%      | 71%         |
| Turkish HSs_US      | 83%      | 82%         | 41%      | 25%         |
| Turkish HSs_Germany | 92%      | 85%         | 55%      | 32%         |
| Turkish MSs         | 91%      | 88%         | 65%      | 44%         |

closings was actually because of the very low percentages of closings in adolescents. We thus see again that openings were more similarly distributed across age and language groups, whereas there was more variability in closings, which were, however, less frequent than openings.

### 3.4 Openings: distribution of discourse functions across communicative situations

This section includes the analysis of discourse functions across communicative situations in openings. We conducted this analysis in order to investigate the type of discourse material that MSs and HSs chose to include in their openings, and the extent to which the functions in openings followed the expected norms for each communicative situation. We present the analysis per language group, starting with MSs and HSs of German.

#### 3.4.1 Opening discourse functions in German MSs and HSs

We analyzed the types of functions (intersubjective, subjective, textual) in German narrations across speaker groups and communicative situations (Figure 7). The Kruskal-Wallis test did not yield any significant differences across speakers

( $\chi^2 = 1.19$ , df = 1,  $p = 0.27$ ), which means that the two groups of speakers showed a similar pattern.

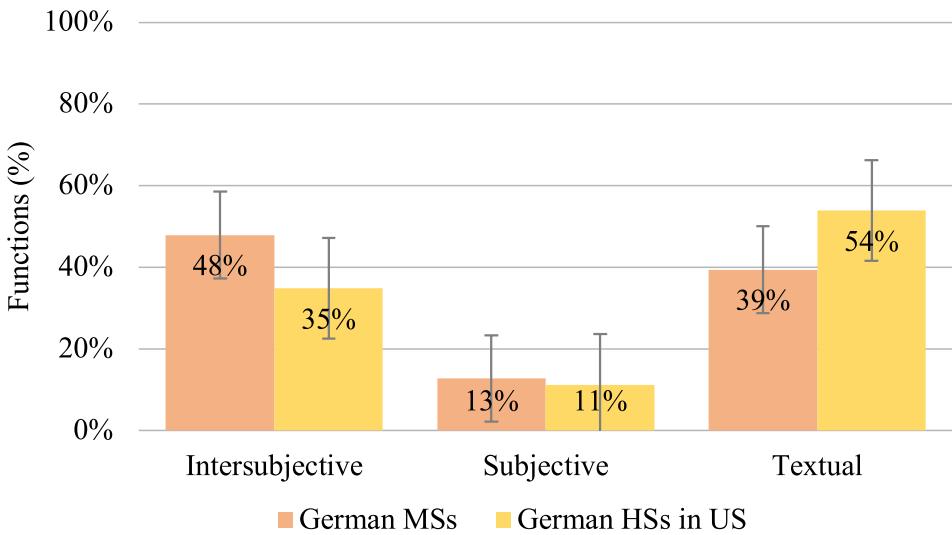


Figure 7: Distribution of functions per speaker group in German openings

As we can see in Figure 7, both groups of speakers produced more intersubjective and textual functions than subjective functions in their openings. Thus, speakers introduced their texts with intersubjective functions such as greetings, justifications and identifications as well as textual functions such as initializing of narrative, contextual and episodic functions and not so often with subjective functions such as evaluations and personal statements. In total, German MSs produced 361 intersubjective, 96 subjective, and 297 textual functions. German HSs produced 99 intersubjective, 32 subjective, and 153 textual functions. Therefore, intersubjective were the most frequent functions in MSs whereas textual were the most frequent functions HSs. The distribution of functions per communicative situation in the two speaker groups can be seen in Table 4.

As already mentioned, intersubjective functions were the most frequent functions in MSs and second most frequent in HSs. A closer look at intersubjective functions showed that MSs and HSs mostly used greetings, identifications and justifications in formal-spoken communicative situations, as illustrated in examples 29-31.

Table 4: Percentages of functions in openings in each communicative situation and speaker group in German openings

| Situation        | Function        | German MSs | HSs_US |
|------------------|-----------------|------------|--------|
| Informal-spoken  | Intersubjective | 44%        | 33%    |
|                  | Subjective      | 22%        | 20%    |
|                  | Textual         | 33%        | 48%    |
| Informal-written | Intersubjective | 45%        | 38%    |
|                  | Subjective      | 17%        | 8%     |
|                  | Textual         | 38%        | 54%    |
| Formal-spoken    | Intersubjective | 62%        | 41%    |
|                  | Subjective      | 7%         | 7%     |
|                  | Textual         | 31%        | 52%    |
| Formal-written   | Intersubjective | 28%        | 14%    |
|                  | Subjective      | 4%         | 7%     |
|                  | Textual         | 68%        | 79%    |

## (29) Greeting

ähm ja schönen guten tag  
 um jeah beautiful good day  
 ‘um yeah good day’ (DEmo21MD\_fsD)

## (30) Identification

mein name ist DEMo19FD  
 my name is DEMo19FD  
 ‘my name is’ (DEmo19FD\_fsD)

## (31) Justification

ich wollte ein tatvorfall berichten den ich eben beobachtet habe  
 I wanted an incident report which I just watched have  
 ‘I wanted to report an incident which I just watched’ (DEmo79MD\_fsD)

Greetings were the standard way for MSs to start an opening in formal-spoken communicative situations (out of the 63 total openings, 54 start with a greeting, i.e. 86%). Half of these greetings (27 greetings) were of the more “formal” *Guten Tag* (‘good day’) type as in examples (2) and (29), almost half of the greetings (25 greetings) involved the neutral greeting *Hello* (‘hello’), and two greetings were a combination of *Hello* and *Guten Tag* (see also Wiese et al. 2025 [this volume])

for an analysis of discourse organization and lexical choices in openings with a focus on register distinctions). Interestingly, in HSs, only half of the openings included a greeting (16 greetings out of 32 openings), and most of these greetings were more informal (13 out of 16 greetings, 81%). Therefore, HSs did not follow the prevalent monolingual strategy of starting a formal-spoken opening with a greeting. And even when HSs started an opening with a greeting, they consisted of the more neutral greeting “hallo”. This result should be treated with caution, however, as the HS German data only include 7 adults and 29 adolescents, and the lack of formal greeting may be a characteristic of age rather than speaker group. When checking the adolescent MS greetings, we saw that half of them (13/26) were informal or neutral, and half of them formal. Adult MSs used a slightly higher number of formal greetings (16/28, 57%) but neither adolescent nor adult MSs reached 81% of informal/neutral greetings found in HSs. This leads us to the conclusion that HSs’ preferred discourse strategy is to use informal/neutral greetings in both informal and formal communicative situations.

Turning to identifications in formal-spoken settings, MSs used identification sequences such as the one in example (30), that is, they provided their name (replaced with their code number for anonymity reasons as stated previously) most of the time (46/53, 87%), and sometimes instead of their name, they provided the case number F16, which was given to them as part of the experimental instructions. HSs provided only 8 identifications, and out of these eight, five referred to the case number and not their name. We thus see here that HSs were reluctant to provide their own name, and they were more likely to use the case number instead. Finally, justifications (31) did not differ much across the two speaker groups except for the fact that HSs provided justifications to a much lesser extent than MSs.

In terms of textual functions in German openings, these were the most frequent in HSs and second most frequent in MSs. In both groups, the two most common textual functions across all communicative situations were contextual (specifically mentioning the incident) and episodic (mentioning time and/or place of incident) functions, as in examples (32) and (33).

(32) Contextual

- Gerade habe ich einen autounfall draussen vor die wohnung gesehen  
Just have I a car.accident outside of the building saw  
'I just saw a car accident outside of the building' (USbi01FD\_iwD)

## (33) Episodic

Um Mittag am 7.5.19 war am Parkplatz hinter dem Wohngebäude an at noon on 7.5.19 was at the parking lot behind the building on der Unteren Hauptbahnstrasse ein Autounfall  
the Unteren Hauptbahnstrasse an accident

'At noon on 7.5.19 at the parking lot behind the residential building on Unteren Hapbahnstrasse there was an accident' (USBi63MD\_fwD)

Although HSs overall produced fewer textual functions than MSs, the type and the distribution of textual functions across communicative situations patterned with MSs (see Table 4).

### 3.4.2 Opening discourse functions in Greek MSs and HSs

Greek MSs and HSs had a similar pattern of intersubjective, subjective and textual functions in their openings (Figure 8). Consequently, the statistical analysis showed no significant difference across speaker groups ( $\chi^2 = 1.06$ ,  $df = 2$ ,  $p = 0.58$ ) but a statistically significant effect in terms of differences across functions ( $\chi^2 = 6.48$ ,  $df = 2$ ,  $p = 0.03$ ). This difference stems from the fact that subjective functions were used very sparingly across all speaker groups.

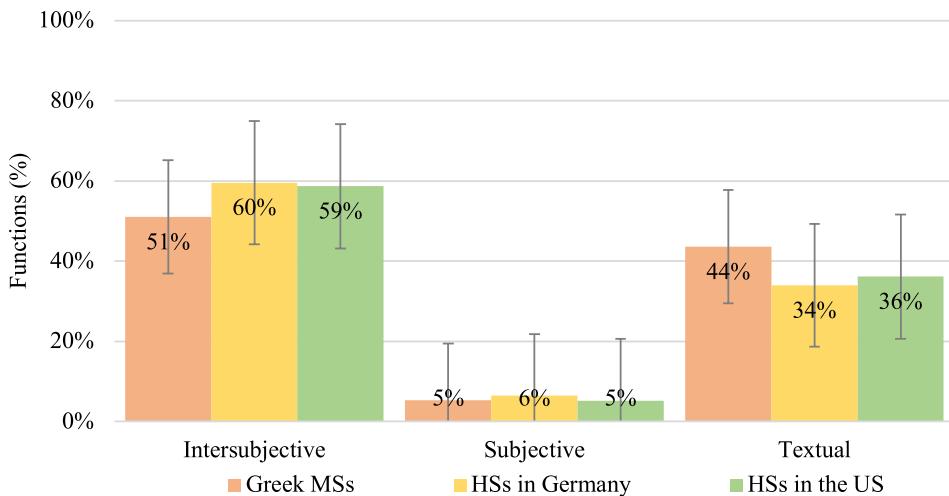


Figure 8: Distribution of functions per speaker group in Greek openings

The distribution of subjective, intersubjective and textual functions in Greek MSs and HSs per communicative situation is presented in Table 5.

Table 5: Percentages of functions in openings in each communicative situation and speaker group in Greek openings

| Situation        | Function        | Greek MSs | HSs_Germany | HSs_US |
|------------------|-----------------|-----------|-------------|--------|
| Informal-spoken  | Intersubjective | 57%       | 64%         | 57%    |
|                  | Subjective      | 10%       | 8%          | 11%    |
|                  | Textual         | 33%       | 28%         | 32%    |
| Informal-written | Intersubjective | 61%       | 60%         | 60%    |
|                  | Subjective      | 6%        | 9%          | 6%     |
|                  | Textual         | 34%       | 32%         | 34%    |
| Formal-spoken    | Intersubjective | 44%       | 61%         | 61%    |
|                  | Subjective      | 2%        | 3%          | 2%     |
|                  | Textual         | 54%       | 36%         | 37%    |
| Formal-written   | Intersubjective | 39%       | 50%         | 55%    |
|                  | Subjective      | 3%        | 6%          | 0%     |
|                  | Textual         | 59%       | 44%         | 45%    |

As we can see in Table 5, the distribution of intersubjective functions across communicative situations in informal communicative situations is similar across speaker groups. In formal communicative situations, MSs produced fewer intersubjective functions than the two HS groups. This means that Greek HSs in Germany and the US inserted more conversational elements in their formal openings than Greek MSs. In addition, HSs in Germany produced a relatively large number of identifications in formal-spoken situations (19/78: 24%) in comparison to MSs (1/56: 2%) and HSs in the US (8/58: 13%). All of these cases involved “my name is...” instances as in example (34):

(34) Identification

legome DEbi17FG; onomazome Debi26MG; t=onoma mu ine  
 I.am.named DEbi17FG; I.am.named DEbi26MG; the.name.of.me is  
 DEbi14MG  
 DEbi14MG  
 ‘my name is DEbi17FG\_P’ (DEbi17FG\_fsG); ‘DEbi26MG’  
 (DEbi26MG\_fsG); ‘DEbi14MG’(DEbi14MG\_fsG)

(35) Identification

ime o USbi17MG

I.am the USbi17MG

'I am USBi17MG' (USbi17MG\_fsG)

All instances of identification in Greek HSs in Germany look like example 34 and constitute formal ways to introduce oneself in Greek. In contrast, all of the 8 instances used by Greek HSs in the US look like example (35) and are primarily informal ways of introducing oneself in Greek. These results lead us to two conclusions; firstly, identifying oneself when making a phone call to the police to leave a voice message is not a discourse strategy that Greek MSs use. Secondly, the two HS groups differ in that the HSs in Germany used the "appropriate" or expected way to introduce themselves, whereas HSs in the US used an informal way to introduce themselves in a formal setting. We thus see that both groups of HSs used a discourse strategy that is not common across Greek MSs, but the two HS groups differ in that only the HSs in Germany identified themselves in a formal way in a formal environment. The second noticeable result in the analysis of intersubjective functions in Greek was the relatively low number of intersubjective functions in informal-written texts in HSs in the US which was due to the fact that only one HS in the US used a justification sequence, whereas for Greek MSs justifications (such as (36)) were the primary way to introduce informal-written settings (24/65: 37%).

(36) Justification

Tha kathisteriso na ertho, giati opos erhomun egine ena  
 Will delay to come, because as I.was.coming happened an  
 atihima!  
 accident

'I will come later because there was an accident as I was  
 coming'(GRmo02FG\_iwG)

In informal-written texts, the distribution of discourse functions in Greek MSs and HSs was the following:

- Greek MSs: justifications > attention getters > greetings
- Greek HSs in Germany: greetings > justifications > attention getters
- Greek HSs in the US: greetings > attention getters

Although HSs primarily used greetings in their instant messages, a greeting was not the preferred strategy to begin a text message in Greek MSs. The HS groups were similar in that both groups placed priority on greetings, and also both of them used at least one discourse strategy that was also found in MSs. Textual and subjective functions had a similar distribution across groups, and therefore we do not discuss them further.

### 3.4.3 Opening discourse functions in Russian MSs and HSs

The data analysis of Russian MSs and HSs revealed a similar distribution of discourse functions across the three speaker groups (Figure 9). There was no statistical significance across groups ( $\chi^2 = 0.62$ , df = 2,  $p = 0.73$ ) but there was a statistically significant effect of functions ( $\chi^2 = 6.48$ , df = 2,  $p = 0.03$ ), which reflected the relatively high distribution of intersubjective and textual functions compared to the low distribution of subjective functions.

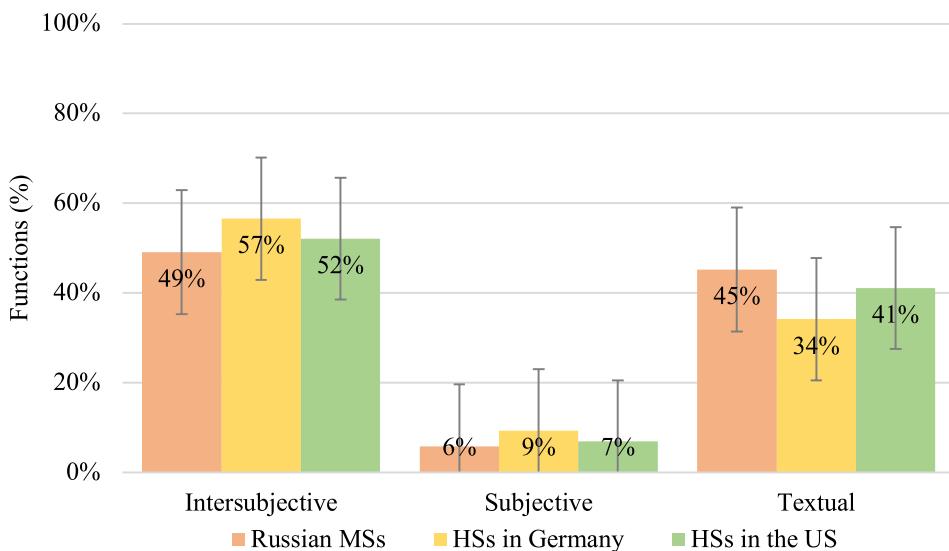


Figure 9: Distribution of functions per speaker group in Russian openings

The most frequent type of function in Russian MSs and HSs openings was intersubjective. Intersubjective functions were most frequent in Russian HSs in Germany who produced a total of 401 intersubjective functions, followed by MSs and HSs in the US who produced almost a similar amount of intersubjective functions (320 and 300 respectively). Although the distribution discourse functions

per communicative situations did not differ across the three speaker groups as can be seen in Table 6, Russian HSs in Germany produced a lot more intersubjective functions in informal than in formal settings (HSs in Germany: 212, HSs in US: 160, MSs: 154 intersubjective functions in informal settings).

Table 6: Percentages of functions in openings in each communicative situation and speaker group in Russian openings

| Situation        | Function        | Russian<br>MSs | HSs_-<br>Germany | HSs_US |
|------------------|-----------------|----------------|------------------|--------|
| Informal-spoken  | Intersubjective | 46%            | 55%              | 52%    |
|                  | Subjective      | 9%             | 16%              | 9%     |
|                  | Textual         | 45%            | 30%              | 39%    |
| Informal-written | Intersubjective | 45%            | 61%              | 46%    |
|                  | Subjective      | 7%             | 6%               | 10%    |
|                  | Textual         | 48%            | 33%              | 45%    |
| Formal-spoken    | Intersubjective | 60%            | 63%              | 59%    |
|                  | Subjective      | 2%             | 8%               | 5%     |
|                  | Textual         | 38%            | 29%              | 37%    |
| Formal-written   | Intersubjective | 35%            | 45%              | 43%    |
|                  | Subjective      | 3%             | 2%               | 2%     |
|                  | Textual         | 62%            | 53%              | 55%    |

A closer examination of intersubjective functions revealed that HSs in Germany used more justifications (see example (37)) than Russian MSs across all communicative situations, and they also had a higher variety of functions in informal communicative situations. Russian MSs used only one justification in the informal-spoken situation, 8 (24%) in formal-written, and 42 (32%) justifications in formal-spoken. In contrast, HSs in Germany used 13 instances of justifications (6%) in informal settings, 11 (28%) in formal-written and 55 (37%) in formal-spoken situations.

(37) Justification

- i ja zvonju potomu chto segodnya ja uvidela  
 and I am calling because of which today I saw  
 'and I am calling because of what I saw today' (DEbi03FR\_fsR)

Another characteristic of Russian openings was that the two HS groups produced more summonings such as (38)<sup>3</sup>, and they also used a wider variety of intersubjective functions in informal communicative situations (such as common ground (39) and reaction seeking (40)) than MSs.

(38) Summoning

DEBi26FR privat

DEBi26FR hello

‘DEBi26FR hello’ (DEBi26FR\_isR)

(39) Common ground

ty znaesh’ zdes’ v centre gorode da

you know here in center downtown jeah

‘You know, here in the center of the town, yeah’ (DEBi64MR\_isR)

(40) Reaction seeking

pozhalujsta priezzhajte

please come.over

‘Please come over’ (USBi12MR\_fsR)

We also checked Russian HSs in the US to see if we would find the same pattern in identifications as in Greek HSs in the US, that is, using informal means to identify themselves in formal-spoken communicative situations. Our search did not show a similar pattern in the Russian data. Russian HSs in the US used similar ways of identifying themselves in formal-spoken communicative situations as MSs (41).

(41) Identification

menja zovut USBi20MR

my name USBi20MR

‘my name is USBi20MR’ (USBi20MR\_fsR)

In terms of subjective functions, Russian HSs in Germany produced more subjective functions – such as personal statements (43) and evaluations (42) – than MSs and HSs in the US, especially in spoken communicative situations: HSs in Germany: 53, HSs in US: 25, MSs: 22. This pattern may reflect a majority (German) language pattern.

---

<sup>3</sup>Please note that summonings differ from greetings in that in summonings the name occurs first and then the greeting, whereas in greetings the greeting precedes the name.

## (42) Evaluation

i vygljadelo jeto vneshe chisto obiektivno kak chistyj sluchaj  
 and looked it outwardly purely objectively like pure accident  
 'and it looked outwardly purely objectively like a pure accident'  
 (DEbi01MR\_isR)

## (43) Personal statement

ja zhe hotela v kino pojti  
 I also wanted to cinema go  
 'I also wanted to go to the cinema' (DEbi63FR\_isR)

Regarding textual functions (Table 6), all three speaker groups showed the same pattern in that they all produced the highest percentages of subjective functions in formal-written situations. At the same time, the three groups differed slightly in that MSs produced a lot more textual functions (62%) than intersubjective functions (35%) in formal-written situations, whereas this difference was less pronounced in HSs in Germany and the US.

### 3.4.4 Opening discourse functions in Turkish MSs and HSs

We also examined intersubjective, subjective and textual functions in the openings of Turkish speakers (Figure 10). The statistical analyses revealed no effect of group ( $\chi^2 = 2.22$ , df = 2,  $p = 0.32$ ) and a marginally significant effect of functions ( $\chi^2 = 5.42$ , df = 2,  $p = 0.06$ ).

The distribution of intersubjective, subjective and textual functions in the Turkish texts per communicative situation and speaker group can be seen in Table 7.

Overall, Turkish speakers produced high numbers of intersubjective and textual functions in their openings. HSs in Germany produced the highest number of intersubjective functions, followed by MSs, and HSs in the US. Further analysis showed that the two HS groups produced more intersubjective functions in formal than in informal communicative situations. For example, HSs used a very high number of identifications in formal communicative situations. All identification instances, however, included the case number F16, which was the number that participants were explicitly asked to provide (see example 44). We thus did not explore this result further.

## (44) Identification

ben ef on altı  
 I am F sixteen  
 'I am F16' (USbi68FT\_fsT)

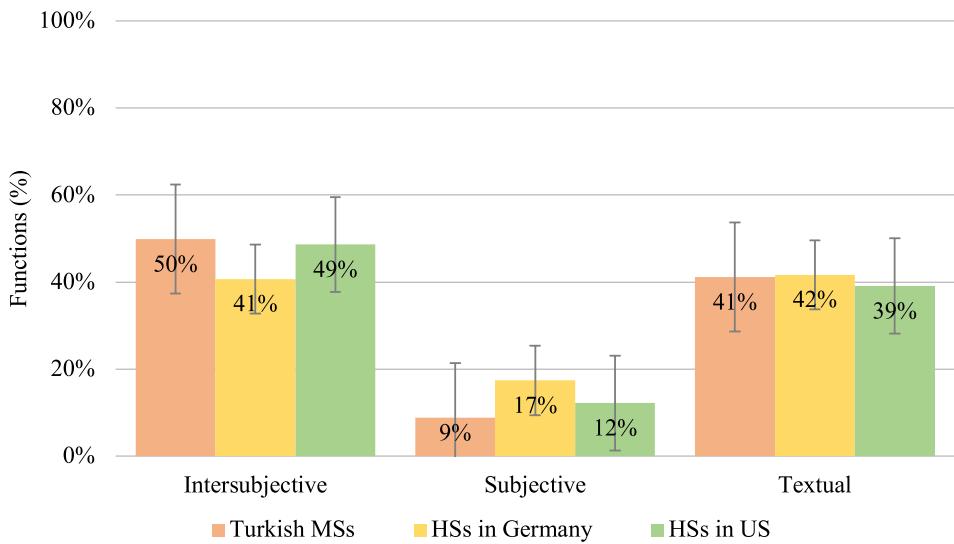


Figure 10: Distribution of functions per speaker group in Turkish openings

Table 7: Percentages of functions in openings in each communicative situation and speaker group in Turkish openings

| Situation        | Function        | Turkish MSS | HSs - Germany | HSs_US |
|------------------|-----------------|-------------|---------------|--------|
| Informal-spoken  | Intersubjective | 48%         | 46%           | 40%    |
|                  | Subjective      | 14%         | 15%           | 27%    |
|                  | Textual         | 38%         | 39%           | 33%    |
| Informal-written | Intersubjective | 44%         | 44%           | 37%    |
|                  | Subjective      | 12%         | 19%           | 22%    |
|                  | Textual         | 44%         | 36%           | 41%    |
| Formal-spoken    | Intersubjective | 58%         | 51%           | 57%    |
|                  | Subjective      | 4%          | 5%            | 4%     |
|                  | Textual         | 38%         | 44%           | 40%    |
| Formal-written   | Intersubjective | 48%         | 51%           | 56%    |
|                  | Subjective      | 4%          | 2%            | 2%     |
|                  | Textual         | 48%         | 47%           | 42%    |

In addition, our analysis showed that all three groups mostly used attention getters (45), summonings (46), greetings (47) and initial inquiries (48) in informal communicative situations. In contrast, they primarily used greetings, identifications and justifications in formal situations. The distribution of these functions across spoken and written situations was similar for all groups.

(45) Attention getters

- a. bugün n=oldu biliyo musunuz  
today what.happened you.know  
'Do you know what happened today' (TUMo23FT\_isT)
- b. sana çok önemli bi şey anlatmam gerek  
to.you very important something I.tell need  
'I have something very important to tell you' (DEbi11FT\_isT)
- c. bugün noldu inanamicaksın  
today what.happened you.won't.believe  
'You won't believe what happened today' (USbi15FT\_isT)

(46) Summonings

- a. kanka  
dude  
'Dude' (TUMo60MT\_isT)
- b. hayatım  
honey  
'Honey' (DEbi04MT\_isT)
- c. kızla:r  
girls  
'Girls' (USbi66FT\_isT)

(47) Greetings

- a. ii günler  
good day  
'Good day' (TUMo24FT\_fsT)
- b. kolay gelsin memur bey  
easy come.in officer  
'Good day officer' or 'Let your work be easy officer'  
(TUMo25MT\_fsT)

- c. merhabalar  
hello.there  
'Hello there' (TUMo61FT\_fsT)
- (48) Initial inquiry  
canım n=apiyon nasılsın  
dear how.doing.you how.are.you  
'Dear, how are you doing, how are you?' (DEbi52FT\_isT)

The fact that we found no attention getters and initial inquiries and only a very small number of summonings in formal communicative situations indicates that these discourse functions are characteristic of informal situations in Turkish. Similarly, there were almost no justifications and a much lower number of greetings (informal: 52, formal: 128 in total) in informal communicative situations, and this implies that justifications (49) and greetings (47) are characteristics of formal communicative situations for Turkish speakers.

- (49) Justification  
ben bi kazayı bildirmek istiyordum da  
I an accident report have.been.wanting well  
'I wanted to report an accident' (TUMo60MT\_fsT)

Regarding subjective and textual functions, the distribution of functions was overall similar across groups and we did not analyze them further.

### **3.4.5 Summary**

In sum, our analysis revealed both similarities and differences in the distribution of discourse functions in openings across speaker groups and languages. In general, HSs and MSs had a similar distribution of discourse functions in openings. In terms of differences, the fact that German and Greek HSs in the US did not use formal greetings in formal situations may be interpreted as evidence for register leveling in HSs. In addition, the relatively high percentages of subjective functions in Russian HSs in Germany may reflect majority language influence as German MSs also produced a relatively high percentage of subjective functions in informal situations (see Table 4).

## **3.5 Closings: distribution of discourse functions across communicative situations**

All speaker groups produced fewer closings than openings (see Section 3.2). The current section includes a basic summary of the main results without statistical

analyses because of the low number of closing instances. As in openings, we present the main results of our analysis starting with the German speakers, then the Greek, Russian and Turkish speakers. Overall, the general pattern across all speaker groups was a high number of subjective functions in informal compared to formal narrations.

### 3.5.1 Closing discourse functions in German MSs and HSs

As can be seen in Figure 11, German MSs produced primarily subjective functions, such as evaluations and personal statements, in their closings. In contrast, German HSs in the US produced textual functions, such as codas and resolutions, more frequently than subjective and intersubjective functions. This difference indicates that HSs were more reluctant than MSs to provide their personal opinions at the end of their narrations.

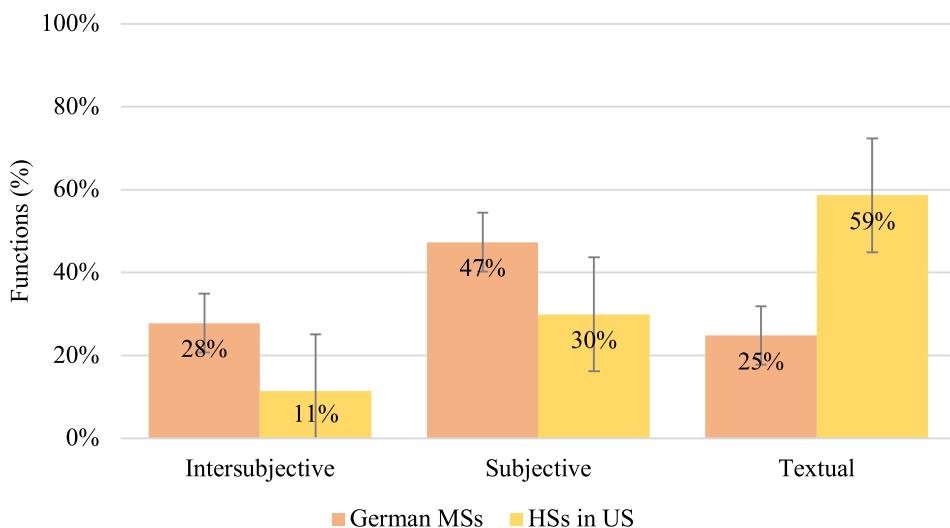


Figure 11: Distribution of functions per speaker group in German closings

We further analyzed the distribution of functions per communicative situation. In informal situations, German MSs produced 203, and HSs produced 46 subjective functions (MSs: 203/322: 63%, HSs: 46/106: 43%). This shows that German MSs' main strategy to end their spoken and written informal narrations was by an evaluation or personal comment. HSs also did so but to a lesser extent. In formal situations, MSs produced a relatively large number of valedictions (45/72: 62%) as in example (50), but this was not the case for HSs who actually produced

no valedictions at all. HSs mostly used textual functions to end their formal texts, such as codas and resolutions (51). In fact, HSs produced more resolutions (65%) than MSs (46%) whereas MSs produced more codas than HSs (MSs: 48%, HSs: 25%) across all communicative situations. Therefore, although MSs mostly used codas to close their narrations, HSs primarily used resolutions.

(50) Valedictions

- a. auf wiederhören  
to listen.again  
'Goodbye' (DEmo70MD\_fsD)
- b. einen wunderschönen tag noch  
a wonderful day still  
'Have a wonderful day' (DEmo12MD\_fsD)

(51) Resolution

- es ist aber keiner verletzt worden  
it is but nobody hurt gotten  
'But nobody got hurt' (USbi01FD\_fsD)

Overall, German MSs and HSs mostly produced evaluations, personal statements and reactions in informal situations. In formal situations, the two groups differed in that HSs used mostly textual functions in their closings. This difference indicates that HSs behaved in a more "formal" way than MSs in formal communicative situations.

### **3.5.2 Closing discourse functions in Greek MSs and HSs**

The distribution of functions in Greek closings was similar across MSs and HSs in Germany and the US (see Figure 12).

The analysis of functions per communicative situation showed that all three speaker groups mostly produced textual functions (codas and resolutions) in formal communicative situations, and the distribution of codas and resolutions was similar in the three groups. In contrast, distributional differences were found in informal situations. Specifically, MSs produced 57 (39%) textual, 52 (36%) subjective, and 37 (25%) intersubjective functions. In HSs in Germany, informal discourse functions were equally distributed (intersubjective: 32/96 (33%), subjective: 33/96 (34%), textual: 31/96 (32%)). HSs in the US had a similar distribution to MSs in that they produced more textual functions (31/58: 53%) than subjective (20/58: 34%) and intersubjective (7/58: 12%), but they differed from MSs and

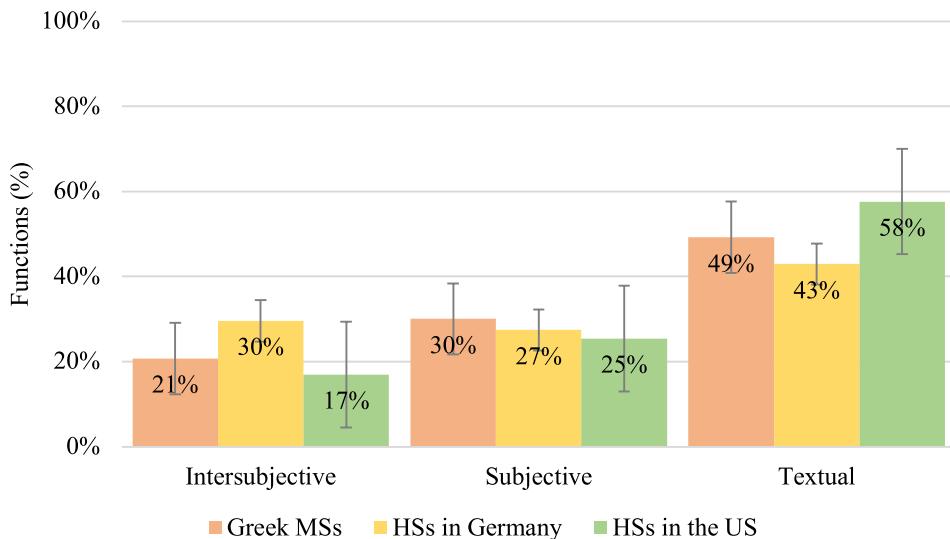


Figure 12: Distribution of functions per speaker group in Greek closings

HSs in Germany in that they did not produce any justifications in their informal closings.

- (52) Justification  
 that argiso ligo tha kathisteriso  
 will delay a.bit will delay  
 'I will be a bit late, I will be delayed' (DEbi15MG\_isG)

The justifications in closings for MSs and HSs in Germany were mostly statements that they would be late, as shown in (52), and often a repetition of the same justification function in the opening. In these cases, speakers opened and ended their narrations in a similar way.

### 3.5.3 Closing discourse functions in Russian MSs and HSs

The analysis of the Russian closings showed that all groups mostly used subjective and textual functions in their closings (see Figure 13).

The analysis of functions across communicative situations showed that all groups used codas and resolutions to a similar extent in formal situations. In informal situations, HSs used more intersubjective functions such as reaction seeking (example (53)) and valedictions than MSs (reaction seeking: MSs: 4/14 (29%), HSs in Germany: 14/37 (38%), HSs in the US: 18/42 (42%); valediction: MSs:

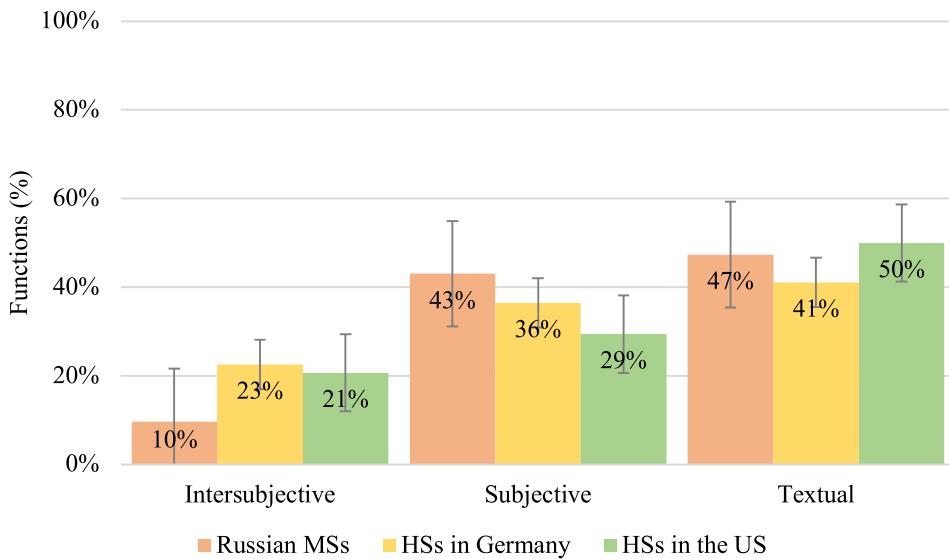


Figure 13: Distribution of functions per speaker group in Russian closings

2/14 (4%), HSs in Germany: 15/37 (41%), HSs in the US: 15/42 (36%). This shows that in informal situations, HSs tended to be more interactive than MSs.

- (53) Reaction seeking
  - a perezvoni mne
  - and call.back me
  - ‘And call me back’ (USbi14MR\_isR)
- (54) Resolution
  - nu k.schast’ju nikto ne postradal
  - well luckily no one was hurt
  - ‘Well, luckily noone got hurt’ (RUmo07FR\_isR)
- (55) Coda
  - vot vsjo chto ja videla
  - here all that I seen
  - ‘That’s all I saw’ (USbi18FR\_fsR)

Regarding textual functions, MSs produced more resolutions (54) than HSs (MSs: 62/99 (63%), HSs in Germany: 18/72 (25%), HSs in the US: 26/74 (35%)), whereas HSs produced more codas (55) than MSs (MSs: 29/99 (29%), HS in Germany: 46/72 (64%), HSs in the US: 39/74 (53%)). Therefore, Russian HSs did not

follow the Russian MS strategy of closing their narrations with resolutions as they produced more codas than resolutions in their closings. Future examination of the majority language texts (German, English) of Russian HSs will shed more light onto these results. The current results showed that German MSs produced slightly more codas than resolutions and thus the overall distribution of codas and resolutions was quite balanced in German MSs (see Section 3.5.1). Thus, the strategy of the Russian HSs in Germany to produce more codas than resolutions does not seem to stem from their majority language. A closer look into the distribution of codas and resolution across communicative situations showed that although Russian MSs overall produced more resolutions than codas, they actually produced more codas than resolutions in formal situations (codas: 57%, resolutions: 20%). Russian HSs in Germany produced more codas than resolutions both in formal (codas: 70%, resolutions: 21%) and informal situations (codas: 64%, resolutions: 25%). It may thus be the case that Russian HSs in Germany applied a formal situation strategy to informal situations.

### 3.5.4 Closing discourse functions in Turkish MSs and HSs

The analysis of functions in the Turkish closings showed that all groups mostly produced textual and subjective functions to a similar extent. In addition, Turkish HSs produced fewer intersubjective functions than MSs. The distribution of functions in Turkish closings per speaker group can be seen in Figure 14.

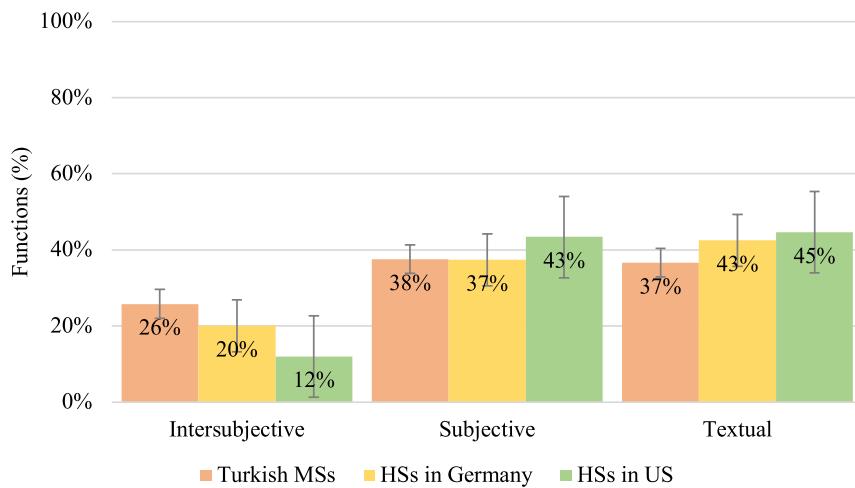


Figure 14: Distribution of functions per speaker group in Turkish closings

The analysis of closing functions per communicative situation showed that textual functions had the highest frequency across all three groups in formal situations. There was only a small difference between MSs and HSs, as MSs produced a slightly higher amount of resolutions (MSs: 25/57 (44%), HSs in Germany: 17/49 (35%), HSs in the US: 12/34 (35%)) whereas HSs produced more codas (MSs: 18/57 (32%), HSs in Germany: 24/49 (49%), HSs in the US: 15/34 (44%)). In informal situations, all Turkish groups produced a large number of subjective functions, which was mostly evaluations (57) and personal statements (56). The distribution of these functions was similar in all three groups: all produced more personal statements than evaluations (personal statements: MSs: 51/97 (52%), HSs in Germany: 39/86 (45%), HSs in the US: 32/65 (57%); evaluations: MSs: 39/97 (40%), HSs in Germany: 31/86 (36%), HSs in the US: 17/56 (30%)). Intersubjective functions were quite limited in informal communicative situations for all groups with reaction seeking being the relatively most frequent function. Regarding textual functions, all groups had a similar distribution of resolutions, codas and encapsulations, with resolutions being slightly more than codas.

(56) Personal statement

film izlemiş gibi oldum bi dakikada  
movie watched like have.been in one.minute

'I felt like watching a movie in a minute' (TUMo76MT\_isT)

(57) Evaluation

adamin yüzünden bence kesinlikle o topla oynayan adamin  
man's because.of I.think definitely that ball.with playing man's  
yüzünden yani  
because.of well

'I think it's the guy's fault, it's definitely the guy playing with the ball'  
(DEBi05FT\_isT)

In formal communicative situations, textual functions had the highest frequency, and we found only a small difference between MSs and HSs as MSs produced a slightly higher amount of resolutions (MSs: 25/57 (44%), HSs in Germany: 17/49 (35%), HSs in the US: 12/34 (35%)) whereas HSs produced more codas (MSs: 18/57 (32%), HSs in Germany: 24/49 (49%), HSs in the US: 15/34 (44%)). In sum, all three Turkish-speaking groups had a similar discourse function distribution in informal situations.

### **3.5.5 Summary**

Overall, the analysis of closings showed a higher use of subjective functions in relation to openings, and very few differences across speaker groups. In formal communicative situations, no major differences were found between MSs and HSs in Greek, Russian and Turkish narrations. In German formal situations, HSs differed from MSs in that HSs did not produce any valedictions in spoken and written formal situations. In informal situations, German and Greek HSs in the US produced a high number of textual functions but only German HSs differed from MSs. Russian HSs differed from MSs in that they produced more intersubjective functions than MSs, and Turkish HSs did not show differences from MSs in the distribution of functions in informal situations. Finally, we did not find a consistent pattern across HS groups of different languages concerning the production of resolutions and codas. In German narrations, HSs produced more resolutions than MSs, in Russian narrations MSs produced more resolutions than HSs, and there were no differences between MSs and HSs in Greek and Turkish narrations.

## **4 Discussion and conclusions**

The present study explored the distribution of openings and closings in narratives produced by monolingually raised speakers (MSs) and heritage speakers (HSs) of German, Greek, Russian, and Turkish across different communicative situations. Our main aim was to investigate the universality of macro discourse structures and explore the use of discourse functions in openings and closings across speaker groups and communicative situations. Research Question 1 aimed to explore whether MSs and HSs of German, Greek, Russian, and Turkish use openings and closings in their narrations with a similar frequency. Our analysis partially confirmed the results of previous research on openings and closings arguing for the optionality of openings and closings, especially in electronic interactions (Baron 1998, Crystal 2001, Herring 1996). Our data analysis showed that although not all texts included an opening and a closing, openings occurred in our data with a frequency of 80% and above. This result is similar to that of Bou-Franch (2011) who found 85% of openings in Spanish e-mail conversations of various types. At the same time, other studies examining e-mail interactions such as the one by Waldvogel (2007) found a much lower occurrence of greetings in openings (59%) (see also Wiese et al. 2025 [this volume]). We believe that the very high percentages of openings in our study are most possibly linked to the experimental method. Participants were asked to describe an incident in four different

communicative situations (voice message to friend, instant message to friend, voice message to police, written report to police). It is therefore very possible that participants used the opening sequences to establish each communicative situation.

We then further analyzed our data in order to investigate the discourse functions that are found in openings and closings (Research Question 2). We identified three types of function categories, intersubjective, subjective and textual – with each function category including several more specific discourse functions. This analysis revealed interesting patterns across speaker groups. In the German-speaking group, a greeting was the most common way to introduce an opening sequence and, similarly, a valediction was the most common way to end a closing sequence in MSs. HSs showed a similar pattern as MSs in openings, but not in closings where they produced more codas and resolutions than valedictions. Informal greetings in German spoken and written narrations were used to express closeness and familiarity (see Bou-Franch (2011) for an analysis of informal greetings in e-mail interactions), and formal greetings were used in formal communicative situations by MSs. In contrast, HSs did not use formal greetings in formal communicative situations, and this may constitute evidence for register leveling. A similar strategy was observed in Greek HSs in the US who used informal self-identifications in formal situations. Our data suggest that HSs may use informal language in formal settings due to the familial context of heritage language acquisition. Heritage speakers acquire the language in the family and thus are exposed to more informal language use, and this can be the reason why they may use informal language in formal settings.

As pointed out by Wiese et al. (2022), setting (or “register”) differences between HSs and MSs should not be taken to mean that HSs are less “native” speakers than MSs because formality is not a determining factor for HSs’ “nativeness” (cf. Montrul & Ionin 2012, Polinsky 2018). According to Wiese et al. (2022), formality distinctions are closely tied to social and communicative needs, which may vary among different social groups. This aspect is independent of whether someone is bilingual or monolingual, and in the case of monolinguals, it typically does not affect our perception of individuals as native speakers (Wiese et al. 2022). The results of the present study show that in openings and closings, HSs follow formality conventions of the standard language to a large extent, and even in cases that they exhibit distinct patterns, this does not mean that their German language ability is more restricted in relation to German MSs.

Evidence from the analysis of the Russian and the Turkish data did not show any register leveling effects in HSs. A main strategy found in Russian HSs in Germany was to use justifications in their opening sequences, that is, to provide

a reason why they are calling or writing. This strategy was not prevalent in either the German or the Russian MSs, and thus may be a discourse strategy specific to HSs in Germany. In addition, Russian HSs in Germany and in the US use more intersubjective functions in their openings than MSs, and this may indicate that HSs are more interactive than MSs, or HSs feel much more confident using intersubjective functions in informal settings compared to formal settings and thus make extensive use of them. A large variety of intersubjective functions was also found in the openings of the Turkish-speaking group with a similar distribution across MSs and HSs.

Turning to closings, a general finding was that closings were less frequent than openings in our data. In addition, there was more variation in the distribution of closings than in the distribution of openings across speaker groups. This result contrasts with previous research that showed a high frequency of closings (over 75%) in e-mail interactions (Bou-Franch 2011). This difference is not surprising as e-mails have very different characteristics in general. Being quasi-synchronous in nature, IM always leaves the communication possibility open. In addition, following our argument that speakers in our study may have used the opening sequences to establish the communicative situation, it can be argued here that closings were not so frequent because the communicative situation had already been established through the opening. Previous studies have also found minimal occurrence of farewell closings present (Herring 1996, Waldvogel 2002). Although valediction sequences were quite frequent in the German and Russian texts, they were not so frequent in Greek and Turkish texts. The main functions across all closings in the present study were evaluations of the incident that the participants described and personal statements, that is, switching the topic to something that focuses on the individual and not the situation that participants were asked to describe. In addition, closing sequences included many instances of textual functions such as resolutions and codas. In contrast to opening sequences, closing sequences included much fewer intersubjective functions across all speaker groups.

We further examined the role of age in the frequency of openings and closings (Research Question 3). Previous research has indicated age-related differences in openings and closings, particularly between child and adolescent groups (Dollnick & Pfaff 2013, Tolchinsky et al. 2002). Contrary to previous findings, our study did not show any statistically significant differences between adolescents and adults in the use of openings and closings. It can thus be that there are fewer differences between adolescents and adults than children and adolescents when it comes to macro discourse structures. Future analysis of discourse functions in

openings and closings across adolescents and adults would shed more light on possible age differences in our data.

In summary, the annotation and analysis of openings and closings in the narrative texts of German, Greek, Russian and Turkish MSs and HSs across four different communicative situations provided insights into the macro discourse strategies that MSs and HSs of different languages use to narrate events. Our results showed similar patterns as well as variation relating to cultural and language status differences across different speaker groups, and support the view that although core discourse functions may occur cross-linguistically, discourse functions are also language-specific (see e.g. Márquez Reiter & Luke 2010). We hope that the findings of this exploratory study will contribute to a better understanding of macro discourse structures in cross-cultural and language contact contexts.

## Acknowledgements

This work would not have been possible without the contribution of the brilliant student assistants who spent many hours annotating, discussing issues, and providing valuable feedback. We would thus like to thank: Barbara Zeyer (German and English annotations), Aban Alehojat, Sadaf Rezai, Mateo Vargas-Nuñez (English annotations), Iliuza Akhmetzianova, Ksenia Idrisova, Nadezhda Kushina (Russian annotations), Hannah Plückebaum, Josefine Hundelt, Franziska Groth (German annotations), Aikaterini Tsaroucha (Greek annotations and co-presenting in AMLaP 2023). Special thanks to the Turkish annotation team: Mert Can, Oğuzhan Kuyrukçu, and Simge Türe for taking up the extra task to present in the PiF and HL@Cross conferences, and to Lea Coy for help with pre-publication formatting. We would also like to thank Nadine Zürn, Vicky Rizou and Anke Lüdeling for their feedback on a previous version of this chapter, as well as two anonymous reviewers and the audiences of PiF 2023, HL@cross 2023, and AMLaP 2023 conferences for their helpful and constructive comments and suggestions. This research was supported financially by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit *Emerging Grammars in Language Contact Situations*, project P9 (grant number: 313607803).

## References

- Baron, Naomi S. 1998. Letters by phone or speech by other means: The linguistics of email. *Language & Communication* 18(2). 133–170. DOI: 10.1016/S0271-5309(98)00005-6.
- Berman, Ruth A. 1997. Narrative theory and narrative development: The Labovian impact. *Journal of Narrative and Life History* 7(1-4). 235–244. DOI: 10.1075/jnlh.7.29nar.
- Berman, Ruth A. & Irit Katzenberger. 2004. Form and function in introducing narrative and expository texts: A developmental perspective. *Discourse Processes* 38(1). 57–94. DOI: 10.1207/s15326950dp3801\_3.
- Berman, Ruth A. & Ludo Verhoeven. 2002. Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy* 5(1). 1–43. DOI: 10.1075/wll.5.1.02ber.
- Bou-Franch, Patricia. 2011. Openings and closings in Spanish email conversations. *Journal of Pragmatics* 43(6). 1772–1785. DOI: 10.1016/j.pragma.2010.11.002.
- Brinton, Laurel J. 1996. *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110907582.
- Coronel-Molina, Serafin M. 1998. Openings and closings in telephone conversations between native Spanish speakers. 14. 49–68.
- Crystal, David. 2001. *Language and the Internet*. 1st edn. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139164771.
- Do, Hoa, Tran Huu Thuy Giang & Ket Mai. 2018. Vietnamese telephone openings: Both universals and particulars. *Language and Dialogue* 8(3). 363–389. DOI: 10.1075/ld.00022.do.
- Dollnick, Meral & Carol Wollman Pfaff. 2013. Entwicklung von Diskursorganisation in verschiedenen Sprachen: Einführung und Schluss in narrativen und in erörternden Texten in Türkisch, Deutsch und Englisch bei Schülern mit türkischer Erstsprache in Berlin. In Tabea Becker & Petra Wieler (eds.), *Erzählforschung und Erzähldidaktik heute: Entwicklungslinien, Konzepte, Perspektiven*, 147–162. Tübingen: Stauffenburg.
- Gkouma, Athanasia. 2023. “σε υπερευχαριστώ, με σώξεις!”: Η γλωσσική πράξη της ευχαριστίας από φυσικούς/ές ομιλητές/τριες της ελληνικής [“Thank you, you’re saving me!”: The linguistic act of thanking in native Greek speakers”]. *Studies in Greek Linguistics* 42. 105–114.
- Godard, Daniel. 1977. Same setting, different norms: Phone call beginnings in France and the United States. *Language in Society* 6(2). 209–219. <https://www.jstor.org/stable/4166925>.

- Herring, Susan C. 1996. Two variants of an electronic message schema. In Susan C. Herring (ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (Pragmatics & Beyond New Series 39), 81–108. Amsterdam: John Benjamins. DOI: 10.1075/pbns.39.08her.
- Hopper, Robert. 1989. Speech in telephone openings: Emergent interaction v. routines. *Western Journal of Speech Communication* 53(2). 178–194. DOI: 10.1080/10570318909374299.
- Hopper, Robert & Chia-Hui Chen. 1996. Languages, cultures, relationships: Telephone openings in Taiwan. *Research on Language & Social Interaction* 29(4). 291–313. DOI: 10.1207/s15327973rlsi2904\_1.
- Hopper, Robert & Nada Koleilat-Doany. 1989. Telephone openings and conversation universals: A study in three languages. In Stella Ting-Toomey & Felipe Korzenny (eds.), *Language, communication and culture: Current directions*, 157–179. Newbury Park, CA: Sage.
- Houtkoop-Steenstra, Hanneke. 1991. Opening sequences in Dutch telephone conversation. In Deirdre Boden & Don Zimmerman (eds.), *Talk and social structure*, 232–250. Berkeley: University of California Press.
- Houtkoop-Steenstra, Hanneke. 2002. Questioning turn format and turn-taking problems in standardized interviews. In Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer & Jonannes van der Zouwen (eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview*, 243–261. New York: Wiley.
- Krause, Thomas & Amir Zeldes. 2014. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1). 118–139. DOI: 10.1093/llc/fqu057.
- Labov, William. 1972. The transformation of experience in narrative syntax. In William Labov (ed.), *Language in the inner city: Studies in the black English vernacular* (Conduct and communication 3), 355–396. Philadelphia: Univ. of Pennsylvania Press.
- Lindström, Anna. 1994. Identification and recognition in Swedish telephone conversation openings. *Language in Society* 23(2). 231–252. <https://www.jstor.org/stable/4168515>.
- Lüdeling, Anke, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Sofia Grigoriadou, Rahel Gajaneh Hartz, Kateryna Iefremenko, Esther Jahns, Kalliopi Katsika, Mareike Keller, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Onur Özsoy, Tatiana Pashkova, Maria Pohle, Judith Purkarthofer, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehaye, Heike Wiese, Sabine Zerbian, Yu-

- lia Zuban & Nadine Zürn. 2024. *RUEG Corpus*. Version 1.0. Zenodo. DOI: 10.5281/zenodo.11234583.
- Luke, Kang-kwong. 2002. The initiation and introduction of first topics in Hong Kong telephone calls. In Kang-kwong Luke & Theodossia-Soula Pavlidou (eds.), *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (Pragmatics & Beyond New Series 101), 171–200. Amsterdam: John Benjamins. DOI: 10.1075/pbns.101.12luk.
- Luke, Kang-kwong & Theodossia-Soula Pavlidou (eds.). 2002. *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (Pragmatics & Beyond New Series 101). Amsterdam: John Benjamins. DOI: 10.1075/pbns.101.
- Márquez Reiter, Rosina & Kang-kwong Luke. 2010. Telephone conversation openings across languages, cultures and settings. In Anna Trosborg (ed.), *Pragmatics across languages and cultures*, 103–138. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110214444.1.103.
- Montrul, Silvina & Tania Ionin. 2012. Dominant language transfer in Spanish heritage speakers and second language learners in the interpretation of definite articles. *The Modern Language Journal* 96(1). 70–94. DOI: 10.1111/j.1540-4781.2012.01278.x.
- Özsoy, Onur, Kateryna Iefremenko & Christoph Schroeder. 2022. Shifting and expanding clause combining strategies in heritage Turkish varieties. *Languages* 7(3). 242. DOI: 10.3390/languages7030242.
- Park, Yong-Yae. 2002. Recognition and identification in Japanese and Korean telephone conversation openings. In Kang-kwong Luke & Theodossia-Soula Pavlidou (eds.), *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (Pragmatics & Beyond New Series 101), 25–47. Amsterdam: John Benjamins. DOI: 10.1075/pbns.101.06par.
- Pavlidou, Theodossia-Soula. 1994. Contrasting German-Greek politeness and the consequences. *Journal of Pragmatics* 21(5). 487–511. DOI: 10.1016/0378-2166(94)90026-4.
- Pavlidou, Theodossia-Soula. 1997. The last five turns: Preliminary remarks on closings in Greek and German telephone calls. *International Journal of the Sociology of Language* 126(1). DOI: 10.1515/ijsl.1997.126.145.
- Pavlidou, Theodossia-Soula. 2014. Phases in discourse. In Klaus P. Schneider & Anne Barron (eds.), *Pragmatics of discourse*, 353–384. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110214406-014.
- Polinsky, Maria. 2018. *Heritage languages and their speakers* (Cambridge Studies in Linguistics 159). Cambridge: Cambridge University Press. DOI: 10.1017/9781107252349.

- Raclaw, Joshua. 2008. Two patterns for conversational closings in instant message discourse. *Colorado Research in Linguistics* 21. DOI: 10.25810/5E6R-FK53.
- Ravid, Dorit & Liliana Tolchinsky. 2002. Developing linguistic literacy: A comprehensive model. *Journal of Child Language* 29(2). 417–447. DOI: 10.1017/S0305000902005111.
- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2). 155–163. DOI: 10.1177/1367006909339814.
- RStudioTeam. 2020. *RStudio: Integrated development for R*. Boston, MA. <http://www.rstudio.com/>.
- Sacks, Harvey. 1967. *Harvey Sacks lecture transcripts*. [https://oac.cdlib.org/findaid/ark:/13030/kt4489r9p9/entire\\_text/](https://oac.cdlib.org/findaid/ark:/13030/kt4489r9p9/entire_text/).
- Schegloff, Emanuel A. 1968. Sequencing in conversational openings. *American Anthropologist* 70(6). 1075–1095. DOI: 10.1525/aa.1968.70.6.02a00030.
- Schegloff, Emanuel A. 1979. Identification and recognition in telephone conversation openings. In George Psathas (ed.), *Everyday language: Studies in ethnomethodology*, 23–78. New York: Irvington Publishers.
- Schegloff, Emanuel A. 2007. *Sequence organization in interaction: A primer in conversation analysis*. 1st edn. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511791208.
- Schegloff, Emanuel A. & Harvey Sacks. 1973. Opening up closings. *Semiotica* 8(4). DOI: 10.1515/semi.1973.8.4.289.
- Schmidt, Thomas & Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *Handbook on corpus phonology*, 402–419. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199571932.013.030.
- So'o, Ainslie & Anthony J. Liddicoat. 2000. Telephone openings in Samoan. *Australian Review of Applied Linguistics* 23(1). 95–107. DOI: 10.1075/aral.23.1.06soo.
- Taleghani-Nikazm, Carmen. 2002. A conversation analytical study of telephone conversation openings between native and nonnative speakers. *Journal of Pragmatics* 34(12). 1807–1832. DOI: 10.1016/S0378-2166(02)00049-8.
- Tolchinsky, Liliana, Victoria Johansson & Anita Zamora. 2002. Text openings and closings in writing and speech: Autonomy and differentiation. *Written Language & Literacy* 5(2). 219–253. DOI: 10.1075/wll.5.2.05tol.
- Topaj, Nathalie. 2020. *Grammatical and pragmatic use of referential expressions in picture-based narratives of bilingual and monolingual children in Russian and German*. Humboldt-Universität zu Berlin. (Doctoral dissertation). DOI: 10.18452/21672.

- Tsehaye, Wintai, Tatiana Pashkova, Rosemarie Tracy & Shanley E. M. Allen. 2021. Deconstructing the native speaker: Further evidence from heritage speakers for why this horse should be dead! *Frontiers in Political Science* 12. 1–14. DOI: 10.3389/fpsyg.2021.717352.
- Waldvogel, Joan. 2002. Some features of workplace emails. *New Zealand English Journal* 16. 42–52.
- Waldvogel, Joan. 2007. Greetings and closings in workplace email. *Journal of Computer-Mediated Communication* 12(2). 456–477. DOI: 10.1111/j.1083-6101.2007.00333.x.
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI* (Bamberg Studies in English Linguistics 59), 105–117. Frankfurt am Main: Peter Lang.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Katerina Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehaye, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. 5982. DOI: 10.3389/fpsyg.2021.717973.
- Wiese, Heike, Artemis Alexiadou, Shanley E. M. Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Anke Lüdeling, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Tracy Rosemarie, Christoph Schroeder, Luka Szucsich, Wintai Tsehaye, Sabine Zerbian & Yulia Zuban. 2021. *RUEG Corpus*. Version 0.4.0. Zenodo. DOI: 10.5281/zenodo.3236068.
- Wiese, Heike & Annika Labrenz. 2021. Emoji as graphic discourse markers: Functional and positional associations in German WhatsApp messages. In Daniël Van Olmen & Jolanta Šinkūnienė (eds.), *Pragmatic markers and peripheries* (Pragmatics & Beyond New Series 325), 277–300. Amsterdam: John Benjamins. DOI: 10.1075/pbns.325.10wie.
- Wiese, Heike, Annika Labrenz & Albrun Roy. 2025. Tapping into speakers' repertoires: Elicitation of register-differentiated productions across groups. In Shanley E. M. Allen, Mareike Keller, Artemis Alexiadou & Heike Wiese (eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group*, 33–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.15775159.
- Zimmerman, Don H. 1984. Talk and its occasion: The case of calling the police. In Deborah Schiffrin (ed.), *Meaning, form and use in context: Linguistic applications*, 210–228. Washington, D.C.: Georgetown University Press. [https:](https://)

*Kalliopi Katsika, Annika Labrenz, Kateryna Iefremenko & Shanley E.M. Allen*

//repository.library.georgetown.edu/bitstream/handle/10822/555477/GURT\_1984.pdf.

Zimmerman, Don H. 1992. The interactional organization of calls for emergency. In Paul Drew & John Heritage (eds.), *Talk at work: Interaction in institutional settings*, 418–469. Cambridge: Cambridge University Press.

# Name index

- Aalberse, Suzanne P., 4, 36, 37, 160, 164, 166, 259, 260  
Acarlar, Funda, 221  
Adamović, Milan, 290  
Adamson, Luke James, 194  
Adli, Aria, 35  
Agathopoulou, Eleni, 230  
Ahmad, Dohra, 122  
Aijmer, Karin, 509  
Aksu-Koç, Ayhan, 221, 254  
Alamillo, Rosalva, 329, 335, 338, 355  
Alexiadou, Artemis, 3, 184, 220, 222, 227, 228, 233, 239, 241, 279, 386, 423, 437, 452  
Alexopoulou, Theodora, 102  
Allen, Shanley E. M., 75, 84, 144, 152, 194, 205, 461, 465, 470  
Alm, Maria, 497  
Altmann, Hans, 394, 395  
Anagnostopoulou, Elena, 184, 194  
Andersen, Christiane, 142  
Andersen, Gisle, 496  
Andersen, Roger W., 230  
Andreou, Maria, 155, 185, 204  
Andrews, Molly, 258  
Androutsopoulos, Jannis, 499, 503, 517  
Anthony, Laurence, 339  
Antomo, Mailin, 145  
Antoniou, Kyriakos, 498  
Antonova-Ünlü, Elena, 221, 234  
Anzt, Hartwig, 100  
Argyri, Efrosyni, 180, 203  
Ariel, Mira, 194, 496  
Arnaus Gil, Laia, 112  
Arnaut, Karel, 112  
Arnbjörnsdóttir, Birna, 386  
Arnold, Jennifer E., 194, 380  
Arslan, Seçkin, 237, 451, 453  
Atwell, Eric, 80  
Audring, Jenny, 328  
Auer, Peter, 55, 432, 493  
Austin, Jennifer, 183  
Azar, Zeynep, 180, 187, 203, 204, 451, 453, 461, 462, 477  
Baayen, Rolf Harald, 189  
Babby, Leonard, 255  
Bailyn, John F., 380, 391, 395  
Baker, Colin, 120–122  
Balci, Mustafa, 292  
Bar-Shalom, Eva, 235  
Barabashova, Kristina, 394, 397, 434, 436  
Barbosa, Poliana, 460, 477  
Baron, Naomi S., 499, 567  
Baroncini, Ioli, 185, 204  
Bates, Douglas, 196, 336  
Batiukova, Olga, 339  
Bauer, Laurie, 80  
Baumann, Stefan, 380, 391, 457, 458  
Bayer, Josef, 310  
Bayram, Fatih, 36, 237, 238, 260, 307  
Beckman, Mary E., 458, 459

## *Name index*

- Beckner, Clay, 97  
Belz, Malte, 350  
Benmamoun, Elabbas, 180, 192, 221  
Berman, Ruth A., 258, 472, 528, 530,  
    531  
Bethard, Steven J., 90  
Bialystok, Ellen, 450  
Biber, Douglas, 35, 43, 83, 257, 334,  
    383, 470  
Bilá, Magdaléna, 42  
Birner, Betty, 380  
Bizzarri, Camilla, 185  
Blackledge, Adrian, 112, 401  
Blakemore, Diane, 496  
Blau, Otto, 290  
Blom, Elma, 165  
Blommaert, Jan, 9  
Blühdorn, Hardarik, 388, 496  
Blum, Frederic, 3, 75, 278  
Boas, Hans C., 142, 148, 149, 158, 326,  
    332  
Boas, Hans Christian, 142, 148, 149,  
    158, 326  
Bock, Kathryn J., 380  
Boers, Frank, 329, 336  
Boersma, Paul, 76, 417, 455  
Böhm, Stefanie, 292  
Bolden, Galina B., 496, 505  
Bombaci, Alessio, 290  
Bondarko, Aleksandr V., 235  
Bonefeld, Meike, 122, 128  
Bonfiglio, Thomas Paul, 2  
Bonney, Rick, 129  
Bonvin, Audrey, 329, 336  
Boon, Erin Diane, 183, 228  
Böttcher, Marlene, 150, 350, 419, 423,  
    431, 457–459  
Bou-Franch, Patricia, 567–569  
Bousquette, Joshua, 395, 396  
Bredel, Ursula, 499  
Brehmer, Bernhard, 121, 184, 205,  
    389, 390  
Breillard, Žan, 185  
Breitbarth, Anne, 432, 433  
Brendemoen, Bernt, 289  
Brezina, Vaclav, 189, 291  
Bridges, Kelly, 160  
Brinton, Laurel J., 432, 495, 536  
Brisk, María Estela, 112  
Britain, David, 191  
Bross, Fabian, 388  
Brutt-Griffler, Janina, 2  
Bryant, Doreen, 356  
Bryzgunova, Elena, 424  
Bubenhofer, Noah, 93  
Budennaya, Evgeniya, 185  
Bulut, Christiane, 296  
Bunk, Oliver, 3, 12, 13, 16, 48, 145, 155,  
    291, 386–389, 394–396, 400,  
    401, 415, 430–433, 467, 469,  
    477  
Burbelko, Mariya, 201  
Busch, Brigitta, 112  
Busch, Florian, 503, 517  
Büyükyıldırım, Nisa, 220, 223, 236,  
    237, 239  
Bybee, Joan L., 328, 432  
Bylund, Emanuel, 452, 453, 476  
Caloi, Irene, 122  
Camacho, José, 206  
Canagarajah, Suresh, 2  
Capps, Lisa, 56  
Cejtin, Stella N., 221, 234  
Chafe, Wallace L., 194, 257, 380  
Chang, Charles B., 414, 436, 451  
Chen, Baoguo, 450  
Chen, Chia-Hui, 530

- Chevalier, Joan F., 222  
Chociej, Joanna, 183  
Choi, Soonja, 254  
Chondrogianni, Vasiliki, 450  
Chou, Chao-Ting Tim, 206  
Christodoulidou, Maria, 498  
Christoffersen, Katherine, 417  
Çiçek, Büsra, 238  
Çimpoeş, Lübov S., 309  
Clahsen, Harald, 159  
Clark, Herbert H., 350, 380  
Clavel-Arroitia, Begoña, 329  
Clyne, Michael C., 142, 146, 148, 149,  
    155, 158, 326  
Clyne, Michael G., 149, 349  
Colantoni, Laura, 423, 437  
Colombi, M. Cecelia, 467  
Comrie, Bernard, 181, 220, 239  
Comstock, Lindy Burden, 420–423  
Conrad, Susan, 43, 257, 334  
Contemori, Carla, 451, 453, 461  
Conti, Erica, 394, 401  
Cook, Vivian, 2  
Corbett, Greville G, 181, 205  
Coronel-Molina, Serafin M., 530  
Coşkun Kunduz, Aylin, 237  
Covington, Michael A., 335  
Crible, Ludivine, 495  
Croft, William, 97  
Crystal, David, 567  
Csató, Éva Á., 292, 296, 306  
Cuza, Alejandro, 220  
Cysouw, Michael, 85  
Czapka, Sophia, 329  
  
Daller, Helmut, 329  
Daskalaki, Evangelia, 165, 185, 189,  
    204  
De Houwer, Annick, 119  
  
de Jong, Nivja H., 419, 455  
De Korne, Haley, 123  
de Marneffe, Marie-Catherine, 94  
de Mier, Mariela Vanesa, 257, 260,  
    467  
de Ruiter, Laura E., 457  
Degani, Tamar, 328  
Déhé, Nicole, 414, 422, 423, 432, 451,  
    457  
Delidakı, Sophia, 221  
Demirok, Ömer, 310  
Destanov, Ekrem, 289  
Dewaele, Jean-Marc, 39, 82  
Dewald, Anika, 395  
Dhongde, Ramesh V., 310  
Di Domenico, Elisa, 185, 204  
Dickhäuser, Oliver, 122, 128  
Diewald, Gabriele, 388, 495  
Dillon, Kathleen, 392  
Dixon, Robert MW, 181  
Dikilitaş, Kenan, 187, 204  
Do, Hoa, 530  
Doerfer, Gerhard, 292  
Dollnick, Meral, 528, 530–532, 546,  
    569  
Döpke, Susanne, 167, 188, 203  
Dorleijn, Margreet, 296  
Drach, Erich, 143, 420  
Dressler, Wolfgang U., 222  
Drimba, Vladimir, 290  
Druskat, Stephan, 90  
Dryer, Matthew S., 182  
Du Ryer, André, 290  
Dubinina, Irina, 180, 186, 203, 204,  
    392, 393, 396, 397, 417  
Dündar, Fuat, 296  
Duranti, Alessandro, 394, 395  
Dussias, Paola E., 260

## *Name index*

- Edwards, Viv, 113, 121  
Egbert, Jesse, 334  
Ehlich, Konrad, 497  
Eichinger, Ludwig M., 142  
Elabdali, Rima, 329  
Ellis, Nick C., 97  
Elspaß, Stephan, 388  
Embick, David, 4  
Enç, Mürvet, 187  
Engelberg, Stefan, 328  
  
Fairclough, Marta, 327  
Fehrmann, Dorothee, 185  
Fei, Hao, 90  
Feldhausen, Ingo, 414  
Felfe, Marc, 330, 332  
Felser, Claudia, 451, 453  
Fernández-Dobao, Ana, 233  
Fernández-Villanueva, Marta, 497  
Filiaci, Francesca, 461  
Fischer, Kerstin, 496  
Fishman, Joshua A., 112  
Flege, James Emil, 476  
Flores, Cristina, 4, 158, 184, 205, 414,  
    495  
Flores, Nelson, 122  
Flubacher, Mi-Cha, 112  
Földes, Csaba, 142  
Fougeron, I. I., 185  
Fox Tree, Jean E., 350, 471  
Franks, Steven, 255  
Frascarelli, Mara, 182  
Fraser, Bruce, 496  
Freitag, Amy, 129  
Frey, Werner, 395  
Freywald, Ulrike, 145, 432  
Frick, Karina, 181, 205  
Fridman, Clara, 344, 355  
Friedman, Victor A., 292  
  
Fries, Norbert, 183  
Fritzenschaft, Agnes, 167  
Fromkin, Victoria, 350  
Fuller, Janet M., 142, 326, 494  
  
Gagarina, Natalia, 121, 180, 203, 220,  
    221, 223, 224, 234, 235, 240  
Gallotta, Aldo, 290  
Gamper, Jana, 386  
Garrod, Simon, 152  
Gast, Volker, 255  
Gavarró, Anna, 185  
Gawlitzek, Ira, 131, 450  
Gawlitzek-Maiwald, Ira, 119, 120, 167,  
    188, 203  
Gelman, Andrew, 226  
Geluykens, Ronald, 395, 467  
Georgakopoulou, Alexandra, 498  
Georgievits, Bartholomaeus, 296  
Gharibi, Khadijeh, 329, 336  
Giannakou, Aretousa, 185, 204  
Gibbon, Dafydd, 150, 350  
Gkouma, Athanasia, 529  
Godard, Daniel, 530  
Gogolin, Ingrid, 2, 257  
Göksel, Ash, 307  
Goldberg, Adele E., 328  
Gollan, Tamar H., 419  
Columbic, Yaela N., 128  
Gonser, Monika, 116  
Gorba, Celia, 450  
Goschler, Juliana, 260  
Gosling, Samuel D., 417  
Goss, Emily L., 494  
Götze, Michael, 428, 458, 459  
Gould, Elliot, 80  
Goutsos, Dionysis, 498  
Gozli, Davood, 82  
Granger, Sylviane, 41

- Grant, Anthony P., 258  
Gregory, Michelle L., 395  
Grenoble, Lenore A., 259  
Grice, Martine, 207  
Gries, Stefan Th., 423  
Grimm, Angela, 260  
Grosjean, François, 1, 3, 9  
Guizarro-Fuentes, Pedro, 3, 180, 221  
Gülich, Elisabeth, 56  
Gullberg, Marianne, 461  
Gülsevin, Gürer, 292  
Gülzow, Insa, 235  
Gundel, Jeanette K., 194  
Güngör, Veyis, 287  
Günßen, Ahmet, 292  
Günther, Susanne, 432, 493  
Guryev, Alexander, 395  
Gutiérrez-Cellen, Vera F., 335  
Guy, Gregory R., 35  
  
Haberzettl, Stefanie, 257, 260  
Haegeman, Liliane, 183  
Haider, Hubert, 144  
Haig, Geoffrey, 296  
Haliloglu, Neşe, 289, 309  
Halliday, Michael A. K., 380  
Härtl, Holden, 332  
Harvie, Dawn, 183  
Haspelmath, Martin, 228, 253, 287,  
    433  
Haviland, Susan E., 380  
Hayakawa, Samuel I., 113  
Hazai, György, 289, 293, 296  
Haznedar, Belma, 180, 187, 203, 204  
Heffening, Willi, 290, 296  
Heine, Bernd, 494  
Helbig, Gerhard, 388  
Helmholdt, Wolfgang, 289  
Hendriks, Henriette, 461, 462  
  
Hennig, Mathilde, 497, 505  
Hentschel, Elke, 388  
Herbinus, Johannes, 290  
Herkenrath, Annette, 260  
Herring, Susan C., 567, 569  
Herschensohn, Julia, 233  
Hervé, Coralie, 468  
Hickmann, Maya., 462  
Hieke, Adolf E., 350  
Higby, Eve, 259  
Hilpert, Martin, 380, 383  
Himmelmann, Nikolaus P., 417, 433,  
    457  
Hinrichs, Erhard W., 90  
Hinton, Leanne, 121  
Hirschberg, Julia, 432  
Hirschmann, Hagen, 98  
Hiver, Phil, 97  
Hock, Hans H., 310  
Hoff, Erika, 160  
Hohenstein, Jill, 450  
Höhle, Tilman N., 143, 388, 420  
Holmberg, Anders, 182  
Holton, David, 232  
Hopp, Holger, 113, 142, 148, 149, 157,  
    326  
Hopper, Paul J., 495  
Hopper, Robert, 529, 530  
Housen, Alex, 151, 257  
Houtkoop-Steenstra, Hanneke, 530  
Hržica, Gordana, 329  
Huddleston, Rodney, 383, 470  
Huffines, Marion Lois, 142  
Hughes, Diana L., 397, 417  
Hulk, Aafke, 188  
  
Iefremenko, Kateryna, 14, 254, 256,  
    259, 277, 279, 291, 296, 304,  
    306, 307

## *Name index*

- İğci, Alpay, 292  
Ihsane, Tabea, 183  
Illésházy, Miklós, 296  
Imo, Wolfgang, 55, 144, 157, 388, 432, 495  
Ionin, Tania, 451, 453, 465, 528, 568  
Irwin, David E., 380  
Isurin, Ludmila, 186, 189, 204, 392  
Ivanova, Iva, 451, 453, 461  
Ivanova-Sullivan, Tania, 180, 186, 203, 392  
Iwashita, Noriko, 417
- Jable, Ergin, 292  
Jackendoff, Ray, 328  
Jankowski, Bridget L., 396, 468  
Jarvis, Scott, 329, 335, 339, 450  
Jasinskaja, Katja, 380, 427, 430  
Jia, Ruiting, 450  
Johannessen, Janne Bondi, 4, 142  
Jóhannsdóttir, Kristín M, 233  
Johanson, Lars, 146, 149, 256, 259, 288, 292, 294, 296, 306, 307  
Joseph, Brian D., 255, 288  
Jucker, Andreas H., 419, 432  
Junghanns, Uwe, 185
- Kagan, Olga, 121, 239, 392  
Kahane, Sylvain, 85, 89  
Kakuk, Suzanne, 289  
Kaland, Constantijn, 459  
Kallestinova, Elena, 380, 389, 395  
Karaca, Figen, 238  
Karakoç, Birsel, 260, 307  
Karaşinik, Bahtişen, 309  
Karst, Karina, 123  
Kassambara, Alboukadel, 336  
Kasstan, Jonathan R, 121  
Katona, Louis K., 289, 292
- Katsika, Kalliopi, 13, 14, 50, 261, 471, 473, 502  
Katzenberger, Irit, 472, 528, 531  
Keenan, Elinor O., 467  
Keim, Inken, 128  
Keller, Mareike, 142, 146, 326  
Kerslake, Celia, 256, 307  
Keskin, Cem, 3, 13, 259, 260, 289, 294, 304, 306, 308, 310, 467, 508  
Kim, Ji-Young, 423, 430  
King, Tracy Holloway, 380, 394, 395, 427  
Kinn, Kari, 386  
Kirli, Nadejda, 292  
Kisselev, Olesya, 392, 393  
Kissling, Hans Joachim, 296  
Kistanova, Ekaterina V., 221, 234, 235  
Kiyota, Masaru, 220  
Klotz, Martin, 11, 38, 42, 70, 76, 84, 90, 327, 455  
Knobloch, Clemens, 80  
Koban Koç, Didem, 180, 187, 203, 204  
Kolelat-Doany, Nada, 530  
Konerding, Klaus-Peter, 497  
König, Ekkehard, 255  
König, Katharina, 57  
Konstantzou, Katerina, 221  
Kopotev, Mikhail, 329  
Kornfilt, Jaklin, 307  
Kortmann, Bernd, 255  
Korybski, Tomasz, 329, 335  
Kotsinas, Ulla-Britt, 386, 431  
Kovtunova, Irina I., 389  
Kožina, Margarita N, 257  
Krause, Thomas, 11, 76, 90, 193, 455, 533  
Krifka, Manfred, 122, 380, 457  
Kroll, Judith F., 419  
Kruschke, John, 226

- Küçükcan, Talip, 287  
Kügler, Frank, 458  
Kühl, Karoline, 386  
Kupisch, Tanja, 3, 9, 10, 34, 35, 414, 450–453, 457, 476  
Kuprianova, T. F., 255  
Kyle, Kristopher, 335, 456  
  
Laarmann-Quante, Ronja, 98  
Labov, William, 257, 258, 472, 528, 530  
Labrenz, Annika, 12, 13, 50, 51, 102, 142, 151, 275, 452, 471, 474, 495–497, 502, 503, 515, 517, 540  
Ladd, D. Robert, 427, 458  
Laleko, Oksana, 205, 220, 221, 234, 239, 392, 393, 396, 397, 417  
Lambelet, Amelia, 329  
Lambrecht, Knud, 394  
Lambrou, Marina, 258  
Lanza, Elizabeth, 119  
Lapkin, Sharon, 234  
Larsen-Freeman, Diane, 97  
Larsson, Ida, 386  
Lattey, Elsa, 142, 146, 148, 326, 386, 431  
Laufer, Batia, 329  
Lavoie, Brian F., 100  
Lee, Chungmin, 205  
Lee, Miseon, 451, 452, 476  
Lee-Ellis, Sunyoung, 451  
Lehmann, Christian, 252–254, 278  
Lehmann, Nico, 258  
Lein, Tatjana, 453  
Lenth, Russell V., 336  
Leone, Andrea R., 129  
Leonelli, Sabina, 79, 99  
Levelt, Willem J. M., 350  
  
Li, Xiaoshi, 181, 205  
Liddicoat, Anthony J., 530  
Lieven, Elena, 235  
Lindström, Anna, 530  
Ling, Rich, 499  
Lippi-Green, Rosina, 401  
Litman, Diane, 432  
Lloyd-Smith, Anika, 451, 453, 457  
Loban, Walter, 54  
Lohndal, Terje, 4, 228, 386  
Louden, Mark Laurence, 142, 148  
Lüdecke, Daniel, 336  
Lüdeling, Anke, 8, 11, 42, 70, 75, 98, 150, 193, 225, 240, 260, 330, 332–334, 382, 386, 416, 428, 453, 533  
Luke, Kang-kwong, 471, 528–531, 537, 570  
  
Maas, Utz, 259  
MacWhinney, Brian, 476  
Madariaga, Nerea, 182, 185, 186, 255  
Makarova, Veronika, 424  
Malvern, David, 329, 335  
Margaza, Panagiota, 185  
Marinis, Theodoros, 450  
Márquez Reiter, Rosina, 529, 531, 570  
Martynova, Maria, 181, 257, 260, 277, 279, 389, 390, 392, 393, 400  
Maschler, Yael, 494, 496  
Massam, Diane, 206  
Massip-Bonet, Àngels, 97  
Matras, Yaron, 258, 292, 326, 494, 496, 518, 519  
Matthiessen, Christian M. I. M., 334  
McCardle, Peggy, 450  
McCarthy, Philip M., 335  
McCarus, Ernest N., 296  
McCloskey, Jim, 194

## *Name index*

- McFall, Joe D., 335  
Meelen, Marieke, 386, 431  
Meibauer, Jörg, 499, 503  
Meir, Natalia, 344, 355  
Mennen, Ineke, 414, 420, 437, 438  
Menz, Astrid, 292  
Merino, Barbara J., 9  
Meyer, Roland, 424  
Meyerhoff, Miriam, 181, 189, 205  
Michaelis, Laura A., 395  
Miglio, Viola G., 423  
Mikhaylova, Anna, 233  
Milano, A, 121  
Miller, James Edward, 257  
Minor, Serge, 236, 238  
Mithun, Marianne, 259  
Mleinek, Ina, 424  
Moisl, Hermann, 82  
Molnár, Valéria, 145  
Montrul, Silvina, 3, 36, 44, 75, 121,  
158, 166, 167, 180, 220, 221,  
239, 327, 328, 414, 417, 449–  
451, 453, 471, 476, 528, 568  
Moosegard Hansen, Maj-Britt, 493  
Moreno-Fernandez, Francisco, 327,  
328  
Moser, Amalia, 230  
Müller, Anja, 119  
Müller, Hans Georg, 36, 155, 386, 387,  
431, 433  
Müller, Natascha, 167, 188  
Müller, Simone, 496  
Müller, Stefan, 330, 332  
Murtaza, Durguyul Metin, 309  
Musan, Renate, 380  
Muysken, Pieter, 4, 326, 349  
Myers-Scotton, Carol, 1  
Nagy, Naomi, 422, 438  
Nagy, Naomi G., 183, 186, 204, 396  
Nation, Paul, 329, 339  
Navarro, Samuel, 183  
Nedjalkov, Vladimir P., 255  
Neeleman, Ad, 206  
Németh, Gyula, 290, 296  
Nero, Shondel J., 122  
Niedrig, Heike, 122  
Ochs, Elinor, 56, 394, 395  
Ofner, Daniela, 123  
Ögel-Balaban, Hale, 254  
Ogren, Philip V., 90  
Oikonomou, Despina, 233  
Olsen-Reader, Vincent I., 122  
Onar Valk, Pelin, 260, 307  
Önnerfors, Olaf, 55  
Öpengin, Ergin, 296  
Oppermann, Sophia, 180  
Opsahl, Toril, 386, 431  
Ortega, Lourdes, 2  
Özbek, Nihal, 498  
Özkan, Nevzat, 289, 292, 293, 309  
Özsoy, A. Sumru, 380  
Özsoy, Onur, 3, 12, 13, 47, 61, 75, 201,  
204, 220, 223, 236–239, 260,  
278, 304, 307, 397, 458, 467,  
509, 528  
Öztürk, Balkız, 182, 187, 204, 310  
Özyürek, Aslı, 463  
Pandharipande, Rajeshwari V., 310  
Panitsa, Georgia, 221  
Papadopoulou, Despina, 230  
Paradis, Johanne, 155, 183, 450–453  
Park, Yong-Yae, 530  
Pascual y Cabo, Diego, 3, 75, 449  
Pashkova, Tatiana, 12, 13, 48, 151, 153,  
188, 260, 327, 334, 348, 384,

- 394, 415, 461, 465, 467, 468, 470  
Pavlenko, Aneta, 112, 450  
Pavlidou, Theodosia-Soula, 471, 528–530, 537  
Pekelis, Olga E., 185, 186  
Pennock-Speck, Barry, 329  
Pereltsvaig, Asya, 234  
Perez-Cortes, Silvia, 233  
Perlmutter, David M., 182  
Perniss, Pamela, 463  
Perpiñán, Silvia, 220  
Pescuma, Valentina N., 239  
Petersen, Jan Heegård, 329, 386, 417  
Petrou, Maria, 289, 292  
Petrov, Slav, 80  
Pfaff, Carol Wollman, 528, 530–532, 546, 569  
Pfeiffer, Martin, 497  
Philippaki-Warburton, Irene, 184, 204  
Pichler, Heike, 500  
Pickering, Martin, 152  
Plewnia, Albrecht, 142  
Pohle, Maria, 53  
Polinsky, Maria, 4, 10, 35, 36, 44, 75, 121, 158, 159, 166, 167, 180, 186, 189, 192, 203–205, 221, 240, 260, 278, 327, 328, 344, 355, 356, 380, 392, 413–415, 417, 434, 437, 449–452, 457, 460, 461, 477, 528, 568  
Pons Bordería, Salvador, 496  
Poplack, Shana, 415  
Prasad, Gail, 120  
Prince, Ellen F., 394  
Prior, Anat, 328  
Proske, Nadine, 144  
Pu, Ming-Ming, 181, 205  
Purkarthofer, Judith, 15, 16, 112, 113, 120, 131  
Pustejovsky, James, 331, 339  
Putnam, Michael T., 142, 148, 149, 157, 326  
Quasthoff, Uta, 56  
Queen, Robin, 451, 453, 457  
Quist, Pia, 386, 431  
Rabe, Maximilian M., 336  
Rabinovich, Ella, 328  
Raclaw, Joshua, 499, 530  
Rakhilina, Ekaterina, 327  
Rao, Rajiv, 423, 437  
Rathcke, Tamara, 424  
Rato, Anabela, 414  
Ravid, Dorit, 531  
Redeker, Gisela, 498  
Rehbein, Ines, 380, 386, 396, 431  
Revithiadou, Anthi, 223, 232  
Reznicek, Marc, 98  
Richards, Brian, 329, 335  
Riehl, Claudia M., 142, 326  
Riester, Arndt, 391, 457  
Rinke, Esther, 4, 184, 205  
Ripley, Brian D., 336  
Rizou, Vasiliki, 3, 12–14, 220, 222, 223, 225, 227, 228, 230–233, 235, 239–241, 279, 452  
Rizzi, Luigi, 206  
Roberts, Ian, 182, 385  
Robins, Robert H., 80  
Robinson, Lucy Danielle, 129  
Robles-Puente, Sergio, 423  
Roch, Maja, 329  
Rocker, Maike H., 386–388, 420, 432, 433  
Romaine, Suzanne, 1, 119

## *Name index*

- Romanova, Natalia, 233  
Romary, Laurent, 76, 90  
Rommel, Meike, 422  
Rosa, Jonathan, 122  
Rosenberg, Peter, 142  
Rosenthal, Dietmar E., 499  
Rothman, Jason, 2, 3, 9, 34, 36, 75, 180, 221, 222, 413, 449, 476, 495, 528  
Rothweiler, Monika, 167  
Roussou, Anna, 255, 385  
Roy, Albrun, 54  
Rymes, Betsy, 128, 129  
  
Sacks, Harvey, 471, 528, 529  
Sakel, Jeanette, 494, 519  
Salmons, Joseph C., 142, 326, 494  
Samimy, Keiko K., 2  
Sanchez Abchi, Veronica Soledad, 257, 260, 467  
Sarvasy, Hannah S., 254  
Sauer, Simon, 225  
Saunders, Georg, 115  
Schad, Daniel J., 336  
Schaeder, Burkhard, 80  
Schäfer, Lisa, 180, 182–184  
Schalowski, Sören, 386, 387, 431  
Schegloff, Emanuel A., 471, 528, 529  
Schiffrin, Deborah, 473, 493  
Schlauch, Tobias, 100  
Schleppegrell, Mary J., 257, 467  
Schmid, Helmut, 80, 190, 225  
Schmid, Monika, 142, 148, 326  
Schmidt, Thomas, 76, 193, 226, 261, 455, 533  
Schmitz, Katrin, 3, 180, 184, 205, 221  
Schoonbaert, Sofie, 450  
Schourup, Lawrence, 493, 495  
Schreiber, Laurentia, 310  
  
Schroeder, Christoph, 3, 12–14, 36, 112, 257, 260, 277, 279, 291, 304, 307, 423, 437, 509  
Schubö, Fabian, 458  
Schulz, Petra, 121, 166, 257, 260, 332, 333  
Schumann, Kathleen, 497  
Schwartz, Mila, 112  
Scontras, Gregory, 3, 4, 233, 241, 434, 450, 453, 476  
Seals, Corinne A., 122, 123  
Seaman, P. David, 227  
Sekerina, Irina A., 221, 234, 235, 427, 430, 434, 436  
Selkirk, Elisabeth, 457  
Selting, Margret, 395  
Serratrice, Ludovica, 393, 461, 462, 477  
Sevinç, Yeşim, 39  
Sewell, Alyson, 386, 387  
Shadrova, Anna, 11, 12, 33, 38, 42, 44, 60, 82, 84, 87, 327, 335, 455  
Shaer, Benjamin, 395  
Shagal, Ksenia, 255  
Shah, Sheena, 142  
Shin, Sarah J., 160  
Shirai, Yasuhiro, 230  
Shirk, Jennifer L., 129  
Shushurin, Philip, 185, 186, 204  
Siebenhaar, Beat, 51  
Silva-Corvalán, Carmen, 10, 471  
Silverstein, Michael, 181  
Simon-Cereijido, Gabriela, 335  
Şimşek, Yazgül, 474  
Sinclair, John, 35, 41  
Sirotinina, Olga Borisovna, 381, 389, 391  
Sitaridou, Ioanna, 185, 204  
Skopeteas, Stavros, 292

- Slabakova, Roumyana, 233  
Slioussar, Natalija, 380, 381, 389, 391  
Slobin, Dan, 331, 345  
Slobin, Dan I., 258  
Sluckin, Benjamin Lowell, 155, 385, 394–396  
So'o, Ainslie, 530  
Sorace, Antonella, 13, 180, 189, 203, 387, 393, 414, 437, 438, 453, 460, 461  
Speyer, Augustin, 145  
Spyropoulos, Vassilios, 223, 232  
Stachowiak, Herbert, 71  
Staley, Larssyn, 419  
Stathi, Katerina, 432  
Statistisches Bundesamt, 424  
Stavropoulos, Dimitri N., 498  
Steien, Guri Bordal, 120  
Stein, Heidi, 296  
Steinbach, Markus, 145  
Stephany, Ursula, 221  
Stiebels, Barbara, 332  
Stolberg, Doris, 142, 146, 148, 326  
Stoop, Wessel, 164  
Stowell, Tim, 206  
Sulçevsi, İsa, 289, 309  
Švedova, Natalija J., 389  
Svendsen, Bente Ailin, 129  
Svetozarova, Natalia, 424  
Swain, Merrill, 234  
Szendrői, Kriszta, 206  
Tagliamonte, Sali A., 396, 422, 468  
Taleghani-Nikazm, Carmen, 530  
Talmy, Leonard, 331, 345  
Tausch, Johanna, 16  
te Velde, John R., 155, 432  
Tenny, Carol, 331  
Teza, Emilio, 290  
Thielmann, Winfried, 388  
Thoma, Dieter, 123  
Thurmair, Maria, 388, 497  
Tolchinsky, Liliana, 472, 528–532, 546, 569  
Tomasello, Michael, 235, 328  
Topaj, Nathalie, 535  
Torregrossa, Jacopo, 122, 185, 204  
Tottie, Gunnell, 419  
Tracy, Rosemarie, 112, 113, 119–121, 127, 128, 131, 142, 146, 148–150, 159, 165–167, 188, 203, 257, 326, 332, 333, 350, 386, 431, 450  
Traugott, Elizabeth C., 432, 495  
Treffers-Daller, Jeanine, 3, 121, 260, 307, 329, 335, 476  
Triantafyllidis, Manolis, 498  
Trotzke, Andreas, 474  
Trueswell, John C., 427, 430, 434, 436  
Trutkowski, Ewa, 180, 183, 184  
Tsampoukas, Alexandros, 499, 513  
Tsehaye, Wintai, 3, 12–16, 61, 119, 151, 154, 156, 157, 260, 279, 326, 334, 349, 354, 452, 467, 468, 470, 476, 528  
Tsimpli, Ianthi-Maria, 119, 158, 180, 184, 189, 203, 255, 453  
Tufan, Şirin, 292  
Turan, Dilek, 307  
Turan, Feride, 499  
Tzartzanos, Achilleas, 498  
Usanova, Irina, 389, 390  
Vainikka, Anne, 148  
Valdés, Guadelupe, 180, 221  
van der Auwera, Johan, 287  
Van Gijsel, Sofie, 329, 335, 338, 355

## *Name index*

- Van Hout, Angeliek, 333  
van Rijswijk, Remy, 423, 451, 453,  
457  
Vanpaemel, Wolf, 226  
Vanrell, Maria del Mar, 414  
Varol, Orhan, 296  
Vasilyeva, Marina, 155  
Venables, William N., 336  
Vendler, Zeno, 331, 346  
Ventureyra, Valérie A.G, 476  
Verhoeven, Ludo, 531  
Verschik, Anna, 112  
Villavicencio, Aline, 391  
Vinckel-Roisin, Hélène, 144, 145, 155  
Voet Cornelli, Barbara, 121, 127  
Vogels, Jorrig, 461  
Vollmer, Maria C., 181, 205  
  
Waldvogel, Joan, 567, 569  
Waletzky, Joshua, 257, 258  
Wali, Kashi, 310  
Walkden, George, 155  
Wandrei, Alina, 415  
Ward, Gregory, 380  
Wei, Li, 2, 221, 234  
Weil, Gotthold, 289  
Weinert, Regina, 257  
Westbury, Josh, 394, 395  
Westphal-Fitch, Gesche, 155, 158  
White, Hayden, 71  
Whorf, Benjamin Lee, 287  
Wichmann, Anne, 432  
Wickham, Hadley, 226, 336  
Wiese, Heike, 2, 3, 8, 10, 12, 14, 15, 33–  
36, 38, 39, 44, 51, 53, 59, 61,  
70, 75, 84, 89, 102, 112, 113,  
122, 128, 131, 142, 144, 151,  
152, 155, 166, 180, 190, 192,  
205, 221, 225, 229, 230, 235,  
259–261, 279, 290, 327, 333,  
334, 355, 380, 382, 385–387,  
389, 392, 396, 401, 416, 420,  
423, 431, 433, 438, 454, 455,  
460, 464, 476, 494–497, 499,  
518, 528, 529, 531–533, 540,  
549, 567, 568  
Winford, Donald, 260  
Winkle, Claudia, 384, 470  
Winkler, Julia, 145  
Winkler, Susanne, 145  
Winter, Bodo, 189, 207  
Witt, James, 332  
Wöllstein, Angelika, 143, 144  
Wörfel, Till, 122, 123  
Wörner, Kai, 76, 193, 226, 261, 455,  
533  
Wunderlich, Dieter, 332  
  
Yager, Lisa, 158  
Yagmur, Kutlay, 260  
Yang, Charles, 167  
Yao, Yao, 414  
Yildiz, Cemal, 122  
Yilmaz, Erkan, 498  
Yilmaz-Wörfel, Seda, 260  
Yoshioka, Keiko, 461  
Young-Scholten, Martha, 148  
Yu, Guoxing, 329, 335, 338, 355  
  
Zappen-Thomson, Marianne, 142  
Zeldes, Amir, 11, 76, 90, 193, 455, 533  
Zeller, Jochen, 332  
Zellers, Margaret, 150, 350, 419, 423  
Zenker, Fred, 335, 456  
Zerbian, Sabine, 13, 14, 90, 388, 394,  
397, 417, 421, 423, 427, 428,  
430, 434, 436, 437, 457–459  
Zhang, Xinye, 206

- Zhou, Jiangling, 204  
Zhu, Hao, 336  
Zifonun, Gisela, 157  
Zimmer, Christian, 38, 142, 158, 326,  
    387  
Zimmer, Karin, 116  
Zimmerman, Don H., 531  
Zipser, Florian, 76, 90  
Ziv, Yael, 432  
Zuban, Yulia, 3, 388–393, 400, 417,  
    421–425, 430  
Zürn, Nadine, 12–14, 150, 153, 159





# Linguistic dynamics in heritage speakers

This collective volume investigates linguistic dynamics in language contact, focusing on heritage speakers. The chapters provide new insights into the role of speaker repertoires and the distinction between contact-induced change and language-internal variation by reporting on corpus-linguistic studies across different communicative situations in heritage and majority languages. Conducted in the context of the DFG Research Unit “Emerging Grammars in Language Contact Situations” (FOR 2537), the studies focus on bilingual adolescent and adult speakers of German, Greek, Russian and Turkish as heritage languages, and of English and German as majority languages, and on monolingually raised adolescent and adult speakers of all five languages. Crucially, they are not restricted to standard language, but target broader speaker repertoires that cover informal as well as formal settings in both spoken and written modes. The contributions are united by their positive perspective on language contact and multilingual speakers, a comparative approach across several heritage and majority languages, and a shared methodology that captures variation within repertoires for both heritage speakers and monolinguals. The chapters take various theoretical standpoints, highlighting different facets of the data as well as its potential for enhancing our understanding of language contact and language variation.