

# Contemporary research in minoritized and diaspora languages of Europe

Edited by

Matt Coler

Andrew Nevins

Contact and Multilingualism 6



## Contact and Multilingualism

Editors: Isabelle Léglise (CNRS SeDyL), Stefano Manfredi (CNRS SeDyL)

In this series:

1. Lucas, Christopher & Stefano Manfredi (eds.). Arabic and contact-induced change.
2. Pinto, Jorge & Nélia Alexandre (eds.). Multilingualism and third language acquisition:  
Learning and teaching trends.
3. Hakimov, Nikolay. Explaining Russian-German code-mixing: A usage-based approach.
4. Sánchez Moreano, Santiago & Élodie Blestel (eds.). Prácticas lingüísticas heterogéneas:  
Nuevas perspectivas para el estudio del español en contacto con lenguas amerindias.
5. Migge, Bettina & Shelome Gooden(eds.). Social and structural aspects of language contact  
and change.
6. Coler, Matt & Andrew Nevins (eds.). Contemporary research in minoritized and diaspora  
languages of Europe.

ISSN (print): 2700-8541

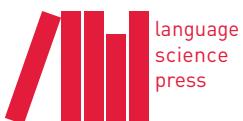
ISSN (electronic): 2700-855X

# Contemporary research in minoritized and diaspora languages of Europe

Edited by

Matt Coler

Andrew Nevins



Matt Coler & Andrew Nevins (eds.). 2022. *Contemporary research in minoritized and diaspora languages of Europe* (Contact and Multilingualism 6). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/332>

© 2022, the authors

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

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

ISBN: 978-3-96110-404-8 (Digital)

978-3-98554-062-4 (Hardcover)

ISSN (print): 2700-8541

ISSN (electronic): 2700-855X

DOI: 10.5281/zenodo.7442323

Source code available from [www.github.com/langsci/332](http://www.github.com/langsci/332)

Collaborative reading: [paperhive.org/documents/remote?type=langsci&id=332](http://paperhive.org/documents/remote?type=langsci&id=332)

Cover and concept of design: Ulrike Harbort

Typesetting: Matt Coler, Yanru Lu, Sebastian Nordhoff, Sjors Weggeman

Proofreading: Agnes Kim, Amy Amoakuh, Andreas Hözl, Christopher

Straughn, Elliott Pearl, Emma Vanden Wyngaerd, Felix Kopecky, Francesco

Screti, Janina Rado, Jeroen van de Weijer, Jorina Fenner, Konstantinos

Sampanis, Lea Schaefer, Ludger Paschen, Jean Nitzke, Sebastian Nordhoff

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: X<sub>\mathbb{E}</sub>T<sub>\mathbb{E}</sub>X

Language Science Press

xHain

Grünberger Str. 16

10243 Berlin, Germany

[langsci-press.org](http://langsci-press.org)

Storage and cataloguing done by FU Berlin



# Contents

Acknowledgments	iii
<b>1 Introduction</b> Andrew Nevins & Matt Coler	1
<b>2 Documenting Italo-Romance minority languages in the Americas: Problems and tentative solutions</b> Luigi Andriani, Jan Casalicchio, Francesco Ciconte, Roberta D'Alessandro, Alberto Frasson, Brechje van Osch, Luana Sorgini & Silvia Terenghi	9
<b>3 Spanish-Nahuatl bilingualism in Indigenous communities in Mexico: Variation in language proficiency and use</b> Justyna Olko, Szymon Gruda, Joanna Maryniak, Elwira Dexter-Sobkowiak, Humberto Iglesias Tepec, Eduardo de la Cruz & Beatriz Cuahutle Bautista	57
<b>4 Trilingual modality: Towards an analysis of mood and modality in Aymara, Quechua and Castellano Andino as a joint systematic concept</b> Philipp Dankel, Mario Soto Rodríguez, Matt Coler & Edwin Banegas-Flores	79
<b>5 What is the role of the addressee in speakers' production? Examples from the Griko- and Greko-speaking communities</b> Manuela Pellegrino & Maria Olimpia Squillaci	121
<b>6 Innovations in the Contemporary Hasidic Yiddish pronominal system</b> Zoë Belk, Lily Kahn, Kriszta Eszter Szendrői & Sonya Yampolskaya	143
<b>7 Validity of crowd-sourced minority language data: Observing variation patterns in the Stimmen recordings</b> Nanna Hilton	189

*Contents*

<b>8 Complexity of endangered minority languages: The sound system of Wymysioeryś</b> Alexander Andrason	<b>213</b>
<b>9 Language variation, language myths, and language ideology as constructive elements of the Wymysioeryś ethnolinguistic identity</b> Tomasz Wicherkiewicz	<b>261</b>
<b>10 Evaluating linguistic variation in light of sparse data in the case of Sorbian</b> Eduard Werner	<b>281</b>
<b>11 Modelling accommodation and dialect convergence formally: Loss of the infinitival prefix <i>tau</i> ‘to’ in Brazilian Pomeranian</b> Gertjan Postma	<b>303</b>
<b>12 Using data of Zeelandic Flemish in Espírito Santo, Brazil for historical reconstruction</b> Kathy Rys & Elizana Schaffel Bremenkamp	<b>341</b>
<b>Index</b>	<b>385</b>

# Acknowledgments

The impulse for *Contemporary research in minority and diaspora languages of Europe* arose from a panel entitled “European minority and diaspora languages”, organized by the two editors. The session was part of the 10th International Conference on Language Variation in Europe (ICLAVE-10), held in Leeuwarden, the Netherlands in 2019. Contributors to that panel offered presentations outlining the wide variety of linguistic variation in diaspora and minoritized languages within and outside of Europe. They presented a broad array of content which touched upon phonetic, phonological, morphological, syntactic and even semantic variation. This was a spirited and engaging event that extended well past the allotted time – but things really picked up in the post conference dinner which inspired several parallel discussions on diaspora languages that lasted late into the night. Those exchanges enabled us to distill some overarching themes across different contexts: Phenomena such as diachronic and synchronic tendencies in variation alongside sociolinguistic and ethnolinguistic factors that are connected with contact phenomena, among others. In the early hours of the following morning, we converged on the idea of this book. Although not every contributor to the book was present at the ICLAVE-10 event, nor is every presenter in that panel included in this book, it was that discussion which served as the impulse for this book. Thank you to all those who stayed up late mulling over language change and contact, and thank you also to all those who participated in the panel and to the conference organizers of ICLAVE-10 at the Fryske Akademy.

It should go without saying that we are grateful to all the contributors to this book who shared their expertise in their own chapter and in the book overall. Collaborating on a collection of articles like this one with so many authors is no easy endeavor – especially when at any given moment, many contributors are doing fieldwork or on sabbatical (or navigating the new academic landscape carved by the coronavirus pandemic). Therefore, we are grateful for the sustained collaboration of all authors who prioritized the writing, revisions, and final edits of their respective chapter. Likewise, we are very appreciative of the anonymous reviewers and copy editors who helped ensure the scientific quality and consistency in style and form. Their input helped transform a collection of essays into a consistent and coherent book. We are also very grateful to Sebastian Nordhoff

### *Acknowledgments*

at Language Science Press who answered innumerable questions about the production process. Finally, a big thank you to Sjors Weggeman, student assistant at the University of Groningen, who provided critical support with the final preparation of the manuscript.

The preparation and publication of this book was funded by the EU Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 778384. It was developed within the project *Minority Languages, Major Opportunities. Collaborative Research, Community Engagement and Innovative Educational Tools* (COLING).

# Chapter 1

## Introduction

Andrew Nevins<sup>a</sup> & Matt Coler<sup>b</sup>

<sup>a</sup>University College London <sup>b</sup>University of Groningen – Campus Fryslân

The introduction is split into three parts. In the first, we provide some background for the book, provide key definitions and reflections on minoritized languages and diaspora languages, and identify overarching themes in the book. In the second, we provide a summary overview of each of the 12 chapters, mentioning the languages investigated, the methods used, and the main findings. Finally, in the third part we reflect on the scope of the book itself and conclude with remarks on multilingualism and minoritization, positing that monolingualism should not be considered the default or status quo – quite the opposite, in fact. The research here suggests that fighting against monolingualism as an increasing trend involves not just scientific work, but also social activism.

### 1 Setting the Stage

These chapters emerged from presentations given at a panel entitled “Variation in European minority and diaspora languages” held at the 10<sup>th</sup> International Conference on Language Variation in Europe in Leeuwarden, Fryslân on June 26<sup>th</sup> 2019.

Critical to all these contributions is the concept of a diaspora and minoritized languages. For the purposes of this book, we consider diaspora languages to refer to languages spoken by people who have resettled in an area outside of their original linguistic community. Speakers of diaspora languages may have diverse origins and might come from different communities, social strata, and even nations.

Comparatively, a minority (or, more properly, *minoritized* language) is a language variety, or a cluster of varieties, that is historically spoken in a particular



region where another language— usually an official majority language—is predominantly spoken.

It is worthwhile to dive deeper into the minority language vs. minoritized language dichotomy, insofar as it is a theme which shows up frequently in the contributions of this book. The term ‘minority language’ can be said to suggest an inherent and static quality of the language itself, and in that sense, it obscures the fact that languages become minoritized as a direct outcome of actions and policies. That is, “minoritization recognizes that systemic inequalities, oppression, and marginalization place individuals into ‘minority’ status rather than their own characteristics”(Sotto-Santiago 2019). Accordingly, many within the field of linguistics promote the term “minoritized language” instead of “minority language” (see Nevins 2022: Ch. 1). Although the latter term is somewhat entrenched in the field, the former has a history going back at least three decades – see Py & Jeanneret 1989 and a full entry in Wikipedia.

This provides an opportune point to reconsider diaspora languages. The common sense definition, that is, “a language spoken by people who have resettled in an area outside of their linguistic community” seems to be lacking. Let us consider two thought experiments. In the first, we can reflect on whether Spanish speakers in Argentina are speakers of a diaspora language in the same way that those in New York are. Next, consider where English is spoken as a diaspora language. Is it accurate to suggest that English is a diaspora language in India, Singapore and Hong Kong? Upon consideration, there seems to be a correlation between colonialism and diaspora status, at least sometimes. That is, people from the Global South are more clearly diaspora whereas less so for the Global North. On the other hand, Yiddish and Pomeranian, spoken in South America but no longer with thriving linguistic communities in their place of origin are clearly diaspora communities as well. By pointing to such contrasts, we hope to raise the issue of the complexity at hand.

There is clearly some overlap between diaspora and minority languages. Many of the languages in our book can be described as both diaspora and minoritized languages (including Pomeranian, Wymysioeryś and Yiddish)<sup>1</sup>. Some others are only diaspora languages, but not, strictly speaking, minoritized languages (Castellano Andino – spoken amongst Amerindian communities, Italo-Romance and Dalmatian varieties spoken in the Americas, and so forth); yet others are minoritized, but not diaspora languages (including Sorbian, Aymara, Quechua, and Nahuatl). These issues entail a gamut of political repercussions. Consider

---

<sup>1</sup>Importantly, in some cases, such as that for Greko and Griko, speakers do not consider themselves as diaspora speakers; see for example Pellegrino (2021).

how the “European Charter for minority or regional languages” of the Council of Europe defines minority and regional languages as those languages which are traditionally used within a given territory of a state by inhabitants of that state who form a group numerically smaller than the rest of the state’s population and which are different from the official language(s) of that state. The charter also protects so-called “non-territorial languages” which are understood as languages spoken by nationals of a particular state, but these languages are distinct from the language(s) used by the rest of the population of the state and which cannot be identified with a particular region of a given state – such languages include Ladino, Romani, and Yiddish, for example. Notably, dialects and migrant languages are not included in the charter; see Woehrling (2005) for further discussion.

Aside from the geographic array, the linguistic variation explored attests to phenomena at many levels. This includes phonetic/phonological variation, relating to historical sound changes in Yiddish and Sorbian morphological and morphosyntactic variation attested in Israeli and US American variants of Yiddish, contact phenomena influencing the expression of grammatical mood in Spanish varieties in contact with Quechua and Aymara, and an array of morphological phenomena in Italo-Romance and -Dalmatian varieties spoken in the Americas. The contribution on Pomeranian is dedicated to developing an analysis of the syntactic structure of this language to account for language change. The contributions further extend beyond specific grammatical phenomena, and additionally touch upon anthropological issues relating to verbal art for Sorbian, sociolinguistic and ethnolinguistic identity (for Wymysiöeryś, and Greko/Griko), and phenomena relating to language contact between minority and majority languages in a new sociological setting at different grammatical levels, and whether it results in complexification (as for Wymysiöeryś in contact with varieties of German and Polish), structural reduction (as for Zeelandic-Flemish in contact with Brazilian Portuguese, or other kinds of change as in Spanish in contact with Nahuatl, Aymara, and Quechua, among the many other examples in these contributions).

## 2 Overview of the Contributions

We provide here an overview of the content and themes covered in Chapters 2 through 12 of this book, highlighting the range of diverse language contact scenarios covered throughout.

In Chapter 2, ‘Documenting Italo-Romance minority languages in the Americas: Problems and tentative solutions’, Luigi Andriani, Jan Casalicchio, Francesco

Ciccone, Roberta D'Alessandro, Alberto Frasson, Brechje van Osch, Luana Sorogini, and Silvia Terenghi look at differential object marking, deixis and demonstratives, and subject clitics and null subjects in seven heritage Italo-Romance varieties (Piedmontese, Venetan, Tuscan, Abruzzese, Neapolitan, Salentino, and Sicilian). They describe the process of preparation and implementation of a data collection enterprise targeting Italo-Romance emigrant languages in North and South America, many of which had never been documented, and which, with the exception of the northern Italian-speaking community, are close to extinction. Their project aims to understand language change in contact. Their article describes the steps they took in assessing the speakers' proficiency, designing and running syntactic questionnaires and picture-sentence matching tasks, and general issues concerning experimental design and statistics.

In Chapter 3, 'Spanish-Nahuatl bilingualism in Indigenous communities in Mexico: Variation in language proficiency and use', Justyna Olko, Szymon Gruda, Joanna Maryniak, Elwira Dexter-Sobkowiak, Humberto Iglesias Tepec, Eduardo de la Cruz and Beatriz Cuahutle Bautista take a historical perspective on bilingualism in Spanish and Nahuatl from 1519 until the present day. They discuss the results of proficiency assessment in both languages, performed with the participation of members of selected Nahua communities. Their work reveals different degrees of assimilation to Mexican identity and shift to Spanish, most salient in more urbanized and less peripheral regions. The authors conclude that factors such as power differentials, economic marginalization, sociopolitical pressures, culture change, ethnic prejudice and discriminatory language policies lead to contemporary Spanish-Indigenous bilingualism at the community level being highly unstable. Using the term 'unstable bilingualism', they suggest that the situation of parallel acquisition and use of Nahuatl and Spanish as well as diminishing and varying proficiency in the heritage language will lead eventually to language shift. Depending on the region, this may occur as quickly as within two to four generations.

In Chapter 4, 'Trilingual modality: Towards an analysis of mood and modality in Aymara, Quechua and Castellano Andino as a joint systematic concept', Philipp Dankel, Mario Soto Rodríguez, Matt Coler and Edwin Banegas-Flores examine how indigenous minoritized languages impact majority European ones. They do this by considering the case of Quechua and Aymara, on the one side, and Castellano Andino (CA) on the other. Their analysis demonstrates that regional varieties of CA reflect Aymara and Quechua mood, even in the speech of those who do not speak either indigenous language. The authors emphasize the complex nature and multiple causality of contact induced change which allows

for potentialities of how minoritized languages can indeed sometimes impact majority languages.

In Chapter 5, ‘What is the role of the addressee in speakers’ production? Examples from the Griko- and Greko-speaking communities’, Manuela Pellegrino and Maria Olimpia Squillaci focus on the two endangered Italo-Greek varieties, Griko and Greko, spoken, respectively, in Salento (Puglia) and Calabria communities in Southern Italy. The authors examine speaker-addressee dynamics, how these affect language use and may potentially lead to ‘temporary variation’, and how the addressee’s linguistic competence, age, and shared linguistic repertoire with the speaker may lead to style-shift in speakers’ production. They then consider how these factors contribute to the emergence of puristic attitudes which may even inhibit the use of Griko and Greko. The authors show how widespread resistance to and monitoring of language variation and change tend to undermine efforts to maintain or revitalise Griko and Greko. This highlights multiple, entangled power struggles embedded in their current revival.

In Chapter 6, ‘Innovations in the Contemporary Hasidic Yiddish pronominal system’, Zoë Belk, Lily Kahn, Kriszta Eszter Szendrői and Sonya Yampolskaya present a study involving 29 native Contemporary Hasidic Yiddish speakers, and demonstrate that significant changes have occurred in the personal pronoun, possessive, and demonstrative systems. While the personal pronoun system has undergone significant levelling in terms of case and gender marking, at the same time a new demonstrative pronoun has emerged which exhibits a novel case distinction. They argue that these innovative features are not determined directly by contact with the dominant co-territorial languages, but rather are internal developments which bear witness to the linguistic vibrancy of Contemporary Hasidic Yiddish.

In Chapter 7, ‘Validity of crowd-sourced minority language data: Observing variation patterns in the Stimmen recordings’, Nanna Hilton considers the usability of crowd sourced minority language data for research. She uses speech recordings and reported dialect knowledge collected with a smartphone application for Frisian, focusing on three phonological variables in Frisian speech. The author considers how minority language communities offer a welcome chance for variationist sociolinguistics to revisit principles of linguistic variation and change. It is often assumed that Frisian is converging towards Dutch on all linguistic levels. However, this assumption is based almost entirely on anecdotal evidence. Very few empirical studies of speech variation in Frisian exist. A way to conduct studies of sound change on a larger scale would be to use crowd-sourced speech data. To this end Hilton considers the usability of data from the

Stimmen application. Stimmen has a picture naming task comprised of 87 images of everyday objects, and a gamified task that provides an estimate of where people hail from within the Province of Fryslân. The author concludes that the data is of high quality and that it can be used for investigating sound change. That said, one must consider whether crowd sourced data include contributions from so-called “new speakers” of minority languages, and whether one has a representative sample of all age groups in such remote studies.

In Chapter 8, ‘Complexity of endangered minority languages: The sound system of Wymysiöeryś’, Alexander Andrason demonstrates that Wymysiöeryś – a severely endangered minority Germanic language – exhibits remarkable complexity despite its moribund status. By analyzing twelve phonetic/phonological properties, the author concludes that the complexity of Wymysiöeryś is greater, both locally and globally, than that of two control languages: Middle High German and Modern Standard German. In most cases, the surplus of complexity attested is attributed to contact with the dominant and aggressive language, i.e. Polish. This confirms the view of language contact as not only having simplifying effects on languages, but also as contributing to their complexification – even in the situation of seemingly imminent language death.

In Chapter 9, Tomasz Wicherkiewicz considers how the Wymysiöeryś language, spoken in Wilamowice, has been frequently classified as a colonial variety of East Central German. He reflects on how such ethnotheories of provenance, including folk linguistic evidence and myths, referred to various Germanic countries as places of origin of the first settlers. However, the microlect of Wilamowice has certainly undergone interactions of various types and intensities with Polish (and its varieties) and standard High German. There is evidence of such contacts, shift, and hybridization in all subsystems of the microlanguage. Wicherkiewicz analyzes this through an approach based on perceptual dialectology and an ethnoscience perspective of language variation.

In Chapter 10, ‘Evaluating linguistic variation in light of sparse data in the case of Sorbian’, Eduard Werner examines one of the oldest Sorbian monuments by applying knowledge on neighboring Germanic and Celtic literature. From the linguistic side, the results of the comparison lend greater insights into historical sound changes in Sorbian. The author shows how historical sound changes help to unearth elements of verbal art. This, in turn, facilitates the possibility of more accurately dating the historical sound changes through their effects on alliterations.

In Chapter 11, ‘Modeling accommodation and dialect convergence formally: Loss of the infinitival prefix *tau* ‘to’ in Brazilian Pomeranian’, Gertjan Postma re-evaluates a well-known, but often ignored mechanism and outcome: retreat

to default settings, the rise of the unmarked, that is whenever the result of the change is not a sum or subset of the input forms, but an innovative pattern. While the original Pomeranian dialects in Europe had a considerable variation in this particular domain, Pomeranian in Brazil has converged to a remarkably uniform new construction, which was not present in Pomerania in the days of emigration. This configuration is reanalyzed as an overt movement relation of T to C, which is the default option in natural language. There are language-internal arguments that the new construction is a result of dialect-convergence to the default setting of the parameters involved. However, when we take the external occurrence rates into account, the data indicate that the similarity in this respect between Pomeranian and Brazilian Portuguese might be analyzed as accommodation of Brazilian Pomeranian to the dominant language.

In Chapter 12, ‘Using data of Zeelandic Flemish in Espírito Santo, Brazil for historical reconstruction’, Kathy Rys and Elizana Schaffel Bremenkamp focus on the case of Zeelandic Flemish in Espírito Santo, an obsolescent language variety spoken by about twenty descendants of Dutch immigrants in the 19<sup>th</sup> century. These speakers are descendants of Dutch immigrants, who left Zeeland in 1858–1862, but have faced deprivation and difficulties in adaptation and integration into Brazilian society, with their language threatened by the majority language Brazilian Portuguese and by another heritage language; namely that of the Pomeranian immigrants who arrived in Espírito Santo at the same time. The speech of rusty speakers can be used to reconstruct the original immigrant language. The authors perform a historical reconstruction of the old Zeelandic Flemish dialect as spoken in the days of emigration, with respect to three linguistic cases: (1) deletion of /l/ in codas and coda clusters, (2) subject doubling in inversion contexts and (3) inflected polarity markers *yes* and *no*. Their findings demonstrate the historical value of transplanted dialects or speech island varieties. Moreover, a comparison of their findings with historical data demonstrates that reliance on rusty speaker data alone may sometimes lead to incorrect conclusions. Instead, such data patterns can also be considered from the perspective of language contact.

### 3 Conclusion

Contact scenarios and minoritization often involve unstable patterns of bilingualism (recall, for example, Chapter 3, which speaks of power differentials, economic marginalization, sociopolitical pressures, culture change, ethnic prejudice and discriminatory language policies). This leads us to conclude that on the basis of the research presented here – through investigations of contact effects on

complexity, on grammatical shift, on power struggles in language revival, and on new methodologies for engaging with communities of diaspora and minoritized languages – it would be fair to accept the claim that monolingualism can be viewed as a destructive force (as Carlos Fuentes once pointed out, "a curable disease") perhaps akin to monoculture farming, which threatens the diversity inherent to a healthy linguistic ecology. The opportunity to consolidate so many different kinds of research happening around the world on questions around contact scenarios and minoritization is scientifically compelling. It is also a testament to the extent to which contributors care about these topics, having devoted untold hours of their time and energy to work towards a careful understanding of the past, present, and future of their linguistic ecologies. Thus, in a real sense, the contributors to this volume offer piece by piece new sources of validation and support for language diversity. The speakers of the languages represented here and others in minoritized or diaspora communities, by continuing linguistic traditions in the face of discrimination are, in a very real sense, social activists.

## References

- Nevins, Andrew. 2022. *When minoritized languages change linguistic theory*. Cambridge: Cambridge University Press.
- Pellegrino, Manuela. 2021. *Greek language, Italian landscape: Griko and the re-storying of a linguistic minority*. Washington DC: Center for Hellenic Studies. <https://chs.harvard.edu/book/greek-language-italian-landscape-griko-and-the-re-storying-of-a-linguistic-minority/>.
- Py, Bernard & René Jeanneret. 1989. *Minorisation linguistique et interaction*. Genève: Droz.
- Sotto-Santiago, Sylk. 2019. Time to reconsider the word minority in academic medicine. *Journal of Best Practices in Health Professions Diversity* 12(1). 72–78.
- Woehrling, Jean-Marie. 2005. *The European Charter for regional or minority languages: A critical commentary*. Council of Europe.

# Chapter 2

## Documenting Italo-Romance minority languages in the Americas: Problems and tentative solutions

Luigi Andriani<sup>a</sup>, Jan Casalicchio<sup>b</sup>, Francesco Ciconte<sup>c</sup>, Roberta D'Alessandro<sup>a</sup>, Alberto Frasson<sup>a</sup>, Brechje van Osch<sup>d</sup>, Luana Sorgini<sup>a</sup> & Silvia Terenghi<sup>a</sup>

<sup>a</sup>Utrecht University, UiL-OTS <sup>b</sup>University of Palermo <sup>c</sup>University of Insubria

<sup>d</sup>University of Tromsø

This article describes the process of preparation and implementation of a data collection enterprise targeting Italo-Romance emigrant languages in North and South America. This data collection is part of the ERC Microcontact project, which aims to understand language change in contact by examining the language of Italian communities in the Americas.

### 1 Introduction

This article describes the process of preparation and implementation of a data collection enterprise targeting Italo-Romance emigrant languages in North and South America. This data collection is part of the ERC Microcontact project, which aims to understand language change in contact by examining the language of Italian communities in the Americas (<https://microcontact.sites.uu.nl>).

The speakers involved in our study are first-generation Italians (so-called *émigrés*; henceforth: “G1”), most of whom emigrated to North and South America between the 1940s and the 1960s, and second- and third-generation speakers (heritage speakers, “HS”). The population of Italian emigrants is close to ideal for a study on language contact, because most of them were tendentially monolingual

Luigi Andriani, Jan Casalicchio, Francesco Ciconte, Roberta D'Alessandro, Alberto Frasson, Brechje van Osch, Luana Sorgini & Silvia Terenghi. 2022. Documenting Italo-Romance minority languages in the Americas: Problems and tentative solutions. In Matt Coler & Andrew Nevins (eds.), *Contemporary research in minoritized and diaspora languages of Europe*, 9–56. Berlin: Language Science Press. DOI: 10.5281/zenodo.4902965



speakers of an Italo-Romance dialect when they arrived in the Americas. Italian was not widely spoken in Italy until the 1960s, and hence they were mostly monolingual speakers of these varieties when they left Italy. When they arrived in the Americas, they entered into sudden intensive contact with other Romance languages: we are focusing here on French (in Quebec and in later fieldwork in Belgium<sup>1</sup>), Spanish (in Argentina), and Portuguese (in Brazil). We also consider these varieties in contact with Italian (in Italy), bearing in mind that this contact is very different from that found in the Americas and Belgium, first and foremost because the contact has been more intense, and because the communities speaking these varieties are generally larger in Italy. Finally, we also investigate Italo-Romance speakers in contact with English in the United States as control group.

The general aim of the project is to draw a predictive analysis of language change in contact by looking at multiple microcontact situations. Contact is investigated here not on a one language-to-one language basis but on a many-to-many basis: in this way, each phenomenon can be checked against multiple, minimally-varying, equivalent phenomena in the contact languages. By the end of the project, we hope to have identified the structural triggers for language change. Furthermore, we also wish to compare change in contact with diachronic change, to ascertain whether they follow similar paths, as is often claimed in the literature. The project follows the evolution of these contact situations by focusing on three language phenomena in seven Italo-Romance varieties. The phenomena that we selected are: differential object marking (“DOM”), deixis and demonstratives, and subject clitics (“SCLs”) / null subjects. Other language features, such as topicalization, are also taken into account to a lesser extent. These phenomena have been selected because they are well documented for the languages at issue and their diachronic evolution can be tracked rather straightforwardly. For each of these phenomena we checked whether they are preserved in the various contact situations, and in which syntactic contexts. The preliminary results are being published in a number of papers (Andriani et al. 2020, Casalicchio & Frasson 2019, Sorgini 2020, Terenghi 2020, D’Alessandro 2021, Frasson In press, Frasson et al. In press, and many other papers in preparation).

The varieties that were originally selected for investigation are Piedmontese, Venetan, Tuscan (Florentine and Sienese), Coastal/Eastern Abruzzese, Neapolitan, Salentino, and Sicilian. They are displayed in the map in Figure 1 (in regular font).<sup>2</sup>

---

<sup>1</sup>See below and Section 3 for details on why we additionally targeted Italian emigrants in the French-speaking part of Belgium.

<sup>2</sup>The map is retrieved from [https://en.wikipedia.org/wiki/Languages\\_of\\_Italy](https://en.wikipedia.org/wiki/Languages_of_Italy).

## 2 Documenting Italo-Romance minority languages in the Americas



Figure 1: Languages of Italy and selected varieties

[https://commons.wikimedia.org/wiki/File:Linguistic\\_map\\_of\\_Italy\\_-\\_Legend.svg](https://commons.wikimedia.org/wiki/File:Linguistic_map_of_Italy_-_Legend.svg) CC-BY-SA-4.0 Mikima

These varieties were chosen for several reasons: they maximally instantiate the variation recorded for our target phenomena across Italo-Romance and are the most spoken by Italian emigrants in the Americas. Moreover, they all have a long literary tradition, with the exception of Abruzzese, which was selected because of the wide documentation on the language available to the PI. This documentation was crucial for us to be able to compare the diachronic evolution of the phenomena that we are considering with their change in contact. While Italian and Italo-Romance languages have been in extensive contact for the last 70 years, not many people could speak Italian at the beginning of the 20<sup>th</sup> century.

The languages selected did not all prove optimal. In particular, we did not manage to find Tuscan, Salentino or Neapolitan speakers. Instead, a very large community of Calabrian and Friulian speakers was identified during fieldwork. Figure 1, in italicized font, shows where those varieties are spoken in Italy. In order to have a large and consistent set of data, it was decided to exclude the varieties with very few speakers and introduce Friulian and Calabrian instead.

One additional change to the original plan was in the locations in which we carried out our fieldwork. Recall that the locations and languages that were originally selected were Argentina/Argentinian Spanish, Brazil/Brazilian Portuguese, Quebec/Quebecois French, and Italy/Italian. English was also included as a control: we selected the English varieties spoken in New York and Boston. However, fieldwork research showed that the Canadian situation was rather different from what we had envisaged. Speakers of this area were in fact mostly Italo-Romance/French/English trilingual. English in particular was very perceptible in their spoken language, and therefore constituted an interference that was difficult to overcome.

After some research, it became clear that Italo-Romance speakers in French-speaking Belgium present a profile that can be compared to that of our target population in Argentina and Brazil. In Belgium, we found speakers who had left Italy in the 1940s–1960s. Despite the geographical proximity between the two countries, the relationship of these speakers with their homeland was as severed as that of the Italians who had emigrated to South America. Moreover, no interfering additional languages (besides the target varieties and the contact language) were detected. It was therefore decided that the data collection for contact with French should be moved from Quebec to Belgium. Finally, fieldwork in Italy has not been prioritized for the reasons given above. Nonetheless, contact data from some Italian regions were collected through online questionnaires: observations on this process are outside the scope of the current study.

This article is based on the fieldwork sessions carried out by the Microcontact team. The first session targeted Argentina only (cf. Section 3 below) and took place in May 2018. This was followed by three parallel fieldwork sessions (March/April 2019) completed in Argentina, Brazil, and Quebec. The control fieldwork took place in New York City between October 2019 and January 2020, while a pilot fieldwork study in Belgium was carried out in November/December 2019. It will be immediately clear that in these early fieldwork sessions the primary focus was not on data elicitation (although we did have a questionnaire to ascertain at least some basic facts regarding heritage syntax), but rather on checking the status of these languages and looking for speakers with the right profile, given that very little to no up-to-date information was available to us regarding these heritage speakers and their languages. Our intention was to gain an initial overview of the syntactic profile of these speakers through the first fieldwork studies, then to return to Europe, analyze the data, and formulate some hypotheses, before carrying out more extensive fieldwork later to verify them. This second, more extended, period of fieldwork was planned to take place in

## 2 Documenting Italo-Romance minority languages in the Americas

spring-summer 2020, but this has been impossible because of the COVID-19 pandemic, which means that our data are mostly incomplete, but we made use of an online data collection that did bring some results.

This article, however, does not present an analysis of the results of our investigation; instead, it provides a report of the organization and realization of the data collection, with specific focus on the fieldwork part.

Table 1 presents an overview of the number of speakers we managed to reach and interview during fieldwork 1 in the various locations. They are listed by generation. Table 2 provides an overview of speakers interviewed remotely during the pandemic.<sup>3</sup> At the moment, it is not clear when fieldwork 2 will be able to take place, nor whether it will be possible to undertake fieldwork before the end of the project, which will be in June 2022.

Table 1: Number of speakers interviewed by generation

	Brazil	Argentina	Canada	US (NYC)	Belgium	Total
G1	7	50	34	32	6	129
HS 1	1	14	2	26	2	45
HS 2 and ff.	42	10	–	–	–	52
Total	50	74	36	58	8	226

Table 2: Number of speakers interviewed remotely during the pandemic

Brazil	Argentina	US	Belgium	Italy	Total
44	61	4	3	228	340

Given the age of the speakers and the conditions under which the fieldwork was planned to take place, we expected the data collection to be quite difficult: in this paper, we discuss the issues that arose during the preparation of the fieldwork and while it was underway. Each section focuses on a specific stage of the data collection and is structured as follows: first, we introduce the background, i.e. the information that is already available in the literature and how we planned to use it to carry out our data collection (“where we started” subsections). Then,

---

<sup>3</sup>Additional data were collected from France (4), Australia (3), and Uruguay (1).

we describe what the actual situation turned out to be (“what we found/did” subsections). We conclude each section with a list of tips and warnings about what needs to be taken into account when setting up similar research.

More specifically, Sections 2 and 3 address issues related to our fieldwork, with a special focus on its practical (Section 2) and theoretical (Section 3) preparation, and on the main problems encountered when working with an elderly population.

## 2 Documenting Italo-Romance varieties in the Americas

### 2.1 Where we started

Fieldwork for Italo-Romance varieties in the Americas is somewhat unique and different from other kinds of fieldwork, in that it targets varieties that have been known, spoken and in many cases also written for centuries, but that are now found outside of their original environment. Furthermore, these languages have undergone contact with other Romance varieties for a considerable amount of time, and are therefore rather difficult to understand even for native speakers of the baseline varieties in Italy. On the one hand, the situation is not comparable to that of documenting a previously undocumented language from an uncertain family; on the other hand, it is not as simple as carrying out a dialectological inquiry in Italy, where people share a common language (Italian) and can understand instructions and translations into Italian, and share at least one language with the interviewer.

In what follows, we report our fieldwork experience, focusing on the make-up of the Italo-Romance speaking communities for this section, and on the results of the syntactic research in the next section.

The initial fieldwork was preceded by a data crowdsourcing enterprise, which consisted in asking younger generation speakers to record the elderly members of the community and upload the recordings on an interactive atlas. The atlas can be found here: <https://microcontact.hum.uu.nl/#home>. While this atlas had a large response from Italy, both North and South America were almost completely unresponsive. The entries that are now visible on the atlas were mainly uploaded by our fieldworkers.

Before turning to the presentation of each fieldwork area, some general considerations are in order regarding data protection protocols that are in place in Europe but not elsewhere. No fieldwork can start without a certified ethical clearance and an approved data protection protocol in compliance with the latest GDPR (General Data Protection Regulation 2016/679; see Leivada et al. 2019),

which enforces strict directives within the EU.<sup>4</sup> These directives may not be entirely consistent with those of the non-EU countries. The challenge lies in checking the GDPR against the regulations of the target country, aiming for an optimal level of mutual adherence. This can be done with the support of the embassies, but responses can be slow or even unforthcoming. An effective alternative is to invite universities in the target countries to co-supervise the fieldwork, thus ensuring that data collection and storage comply with the regulations of both the EU and the target country.

Regarding the data collection itself, it must be kept in mind that Italo-Romance communities in the Americas have very different characteristics. In this subsection, we review the information about Italo-Romance communities that was available in the literature before the start of the project. It will be immediately clear that the type of information available and the level of detail in the data reported in each subsection is significantly different. This is a reflection of the documentation of these varieties and their speakers: North America – particularly the US but also Canada – has a long tradition of heritage studies, mostly in the field of sociology and anthropology, but also in linguistics. Furthermore, the ethnic background of US citizens has always been meticulously recorded; we therefore know exactly how many Italians live in each state, and where they are from, while this is not the case for Argentina and Brazil. Italians in Argentina in particular have mingled with the local population and switched to Spanish much faster than any other group.

In the following subsections, we report the kind of information that was available to us before we planned the first fieldwork.

### **2.1.1 Argentina**

Argentina was a very popular destination for Italian immigrants in the 19<sup>th</sup> and 20<sup>th</sup> centuries. According to the website of the Ministry of Foreign Affairs and International Cooperation of Italy, Argentina received 57% of the total number of Italian people who emigrated overseas between 1946 and 1955.<sup>5</sup> Maurizio (2008) reports data from the Instituto Nacional de Estadística y Censos de la República Argentina, showing that a total of 2,604,447 immigrants lived in Argentina in 1960; 18% of these immigrants were from South America, and 82% from other countries, of which 31% were from Italy.

---

<sup>4</sup>The GDPR became effective after the starting date of our project. Before then, a different set of rules regulated data protection within the EU: we therefore had to modify our original protocol to make it compliant with the new regulations.

<sup>5</sup>These data are taken from [https://www.esteri.it/mae/doc\\_osservatorio/rapporto\\_italiani\\_argentina\\_logo.pdf](https://www.esteri.it/mae/doc_osservatorio/rapporto_italiani_argentina_logo.pdf).

At first, Italian immigrants moved to Argentina only temporarily for economic purposes, with the aim of improving their quality of life once they were back in Italy. This form of immigration began between the 18<sup>th</sup> and the 19<sup>th</sup> century, but became a mass phenomenon in the last quarter of the 19<sup>th</sup> century. Temporary immigrants either stayed in Argentina for several years and then moved back to Italy, or they were seasonal workers, who left Italy during the local autumn/winter and came back during the spring/summer. In the period of mass immigration, this trend was accompanied by permanent immigration, where families would relocate and settle in the new country (Ferrari 2008). Geographically, the first immigrants were predominantly northern Italians; in the final years of the 19<sup>th</sup> century, immigration from the South increased, until southern Italians formed the majority of immigrants just before World War I.

The places that were most influenced by the arrival of southern Italians were cities such as Buenos Aires, Córdoba and Santa Fe. In these cities, the number of immigrants from various countries was extremely high. Spanish was not only the official language, but also the lingua franca for immigrants who had different first languages (“L1s”). This was the optimal condition for the emergence of hybrid varieties like Cocoliche (see a.o. Bagna 2011), a contact variety often described by contemporary sources as a mix of Spanish and Italian, although it should be noted that the Italian elements often came from Italo-Romance varieties rather than from Italian.

In some areas of Argentina, however, linguistically homogeneous communities arose, creating linguistic islands, such as in the Boca and Colonia Caroya. The first of these, in the Boca, a district of Buenos Aires, was created by immigrants from Genoa, whose dialect is described as extremely widely-used and alive from the second half of the 19<sup>th</sup> century onwards, but was reported as essentially dead during the 1980s. In contrast, Colonia Caroya, a town in the province of Córdoba, was home to immigrants from Friuli, and the language was still alive and in popular use in the 1980s (Meo Zilio 1990). This is the only information we had regarding these varieties in Argentina.<sup>6</sup>

With regard to the situation for Italo-Romance varieties from the south of Italy, we had only minimal details before the fieldwork took place.

---

<sup>6</sup>A reviewer suggests that we should include the exact number of speakers and the statistics regarding these varieties in south America in previous centuries. These statistics do not exist, and we are including here everything that we were able to find. Although the information we have is obviously incomplete, we still believe that it provides a useful idea of the language situation of Italian emigrants in Argentina.

### 2.1.2 Brazil

Brazil was one of the main destinations for Italian emigrants in the second part of the 19<sup>th</sup> century. According to Cenni (2003), the main areas of Italian immigration were the states of São Paulo (especially for immigrants from southern Italy) and Rio Grande do Sul, the southernmost Brazilian state (especially for immigrants from northern Italy).

In São Paulo, Italian cultural heritage is still alive, but heritage languages ('HLS') died out fast, as the communities assimilated to the Portuguese-speaking majority. Moreover, Brazilian authorities carried out campaigns against the use of foreign languages (including Italian and Italo-Romance varieties) in the 1930s, which ultimately led to a ban on their use in the 1940s. Consequently, there are no traces left of southern Italo-Romance varieties in São Paulo, nor of the "Paulistano" Italian, a koine variety of Italian strongly influenced by Portuguese that was used in the city at the turn of the 19<sup>th</sup> century (Cenni 2003).

The case of northern immigrants in Rio Grande do Sul is different: they settled in extremely isolated mountain areas, which allowed their varieties to resist the ban on the use of the language imposed in the 1940s and the pressure exerted by Portuguese in recent years. Almost half a million people, descendants of the original settlers, still speak a northern Italo-Romance variety in the area, a phenomenon that has been the focus of multiple sociolinguistic studies conducted in Brazil. These speakers, despite being mainly third- or fourth-generation HSs, are native speakers of an Italo-Romance variety and in most cases have no knowledge of Italian. They are hence good candidates for the study of contact with Portuguese, their second language. Since these communities are particularly isolated and not very easy to reach, during the year prior to the fieldwork relationships were developed with the Venetan Association of Rio Grande do Sul and the Federal University of Santa Maria, as well as a few other contacts in the area, with the goal of developing trusted local contacts.

Italian immigration to Brazil started to decline at the beginning of the 20<sup>th</sup> century and came to an almost complete halt after World War II, earlier than in other American countries. This makes it very difficult to find G1 speakers who are still alive. One exception to this is the city of Porto Alegre, the capital of the state of Rio Grande do Sul, to which immigration from southern Italy continued after World War II. We therefore selected the state of Rio Grande do Sul as our target area for the fieldwork in Brazil, as it is home to both G1 and HSs of both northern and southern varieties.

### **2.1.3 Quebec**

The situation in Quebec was expected to be unlike that found in Argentina and Brazil. To begin with, the demographics of the Italian emigrant population are quite different from what is found in the rest of the Americas: emigration to Canada, and more specifically to Quebec, is relatively more recent than that to the other areas under investigation. Although the first records of Italian emigration to Canada date back to the last quarter of the 19<sup>th</sup> century and the flow of migration never completely stopped, the period of intense movement was fairly brief, lasting from 1951 to 1967. Due to this demographic difference, the majority of first-generation speakers in Canada were typically not (completely) illiterate when they left Italy, as they had received at least some formal education in Italian, and had also been exposed to Italian in the increasingly popular media. Therefore, we knew that Italian, or at least a non-standard variety thereof, would be a not insignificant source of interference on the dialects spoken by our informants. We based our knowledge on the available studies on the language(s) spoken by the Italian community in Montreal (focusing on their Italian: Reinke 2014 and the extensive work by Villata, e.g. Villata 2010).

Moreover, the overall migration flow to Canada was never at a level comparable to that of our other research areas: the peaks, registered in 1956, 1958, 1966, and 1967, were of roughly 28,000 people a year.<sup>7</sup> We therefore expected to find fewer participants in this research area than in our other target areas, even if we did not have the exact numbers for Quebec alone.

Finally, some parts of Quebec are *de facto* bilingual areas: while French is the only official language of the province, English is widely spoken (as well as being an official language of the country), especially in Montreal. We based our decision to include the area in our study on knowledge of pro-French campaigns and policies, which were particularly prominent in the 1970s. However, we were prepared to find some (reduced) instances of speakers also proficient in English, at least to some extent: those would have ideally been excluded from our study, to avoid the confounding factor of an additional variety, particularly a non-Romance one.

### **2.1.4 US**

Italo-Romance varieties have been exported to the US since (at least) the 19<sup>th</sup> century; the 1880 census lists 81,249 Italian migrants, a number that had increased sharply to 4,114,603 by 1920 (cf. Cavaioli 2008). In the census carried out in 2000,

---

<sup>7</sup>Source: ISTAT, [http://seriestoriche.istat.it/fileadmin/documenti/Tavola\\_2.9.1.xls](http://seriestoriche.istat.it/fileadmin/documenti/Tavola_2.9.1.xls).

“Italian” was reported to be spoken by about 1,000,000 people in the US, with the most significant numbers concentrated in the Northeast of the country, where we carried out our fieldwork.<sup>8</sup> However, according to the multi-year American Community Survey 2009–2013, there has been a decrease of about 300,000 speakers, i.e. a drop of a third in just over 10 years. This is most likely due to a rapid language shift to English only, which typically occurs within the third generation in Italian communities abroad (see De Fina 2014; for NYC, see Haller 1987, 1993). For this reason, the vitality of Italo-Romance varieties spoken in the US is endangered, as it is virtually impossible to find Italo-Romance HSs after the first US-born generation.

Moreover, these statistics combine all the languages imported by Italian migrants under the umbrella term “Italian”. This is slightly inaccurate, as pre-World War II migrants mainly exported their local languages, and had minimal knowledge of Italian, or perhaps none at all. This situation changed after World War II, particularly after 1965 with the Immigration Reform Law, thanks to which the families of Italian migrants were allowed to legally live and work in the US. These more recent waves of migrants had an “Italianizing” impact on the local Italo-Romance languages, as speakers were educated in Italian and were hence no longer monolingual Italo-Romance speakers as the previous migrants had been (Haller 1991: 391–392, De Fina & Fellin 2010). As a result, the speakers’ competence in their local languages began to be affected by the new wave of Italian, as well as by English. Moreover, the coexistence of more-or-less-intelligible local varieties brought about the need for a linguistic koine, i.e. a shared Italo-Romance variety intelligible to everyone. This koine has since been the focus of the majority of studies of Italian communities in English-speaking countries. Indeed, this situation is documented for New York by Haller (1987 *et seq.*) and is common to other urban contexts with Italo-Romance–English contact, such as Sydney Italian (Bettoni 1990, 1991) and Montreal Italian (Reinke 2014).

For New York, Haller’s (1987, 1993, 1997a, 1997b, 2002) work provides a solid description of the sociolinguistic situation of the Italian community between 1980 and 2000. He proposes a multilingual continuum for Italo-Romance varieties, which are “used, besides English, with various degrees of competence, according to generation, time of emigration, and education” (Haller 1987: 396): “Standard”

---

<sup>8</sup>New York State (294,271), New Jersey (116,365), Pennsylvania (70,434), Massachusetts (59,811), Connecticut (50,891), Maryland (13,798), Rhode Island (13,759), and Virginia (10,099) [<https://www.census.gov/population/cen2000/phc-t20/tab05.pdf>]. Note that the same level of accuracy and documentation is not available for all countries; we provide here what was available to us prior to our fieldwork, and the information on which we based our planning.

dialectal Italian, Italianized dialect, pidginized American Italian, and archaic dialects". With regard to the koine variety, Haller confirms that

[t]he migration from the depressed South to Rome and Northern Italy and the emigration to the United States both acted as "Schools of Italianization", exposing individuals for the first time to other dialects and languages and forcing them to develop a lingua franca in order to be able to communicate with each other. (Haller 1987: 393)

Hence, while the Italo-Romance local dialects would be employed within the family and closer-knit circles of fellow countrymen, this Italian koine had been functioning as the "community language" for decades (Haller 1991, 1997a: 401). A few decades later, this situation has reached a point at which the Italo-Romance varieties are increasingly fading away, and Italian is taking over or Italo-Romance is being abandoned altogether in favor of English.

### **2.1.5 Interim summary**

Although we were aware of the social differences between the areas in which we planned to collect data, before starting our fieldwork we were working under the hypothesis that the most relevant socio-historical conditions were comparable for all the targeted countries: we expected to find G1 speakers with very low competence in Italian, if any (even in its regional varieties). These G1 speakers would have maintained little contact with their communities of origin in Italy. Moreover, we expected the varieties under analysis to be faithfully preserved by the communities abroad, at least as home languages, and as such passed on to the following generation(s). The advantages of this scenario would have been the possibility of systematically excluding the influence of external factors (e.g. competence, language exposure, age of bilingualism onset, etc., as well as socio-historical variables) on the development of the phenomena under analysis, and to assess, for each HS, the input language (i.e. the language spoken by their parents, the G1) with a good level of detail at the microvariation level, too.

However, in some cases the socio-historical differences proved to be more far-reaching than expected and to have not insignificant impact on the linguistic profile of our informants. These issues are discussed in more detail in the following section, along with some practical matters to be considered when organizing fieldwork, and the solutions that we found to the various problems that arose.

## 2.2 What we found

### 2.2.1 Argentina

As discussed above, Argentina was the destination of huge numbers of immigrants from the end of the 19<sup>th</sup> century onwards. However, when we tried to contact associations of Sicilian, Neapolitan and Abruzzese speakers by email, we received a large number of Non-Delivery Reports (NDRs) that led us to conclude that they were no longer active. We also asked for information from Facebook groups dedicated to Italian descendants in Argentina, as well as from some distant relatives of ours and of other contacts with relatives in Argentina.<sup>9</sup> Neither crowdsourcing nor any subsequent attempt helped to identify speakers who would be eligible to take part in our study.

Due to this lack of information, we decided to carry out a pre-fieldwork exercise, with the specific aim of checking the current situation within the Italo-Romance communities and establishing a network of informants for the following fieldwork. This pre-fieldwork was carried out only in Argentina as this was where finding contacts in advance had proved most challenging. Once in Argentina, our researcher was able to establish a good network after visiting the presidents of some associations and some colleagues at local universities.

The pre-fieldwork was followed by the first main fieldwork session, during which we targeted immigrants who had moved to Argentina after World War II, as well as the small number of their descendants who had acquired the Italo-Romance variety. As in other American countries, institutions, especially schools, played a major role in the diffusion of the monolingual Spanish model. The researchers were told multiple times that teachers explicitly suggested, or even ordered, that the parents should speak only Spanish to their children, in order to avoid “confusion” for the child.<sup>10</sup>

Unfortunately, these interventions were effective in the majority of cases, meaning that the Italo-Romance varieties were abandoned by almost all immigrants.

<sup>9</sup>In the case of Abruzzese, one of our contacts wrote to us: “La Argentina ha recibido inmigrantes de todo el mundo que han traído sus idiomas y dialectos, pero al haberse mezclado con toda la sociedad no sabría si continúan hablando el dialecto. Por ejemplo mi abuelo Francisco no hablaba su dialecto, hablaba español.” [‘Argentina has received immigrants from all over the world, who brought their languages and dialects with them, but as they have integrated into society, I’m not sure whether they still speak their dialects. For instance, my grandfather Francisco didn’t speak his dialect, he spoke Spanish.’]

<sup>10</sup>As one of our speakers told us during the interview: “In taule a si fevelave simpri furlan, ai vut tancj di chei problems ta scuele, parcè che i disevi peraules in furlan, e an clamat a me mari, che no si feveli plui il furlan parcè che no si podeve.” [‘We would always speak Friulian at home, I had so many problems at school because I would say words in Friulian, and they called my mum so that we would not speak Friulian anymore because it was not allowed.’]

Exceptions are found primarily when there were elderly family members (especially grandparents) who never learned Spanish and thus kept speaking their Italo-Romance L1 to their grandchildren.

Despite the geographical distance, during both fieldwork studies in Argentina we found that the local Italian community managed to keep strong bonds to their hometowns through frequent visits to Italy (particularly from the 1980s onwards) and through the countless regional and local associations. According to our informants, these associations were particularly active in the 1940s-1970s, and helped to recreate a sense of Italian community through recurrent parties and celebrations. The members of the associations have tried to maintain the traditions of their home regions, such as religious celebrations, typical food and even clothing. Curiously enough, the only thing they usually did not maintain is their language, switching to Italian or to Spanish.

More generally, Italian gained ground because of marriages between people from different regions of Italy, who chose to speak Italian to their children, when they could, in order to keep a stronger bond with their home country. In the bigger cities, there are also Italian schools that some of our speakers attended. Finally, although the immigrants and their descendants feel a particular link to their region, they feel proud of Italy as a whole and identify with it, particularly when they talk to people who are not descended from Italian immigrants. As a consequence, many first-generation immigrants (as well as the subsequent generations) speak Italian alongside their local variety, and they all insisted on speaking Italian to the fieldworkers.

The various Italian associations were very useful in our search for informants, since they know most members of the community, but even they could not identify more than a couple of speakers each, as there are not many speakers left. However, some associations do offer courses in their Italo-Romance variety. These courses are attended by second- or third-generation immigrants who never developed a high proficiency in the language and wish to improve it or even learn it from scratch. In the Friulian association of Buenos Aires and in the Piedmontese association of Córdoba, for example, 8-10 people followed an Italo-Romance language course. These courses were a useful source for our search for HSs.

The informants we interviewed were found mainly through members of these associations. In Santa Fe, we found some speakers thanks to the Italian scholars at the Universidad Nacional del Litoral, who are running a project on the Italian cultural heritage of the city.<sup>11</sup> In three cases we found speakers of Cocoliche (see

---

<sup>11</sup>The person responsible for the project is Prof. Adriana Crolla (<http://www.fhuc.unl.edu.ar/portalgringo/crear/gringa/>).

Section 2.1.1 above): unfortunately we could not interview them, because they were some of the oldest members of the community and they were not able to perform our tasks. Interestingly, they were usually no longer able to distinguish between Cocoliche and their own Italo-Romance variety, a situation that we also found in New York with the koine and the Italo-Romance varieties.

Overall, most informants were kind but somewhat suspicious at the beginning, especially in the bigger cities: they all refused to allow the fieldworker to visit them at home, unless they were accompanied by a member of the community whom they already knew. As a result, when possible, the interviews were held in emigrant association premises. However, in some cases they had to be carried out in bars, which made for a less than ideal situation, as informants could be distracted, the audio stimuli of the questionnaire were difficult to understand because we had to lower the volume, and the recordings were affected by background noises.<sup>12</sup>

With regard to the geographic distribution, we observed that the Italo-Romance varieties are still found in the main cities. Nowadays, however, their use is limited to the family group, and we could find very few second- or third-generation speakers with a high proficiency in the language. In the smaller towns and villages, on the other hand, the Italo-Romance varieties have virtually died out. One example is Colonia Caroya, in the province of Córdoba: until a few decades ago, Friulian was the main language (Spanish being the only official language), to the extent that, according to our informants, it was impossible to find a job there if you did not speak Friulian. The situation today, however, has radically changed: we could only find three informants (all above 70 years old), while the rest of the community speaks neither Friulian nor Italian.

### **2.2.2 Brazil**

Unlike in Argentina, Italo-Romance varieties in Brazil have survived mainly in the countryside and, to some extent, in bigger cities in southern areas of the country, as we highlighted in Section 2.1.2. Immigration from Italy after World War II was very limited, and it was hence challenging to find first-generation immigrants with the right profile for our research. The state of Rio Grande do

---

<sup>12</sup>One reviewer observes that it would have been possible to train local members of the community to perform the data collection for us. This was not possible due to a lack of time, during the first round of fieldwork, given that the researchers had to travel extensively to meet the various speakers. The second round of interviews, which was performed online, was instead realized with the help of trained local community members, who knew the interviewers by then and gladly agreed to help when possible.

Sul was the best option, as both northern and southern varieties are still spoken in the area by G1 and HSs.

Ten interviews were carried out in Porto Alegre, which is home to a large community of Calabro-Lucanian speakers, as well as smaller Sicilian and Abruzzese communities. The language used by the interviewer was mainly Portuguese. The first problem encountered in Porto Alegre was the general sense of mistrust shown by local associations of HSs towards the research; in particular, one of the local associations refused to help the researcher to find informants. A few participants were found as a result of posting on Facebook groups of descendants. The best result in Porto Alegre, however, came from the successful cooperation with the Calabrian Center of Rio Grande do Sul; this contact was established prior to the fieldwork and was helpful in finding Calabro-Lucanian G1 and HSs. The Calabrian community in Porto Alegre is made up of immigrants (and their descendants) from the town of Morano Calabro; they all therefore speak exactly the same variety, the Moranese dialect. However, the general sense of distrust was also evident here; obtaining signed informed consent was particularly problematic, since most of the participants were afraid that it could be a scam.

The interior of the state of Rio Grande do Sul is home to almost two million descendants of immigrants from northern Italy, most of whom have maintained the language of their ancestors across generations, as a result of the isolated nature of these communities. 42 interviews were carried out in the area of the Serra Gaucha and in the Fourth Colony of Italian Immigration. The languages used by the interviewer were Venetan and Portuguese. The isolated nature of these communities, which helped to preserve three northern varieties (Venetan, Friulian, and Eastern Lombard), also represented the main obstacle in reaching the speakers. In some cases, the only way to get to the villages was by car. The contacts we established in the area not only made it possible to physically reach the speakers, but were also of great help in communicating with older people who, in some cases, had never left their villages and were hence not always willing to talk to a foreigner or to be recorded. Unlike the Calabro-Lucanian informants in Porto Alegre, speakers of northern varieties in Rio Grande do Sul are the descendants of immigrants from different areas in Italy. Their dialects are not exactly identical to each other, nor to the varieties of the languages spoken in Italy. Moreover, isolation from Italy has allowed the preservation of archaic features of the languages in these varieties. Obtaining written informed consent also proved problematic for informants in the Serra Gaucha and in the Fourth Colony of Immigration. The interviewer decided to opt for recorded oral consent by the informants: the whole informed consent form was read out loud during the recording and participants consequently agreed to be interviewed.

Overall, the good outcome of the data collection in Brazil was heavily dependent on the contacts the interviewer established prior to the fieldwork. The need to have trusted relationships with local members of the communities proved to be essential.

### 2.2.3 From Quebec to Belgium

In Quebec, the socio-historical issues highlighted above (younger emigration, bilingualism) turned out to be significant, as they substantially altered the speakers' profiles. The main differences with respect to the speakers we found in Brazil and Argentina were the age of the speakers (younger, which in turn entailed physical and cognitive difference – cf. Section 3.3.1) and their knowledge of other languages: most of our speakers had full knowledge of (regional) Italian and English, in some cases at the expense of French (the target contact variety for the area). Moreover, the Italian immigrants in Quebec had managed to achieve a reasonable level of economic security, which had led, among other things, to increased contact with Italy. Most of our informants had been to Italy several times since they had left the country, and some of them regularly spent their holidays there, even for several months a year. Even those who had no direct or prolonged contact with their hometowns had been in regular contact with their relatives in Italy via long-distance communication devices over the years.

The different sociolinguistic settings of the two main areas of investigation, Montreal and Quebec City, posed a further problem in Quebec. The latter has a small Italian community whose members mostly switched to French on a daily basis because of marriage with their local partners: on the whole, this had a negative impact on the transmission of their languages to the subsequent generation(s), as attested by the fact that we could only find one HS of an Italo-Romance variety in the whole city (Venetan), compared to 9 G1 speakers of different Italo-Romance varieties, mostly Piedmontese, Venetan, and Friulian. Montreal, on the other hand, was the destination of a massive flow of emigration from Italy: the traditional Italian area, la *petite Italie* ('little Italy'), is nowadays a sheer memory of the Italian emigration. Despite being home to the Church of the Madonna della Difesa, built by the Molisano community in the late 1910s, and to some Italian shops and cafés, over the years the bulk of the Italian community has moved away from the area, to the outskirts of the city (e.g. Saint-Léonard and Rivière-des-Prairies, in the northern part of the island) as well as to its neighboring municipalities, especially Laval. However, this setting also proved detrimental to the preservation of the Italian dialects: the presence of emigrants from linguistically different areas of Italy and the need for mutual help and support resulted in the

development of an Italian koine, which, alongside English and French lexical borrowings, includes many regional structural and lexical features and a (broadly speaking) Italian structure, somewhat resembling that which is found in New York. This variety (“Italianese”, e.g. Villata 2010) is most readily available to our informants for daily communication within the community and, over time, has come to overshadow the original dialectal richness of the city: as a result, once again, we only managed to interview one HS of one of the target Italo-Romance varieties (Sicilian), compared to 22 G1 speakers of various target languages.

A further difference between the two areas of interest in Quebec is related to the contact varieties, as mentioned above. While French is the contact variety for the Italians who emigrated to Quebec City, English is the most widely spoken local language for the Italian community in Montreal. The widespread use of English, despite efforts (particularly via education policies) to make all new residents and their descendants use French, is due to the prestige of English in the wider North American context. Due to these differences, our informants in Quebec were very far away from our ideal speaker profile: instead of illiterate, dialectal speakers with a working knowledge of the contact variety, we found quite literate informants, with a good knowledge of some variety of Italian (a koine one, with specific regionalisms, if not a variety closer to the standard), a passive to good knowledge of French and a good general knowledge of English, the most widely used language outside of (but sometimes also inside) the community, especially in Montreal.

In an attempt to find a speaker profile that more closely matched the ideal speaker for this research, we turned instead to Belgium. Italian emigration to Belgium reached its peak right after World War II, with the bilateral agreements between the two nations. First signed in 1946, but effective until 1956, these agreements regulated the transfer of Italian workers to the Belgian mines. Despite the well-known presence of Italians in the country and its greater geographical accessibility, Belgium had been originally excluded from the investigation areas because of its (relative) closeness to Italy, which could have led to extensive contact between the emigrants and their hometowns. However, after we had established that the Italian population in Quebec did not sufficiently match the required profile, and given the accessibility of the area, we decided to run a small-scale fieldwork exercise in the French-speaking part of Belgium.

This pre-test on the general feasibility of a more in-depth study of the Italian emigrant community in Belgium, taking into account our original socio-historical and sociolinguistic conditions, proved extremely fruitful. We decided to target the southern mine region around Charleroi and La Louvière, as well as Brussels;

in total, we interviewed eight informants: six G1 and two HSs. Our original concerns regarding the possible stronger connection between this population and their homeland turned out to be unfounded: we discovered that the terms of the contracts for the mine work ensured that the miners did not go back to Italy for the whole duration, reducing the links considerably. After the contracts expired, some workers went back to Italy for good; those who stayed in Belgium continued their previous lives away from their home country.

The general speaker profile was a better match for our ideal speaker. Due to the disruptions to the school system brought about by the war, we found in our interviews that the emigrants, in most cases fairly young men, were mostly illiterate, which was no doubt beneficial to the maintenance of the dialects. Furthermore, the well-organized migration flow brought people from the same areas of Italy together, both to live and to work: this ensured that the dialects, as the most accessible varieties known to the miners, were preserved in daily life and even passed on to the following generation(s). This situation is demonstrated by the presence, even today, of small municipalities in the south of Belgium where Italian dialects are still widely spoken among the members of the Italian community: an example is Sicilian in Morlanwelz, which was the destination of a considerable wave of emigration from Villarosa, in Sicily (Enna province).<sup>13</sup> These local and linguistic clusters, together with the limited contact with Italy over many years, also resulted in a very limited knowledge of Italian. Moreover, and again in contrast to what we found in Quebec, French (less so its local Walloon dialect) is the major language spoken by these emigrants, alongside their original dialects. No other contact language is attested, making the sociolinguistic context of Belgium overall more suitable for our study. The only disadvantage of the research in Belgium is the very limited number of varieties available: the majority of the miners originally came from southern Italy, mostly from Sicily and Campania. Up to this point we have found very few speakers of northern varieties. Unfortunately, it has so far not been possible to continue this small-scale fieldwork with a more extensive data collection.

#### **2.2.4 US**

The search for Italo-Romance speakers in a metropolis such as New York City was not straightforward. This is due to the fact that, over the years, the different communities of Italians experienced disaggregation and displacement within and outside the urban area, as we discuss below. The most fruitful means of finding

---

<sup>13</sup>This is reflected, at institutional level, by the fact that the two are twin-towns: <https://en.wikipedia.org/wiki/Morlanwelz>.

our speakers was to bypass the official routes, such as the (unresponsive) Italian Cultural Institutes and Embassy, and instead to approach:

- clubs, societies and cultural associations of local communities;
- shops run by Italians (pizzerias, restaurants, tailors, barber shops, *etc.*);
- individual contacts (internal or external to these communities) who acted as mediators between the researcher and the Italian communities, e.g. Endangered Language Alliance.

A total of 58 speakers (G1: 32; HSs: 26) from different heritage communities were interviewed in Manhattan, Brooklyn, and Queens, as well as one family in Jersey City, NJ. The people we interviewed were speakers of Friulian, Nònes (a Ladin-Lombard/Venetan transitional variety from Trentino), Eastern Abruzzese, Neapolitan, and Sicilian. We also interviewed speakers of varieties that were not considered in our other field trips: Eastern Campanian, Cilentano (a southern Campanian variety), Apulo-Barese (from central Apulia), and Ciociaro (an upper-southern variety from southern Lazio). Moreover, as mentioned in Section 2.1.4, most of these speakers had some knowledge of the spoken Italian that they exported (in the case of G1) or learnt in the US (some HSs). However, many HSs only had active knowledge of the Italo-American koine, especially those HSs whose families came from the south of Italy, as we discuss below.

G1 speakers migrated between 1940 and 1980 from different areas of Italy, where the local varieties are spoken alongside (regionally marked) Italian. Most of the G1 speakers who arrived immediately after World War II settled in the same place as the historical communities of Italians during previous migration waves. These areas include (but are not limited to) Manhattan (Rose Hill, Little Friuli in Murray Hill), Brooklyn (in and around Bensonhurst, Williamsburg-Green Point), and Queens (Astoria). However, due to the arrival of new migrant communities and the increasing cost of living in the city, most of these Italian communities were forced to move away from these neighborhoods and to relocate to more peripheral areas.<sup>14</sup> These displacements led to the partial or total dissolution of once-compact linguistic communities; moreover, language shift to English-only happened very frequently due to the generational change of “community leaders” in clubs and associations. In fact, finding US-born HSs who are (fully) proficient in their local HLs proved rather challenging, as the large majority (especially from southern Italy) had been forced to switch to English-only by their families

---

<sup>14</sup>Mainly further East in Queens, further South in Brooklyn, and Staten Island, or outside the five boroughs of NYC.

for integration purposes, therefore completely abandoning their HLs and retaining, in the best cases, only a passive knowledge of them.

The linguistic profiles of the US-born HSs we interviewed included: (usually, highly educated) speakers with active competence of both their own local variety and Italian, which they learnt from (educated) family members, or during secondary education; speakers with different degrees of active competence in their own local variety (depending on their age, the type and length of exposure to that variety, and cohesion of their community), but little competence in Italian; and speakers with passive knowledge of the local variety of their own families, which they define as “the archaic dialect”, and active competence in a regionally influenced variety of (Americanized) Italian koine, the lingua franca, which they refer to as “the modern dialect”, or Brooklynese/Queenese Italian (alongside the local Brooklynese/Queenese English).

Unsurprisingly, the most proficient speakers are from the elder generation, i.e. over 70-75 years old (with very few exceptions). The linguistic repertoire of these speakers is extensive, as they learnt their own HL from close-knit communities of Italian-born parents, grandparents and/or other relatives, who had emigrated in the first half of the 20<sup>th</sup> century to the specific Italian areas/neighborhoods of New York City. Many of these elderly speakers also maintained close connections with their families in Italy, allowing them to also be exposed to the home dialect and/or spoken Italian. Indeed, these HSs learnt a conservative, “frozen-in-time” variant of their own HL from their families of G1 *émigrés*; when they visit their family’s birthplaces in Italy and interact with locals there, they are told they still speak an archaic version of the relevant dialect (cf. Aalberse et al. 2019: 114). This is not only due to the fact that older HSs learnt a conservative variant of the HL imported by their families, but also that these speakers were not exposed – as much as later generations in the US have been – to the growing linguistic pressure of Italian over the last 70 years.

One feature shared between some Italian-born G1 speakers who migrated to the US in their early years and US-born HSs is that they all grew up as sequential bilinguals. They first acquired their own Italo-Romance HL (the local dialect and/or the supraregional Italian koine), and later English as “Child L2”, i.e. bilinguals who acquired a second language between 4 years old and puberty (cf. Aalberse et al. 2019: 117).

For speakers younger than 70 years old, the level of proficiency in the HL diminishes rather drastically. This is likely due to a less constant exposure to the relevant HL, or to a drop in usage on a daily basis. Moreover, in addition to the decrease in input, the quality of that input changed as an effect of attrition and cross-linguistic influence from English, resulting from the long-term decreased

activation of the HL (cf. Pascual y Cabo 2013). Indeed, these HSs grew up learning a supra-regional variant of Italian, originally taken overseas by their families after World War II and later influenced by English, while their own local HL was mainly heard from grandparents, older relatives and/or elders in their neighborhood, but was not actively used. Indeed, from the intermediate to the new generations, the lexis, phonetics/phonology and syntax of the relevant baseline dialect appear to have blended to different extents with a spoken variant of Italian, as well as English, resulting in the Italo-American koine. In fact, speakers under 40 years old (mainly from the south) appear to show an “established confusion” about the Italo-Romance language they speak. They self-report that they are not able to speak “proper Italian” and can only speak “the dialect”, but they actually speak this Italo-American (southern-based) koine. However, the Sicilian and Nonesi communities of HSs seemed to have better preserved their local dialects, across all age groups.

As a (partial) result of the issues highlighted above for each of the contact areas under investigation, the sample of speakers we could find is more varied than we were hoping for, sometimes leading to difficulties of comparison that will be addressed in more detail in Section 3.3.

### **2.3 Tips and warnings**

Here are some things to consider if you wish to set up fieldwork outside of Europe, and in the Americas in particular:

- Before you go, set up a risk assessment plan with your university. The plan should contain information about the potential risks and an exit action in case of danger; furthermore, it should contain a safety protocol that you may establish with your PI.<sup>15</sup>
- Make sure you stick to the safety rules agreed upon with your university/PI.  
As soon as you arrive in a new place, contact or identify the relevant consulate in case you need help, and make your presence known to them. Establish a daily routine with your home supervisor, like sending an email or a message at a given hour every day to confirm that everything is okay and you do not need help.

---

<sup>15</sup>For a taxonomy of possible risks, a useful tool is this Advisory note by the International Science Council: <https://council.science/publications/advisory-note-responsibilities-for-preventing-avoiding-and-mitigating-harm-to-researchers-undertaking-fieldwork-in-risky-settings/>.

- Make sure emails to potential referents are short, clear and personalized.  
After gathering and reviewing any available sources for contacts, including outdated webpages that might suggest further connections, it is crucial to approach the possible (source of) informants with an email/message that is at the same time catchy and trustworthy, especially in the subject line. The text should be clear and concise and should include statements about (i) the purely scientific purpose of the research survey, (ii) the non-profit nature of the enterprise, and (iii) compliance with the data protection protocol (i.e. anonymity), although public institutions are usually willing to be acknowledged in research outputs. Whilst the message template should address a varied range of potential (sources of) informants, a plain, somewhat detached invitation may not engage all the intended recipients. It is worth adjusting the text in a way that demonstrates a sensitive understanding of the contact's role within the targeted local community, highlighting the shared benefits of the cooperation (academic for universities, socio-cultural for associations, motivating for individuals, etc.).
- Allow plenty of time to network with the local community (and to build a relationship of mutual trust). Pre-fieldwork could prove extremely useful for this purpose, too.  
Be it a secluded community in the Brazilian mountains or a dynamic community in Brooklyn, it is easier to access the speakers from within the local community. In particular, it is advisable to spend time participating in the life of the regional associations: parties are not only entertaining, but also an excellent means to meet people in a more relaxed environment than an interview offers, and to exchange contacts. In our case, we carried out pre-fieldwork in May as the fieldwork had to take place in the following spring.<sup>16</sup>
- Involve the local communities (even more): try to train local people to carry out the interviews.  
Training local people to carry out interviews when the researcher is not present is a very good idea, if time constraints allow. Ideally, this should happen for each collection of data, but this cannot always be achieved, especially if the time to be spent in one location is too short. In this respect,

---

<sup>16</sup>A reviewer asks how many days are recommended. We do not have a precise answer for that: the more the better, obviously, so that the researcher can get to know the people and the community a little more. In our case, the pre-fieldwork lasted about 20 days, and the limit was determined by budget considerations rather than anything else.

when calculating the budget in your project proposal, make sure that you allocate some money for your fieldwork assistant as well as for the speakers. Note that university-internal regulations might make it very difficult to transfer money overseas, so paying in local money to be reimbursed later is preferable.

- Consider technical problems in isolated/remote areas.  
In the most remote areas that we visited in our fieldwork, especially in Brazil, some basic IT requirements were not met: for instance, the internet connection was sometimes unavailable, but even more importantly power sockets were missing. This is a problem when performing a computer-based questionnaire. One suggestion might be to carry multiple rechargeable batteries for laptops and recorders; alternatively, the questionnaire should be structured in a way that allows it to be carried out without a laptop. In that case, it is a good idea to print out the questionnaire, so that it is possible to ask at least some of the questions if the laptop cannot be charged. Moreover, in some particularly isolated areas, it is advisable for the interviewer to take some water and food, in case they are stranded somewhere, and no transportation is immediately available.
- Consider problems working with communities in big, busy cities, too.  
Some informants might genuinely not have enough time to spend on a questionnaire (in our case: between half an hour and one hour), as “time is money” in their busy schedules, despite their interest and willingness to participate in the study. In New York, this was true for most of the informants, regardless of their age, and was also true to a lesser extent in Quebec. Moreover, in large metropolitan areas such as New York and Montreal, communities are scattered across the city, and travelling on public transport is extremely time-consuming (for instance, between 1 and 3 hours without leaving the metropolitan area of NYC); likewise, taxis might take as long as public transport, and are very expensive. Allow therefore plenty of time to reach your destination.
- Remember that it might be culturally inappropriate to ask speakers to sign a document they cannot read.  
In this case, still complying with the legal requirements, the informed consent form can be read aloud, with speakers giving at least their oral consent to the collection and processing of their personal data. Make sure this procedure is approved in the ethical and data protection protocols before you start your research.

- Consider the mistrust of the informants.

Both in bigger cities and in more isolated villages there may be a general feeling of a lack of safety. To overcome this sense of mistrust and to make the informant feel at ease even in the presence of a stranger, it is preferable to go to interviews with another community member whenever possible. This is also beneficial for the fieldworker.

- Be prepared to interview people in unconventional places.

As much as we tried to make arrangements in advance to carry out our interviews in a quiet environment, this was not always an option (see also the feeling of mistrust mentioned above). A good solution, when private spaces are not available, would be to carry out the interviews in offices belonging to the various cultural associations. However, even this option is sometimes not available. It might then prove necessary to carry out the interview in parks, cafés, restaurants, shops, offices, waiting halls, etc., where a quiet environment is not an option. It is advisable to keep this in mind while designing the questionnaire, as this can affect its feasibility (e.g. audio stimuli are more problematic in such contexts, and the use of a laptop might be challenging).

### **3 Syntactic tests**

#### **3.1 Where we started**

Our research consisted of three parts: the first aimed to assess the proficiency of the speakers; the second tested the presence of the syntactic phenomena under analysis; and the third gathered sociolinguistic information about the informants' language history and use. At this stage, as we had only preliminary hypotheses to test, we did not concentrate on designing the exact method for data elicitation; rather, we tried to understand whether what was reported about heritage languages and the particular phenomena under investigation was true.

In the first fieldwork trip, the main task was to ascertain the existence of subject clitics and DOM, as well as ternary demonstrative systems, and the syntactic conditions under which these phenomena might be present.

##### **3.1.1 Assessing proficiency**

We decided to make proficiency testing the first part of our research because we wanted to make sure that our informants were genuinely able to speak the

target dialect, rather than Italian with some dialectal expressions (for legitimate concerns on the issue, cf. Section 2.1). Once the data were collected, each component of the group listened to them. The group includes researchers who are native or highly fluent speakers of all the varieties under investigation; most of them are also native speakers of Italian.

Before beginning the fieldwork, a data crowdsourcing enterprise was carried out, over the course of several months. Through this crowdsourcing we had hoped to identify and hence pre-select speakers with the right profile, i.e. proficient speakers, to interview during fieldwork. However, we did not obtain the desired results from this type of data collection, as we did not receive sufficient responses from either North or South America. We therefore interviewed all the speakers who made themselves available, on the basis of their own linguistic self-evaluation, and we needed a more reliable method of assessing their proficiency in the dialect. We ultimately used this part of the interview as a pre-selection method to skim the questionnaires. Proficiency was tested in two ways: by means of spontaneous speech (first) and by performing a specific lexical decision task (afterwards). While the actual set-up of the proficiency assessment will be discussed more in detail in the next subsection, here we will introduce the bases upon which we decided to take proficiency into consideration.

Our fieldwork researchers were all native speakers of Italian, and were therefore able to assess whether the speakers were using some form of Italian or the Italo-Romance variety. Afterwards, the data were checked by the rest of the team, which included native or very fluent speakers of all the Italo-Romance varieties under investigation.

Spontaneous speech was chosen as the introductory task as it offers multiple additional advantages: firstly, we were expecting some of our informants not to speak their recessive language on a daily basis, and sometimes in fact not to speak it at all. We therefore thought that a good way to make them feel at ease could be to allow them to talk freely: we found that some of the less proficient speakers had some issues with speaking the language to start with, but became more and more fluent while talking to us. To try and trigger the use of the dialect, we mostly asked them questions about their childhood and their arrival in their new country: we hoped that by asking them to talk about a time in which they spoke the dialect on a regular basis, they could be further encouraged to reproduce it. Secondly, spontaneous speech allowed us to gather sociolinguistic information (year of emigration, level of education, current and past dialect usage, etc.) that we needed to control for, to keep our sample of informants as homogeneous as possible, without bombarding the speakers with very direct and structured questions. Thirdly, the spontaneous speech data were used to complement the

information gathered from the questionnaire, and to cross-check whether spontaneous production might reveal different linguistic patterns with respect to the elicited data gathered in the questionnaire.

In addition to using spontaneous speech, we also assessed the level of proficiency through the HALA test, developed by O’Grady et al. (2009) and widely used in heritage communities to check the level of proficiency of the speakers. The HALA test is a publicly available picture-naming set of tasks aimed at measuring the accessibility times of items and structures in the different languages in the repertoire of multilingual speakers, so as to assess the relative dominance hierarchy. The overall idea is that proficiency correlates with frequency of use in all language domains, which has direct consequences for latency times. We hence decided to include an additional task based on the HALA materials, as will be explained in more detail in Section 3.2.1.

### 3.1.2 Towards the syntactic questionnaires

The second part of our research consisted of a questionnaire testing the different syntactic phenomena in the contact varieties under analysis (for a definition of the phenomena and of the varieties, see Section 1). Our population is extremely varied in nature: G1 of different ages and levels of education, affected by attrition to different extents, and HSs. Nonetheless, we wanted to develop one unified questionnaire, so as to make our results directly comparable. We therefore had to take into account the specific challenges posed by each different group of speakers while designing the questionnaire. According to the literature, the most difficult group to test would be the HSs (for extensive remarks, cf. Polinsky 2018: chapter 3). Of course, the three phenomena have different instantiations in each of the Italo-Romance varieties under investigation; variation between them is however at the microlevel, so most tests were simply translated from one language to the next without losing crucial information, or without changing information structure, for instance.

We excluded grammaticality judgments and translations. Polar grammaticality judgments (Yes/No type) are not granular enough, while scalar judgments (in which each experimental item is rated on an  $n$ -point scale, e.g. Likert scales) detect finer differences among the stimuli, but ultimately also pose some problems (for a wider discussion, cf. Stadthagen-González et al. 2018 and references therein). The major issue is related to the very concept of scale: it cannot be safely assumed that an  $n$  point scale is sensitive enough to faithfully represent the acceptability continuum, both in its extent and in its actual match with the

speaker's own continuum representation. Moreover, and assuming that informants are consistent in applying one and the same scale throughout the task, scalar grammaticality judgments are clearly demanding on the memory load side as well. Finally, grammaticality judgments prove particularly difficult for HSs (Polinsky 2018: chapter 3, and references), which would have made it difficult to use the same questionnaire for all our participants.

Translations have been commonly used in research on Italo-Romance varieties on a large scale (cf. for instance the traditional atlases that document Italo-Romance: AIS, ALI; and in more recent times: ASIT). However, we were particularly concerned with avoiding the interference of any other language while performing the tasks, particularly given that the Italo-Romance varieties were not that frequently used and were hence expected to be the non-dominant languages of our informants. Moreover, our research targets different areas and different types of speakers, so different varieties would have had to be chosen as the starting point of a translation task: the translations could have been performed from the local contact varieties (Spanish, Portuguese, French, and English) or from Italian, leaving the choice up to the speakers (for instance, elderly speakers who migrated as adults might have preferred Italian, while HSs might have had a preference for the local contact variety). However, a flexible format of this type, with different variables to accommodate all possible needs of our speakers, would have made the translations not fully comparable, as they would have primed the informants in different ways. We therefore decided to exclude translations.

Instead, we decided to structure our questionnaire as a two-alternative forced choice task. In this setup, the informants are asked to compare the acceptability of (a list of) pairs of stimuli by choosing, within each pair, the most acceptable item. Following a considerable number of studies on the issue (for a discussion and specific references, cf. Stadthagen-González et al. 2018), we judged that this format would be beneficial for our research in many respects, including the fact that it is less demanding to compare two items than to rate them on a predefined and consistent scale.

For each phenomenon, we identified a number of research questions on the basis of a preliminary review of the available literature. Variations on the two-alternative forced choice task described above were used to test these questions whenever possible and depending on the nature of the phenomenon: subject clitics, DOM, and, to some extent, deixis (with the support of pictures), were included; for deixis we added a semi-guided production task as well.

More information on how we paired tasks and phenomena and on how each task was designed and carried out will be provided in Section 3.2; in the remain-

der of this section, the specific conditions that were tested for each phenomenon will be explained in more detail.

**3.1.2.1 Subject clitics.** Subject clitics are found in most northern Italo-Romance and Rhaeto-Romance varieties. They differ from regular tonic subject pronouns in that they are syntactically deficient elements; they are inflectional heads, on a par with verbal agreement endings (Rizzi 1986, Brandi & Cordin 1989, Poletto 1993, 2000). However, Frasson (*In press*) found that in Brazilian Venetan subject clitics display pronominal behavior (see also Benincà & Poletto 2004). In our questionnaire we firstly tested the agreement-like or pronominal nature of subject clitics, checking:

- i. doubling contexts: an agreement marker obligatorily doubles an overt lexical subject, while a pronoun cannot occur when the subject is already expressed;
- ii. coordinated structures: an agreement marker is obligatorily realized in both conjuncts in coordinated structures, while a pronoun is generally realized only in the first;<sup>17</sup>
- iii. negation: agreement markers normally follow the preverbal negation marker, while pronouns precede it;
- iv. interpolation: agreement markers cannot be separated from the verb by non-clitic material, while pronouns can.

In addition, we checked for three more contexts that generally display some instability in younger speakers of Venetan in Italy (see Casalicchio & Frasson 2019). More precisely, these are:

- v. interrogative sentences: subject clitics are normally enclitic in interrogative sentences, but there is a tendency to realize them proclitically, on a par with declarative sentences;
- vi. impersonal constructions with meteorological verbs: the realization of subject clitics varies substantially in this context; often, both a sentence with a subject clitic and one without a subject clitic are accepted by speakers;

---

<sup>17</sup>As shown in Poletto (2000), the behavior of subject clitics in coordinated structures is very nuanced. Following her analysis, we avoided testing coordinated structures with two distinct inflected verbs that have the same nominal object, as well as coordinated structures involving the same verb with different tense or aspect specifications; subject clitics may behave differently in these types of coordinated structures.

- vii. default agreement constructions: as with *vi*, realization varies substantially in this context. Sentences with post-verbal subjects and restrictive relative clauses normally require a default third person singular agreement on the verb (and no subject clitic); however, speakers often also accept full agreement (with a subject clitic) in these contexts.

Despite not being strictly related to the agreement-like or pronominal nature of subject clitics, the contexts in *v-vii* were added in order to test the stability of heritage varieties with respect to varieties spoken in Italy.

**3.1.2.2 Deixis.** The main focus of our investigation with respect to deixis was the number of deictic contrasts encoded in demonstrative systems. Demonstrative forms and spatial adverbs anchor an object or an area in the external world to (one of) the discourse participants, by defining them in terms of the distance from the speaker and/or the hearer (e.g. *this* means that something is close the speaker, *there* means that an area is far away from the speaker).

Depending on how many discourse participants are available as possible anchoring points, different demonstrative systems are observed: if the only relevant reference point is the speaker, then we typically have a system that encodes a two-way contrast (an object or an area close to the speaker as opposed to an object or an area far from the speaker; e.g. Italian *questo* ‘this’ and *qui* ‘here’ as opposed to *quello* ‘that’ and *là* ‘there’). Since systems of this sort have two forms, they can be referred to as binary systems. However, it is also possible for the first term of a binary system to jointly refer to an object or an area close to both discourse participants, without any further specification as to who is closer to the referent, and conversely for the second term of a binary system to refer to an object or an area that is far from both discourse participants at the same time: this is the case, for instance, of Catalan *aquest* ‘this (close to the speaker and/or to the hearer)’, *aquell* ‘that (far away from the speaker and the hearer)’. If, instead, the hearer is also relevant in the spatial relations, the resulting system will encode a three-way contrast (an object or an area close to the speaker; an object or an area close to the hearer; or an object or an area far from both the speaker and the hearer). The Portuguese system is of this type, and differentiates between *este* ‘this’ and *aqui* ‘here’ (close to the speaker), *esse* ‘that’ and *ai* ‘there’ (close to the hearer), and *aquele* ‘that’ and *alá* ‘there’ (far from both). These systems display three contrastive forms and can therefore be defined as ternary systems. In the Romance domain there are also systems that do not encode any deictic contrast, i.e. that only display one form that can be used in different deictic contexts without yielding any difference in interpretation: this is the case for

French adnominal and pronominal demonstratives (*ce* and *celui*, respectively, in their masculine singular versions).

The literature on deixis in Romance varieties highlights a high level of microvariation (see, for the most extensive overviews, Ledgeway 2015 and Ledgeway & Smith 2016): Romance varieties display all four systems and there is significant variation, especially in the southern Italo-Romance domain. Therefore, we chose to investigate which deictic contrasts are encoded in Italo-Romance varieties in microcontact, by eliciting material related to the three possible deictic domains (close to the speaker, close to the hearer, far from the speech act participants), to test how these systems behave in contact and, ultimately, to better understand how these forms are encoded in the grammar.

**3.1.2.3 Differential Object Marking.** Differential Object Marking ('DOM'; Móravcík 1978, Bossong 1985, 1991), also known in the Romance literature as prepositional accusative (Diez 1874, Meyer-Lübke 1890–1902, 1895–1900), is the phenomenon whereby some Direct Objects ('DOs') are marked differently than others, depending on certain semantic and pragmatic features of the object. The phenomenon has different distribution patterns in Romance: some languages only display DOM with pronouns (e.g. some Eastern Abruzzese varieties, as in Manzini & Savoia 2005), while other varieties only display DOM with a subset of pronouns (e.g. Ariellese, as in D'Alessandro 2017). In other cases, DOM is only possible in clitic doubling contexts (e.g. in Piedmontese, Manzini & Savoia 2005), whereas it is linked to specificity and definiteness in southern Italo-Romance varieties (see Andriani (To appear), for Barese; Ledgeway 2009 for Neapolitan; Ledgeway et al. 2019 for Calabrian; Guardiano 2000, 2010 for some Sicilian varieties) as well as in Peninsular Spanish (Leonetti 2004). Conversely, in Argentinian Spanish DOM is strictly linked to Case (Saab 2018) and in Standard Italian it mostly marks Object Experiencers (see Belletti 2018 for a recent overview and discussion).

It should additionally be noted that the preposition marking DOs in these languages is the same as the preposition that introduces Indirect Objects ("IOs"), namely *a* (notice the contrast between Spanish DO *Veo a Juan* 'I see Juan' and the IO *Le doy el libro a Juan* 'I gave the book to Juan'). These differences have led to a lively discussion in the literature on what really triggers DOM, and whether DOM objects are true accusatives or datives.

The starting point of our investigation was that what we label as DOM might be referring to a range of different phenomena that just happen to share the same

superficial outcome.<sup>18</sup> We wanted to know whether this is actually the case and, if so, whether these differences are simply the product of diachronic evolution, or if they have in fact developed from different starting points in different languages. Furthermore, in the case of contact with Argentinian Spanish, we wanted to investigate whether a possible change in the distribution of DOM in Italo-Romance varieties reflects the change found in the contact variety.

### 3.2 What we did

#### 3.2.1 Assessing proficiency

The production tasks worked well with most informants: we asked them to tell us about their arrival in the Americas and what they found there (if they were G1), or in general about their childhood, parents and links to Italy. Our plan was to collect at least 5-10 minutes of spontaneous conversation, but some of the speakers were so happy to talk to us that they talked for half an hour or even an hour. Their willingness to speak to us may also have come from their knowledge that their recordings would be published (strictly anonymized) on the project's atlas, so they were happy that their story would reach a larger audience. Still, in some cases the informants felt awkward speaking the dialect when the fieldworker was not a speaker of the same language. In these cases, they often mixed it with Italian or with the language of their new country.

As mentioned in Section 3.1.1, we also performed an additional test to assess lexical proficiency on the basis of material designed by the HALA research group. In the HALA test, three sets of pictures (body parts, natural elements, and general pictures to create short sentences) are shown, and participants have to name the objects depicted as quickly as possible in the target language. Not only does this give an indication of vocabulary size, but also of speed of lexical access, both of which are indicators of language proficiency. The speed is measured by calculating the time lapse between the picture appearing on the screen (highlighted by an audio signal) and the participant's naming of that picture. However, for the test to be carried out successfully, it is necessary to compare the speed of lexical retrieval across the different languages in the participant's repertoire, to comparatively assess whether the specific times for a given language are linked to a genuine delay in retrieval (and hence to lower proficiency) or whether they are in line with the access times in other varieties, suggesting that longer times are simply due to external factors. Since the test had to be performed in different languages, we decided to only use a short part of the original HALA test: we used

---

<sup>18</sup>A detailed discussion of this question is included in Luana Sorgini's ongoing PhD dissertation.

six items in the Italo-Romance variety before the questionnaire and then asked participants to repeat the test in the language they felt to be their dominant one after the questionnaire.<sup>19,20</sup>

A complicating factor was the experimental setting: as already mentioned, we had to carry out interviews in unconventional locations, making it difficult to detect the signal sound that was intended to be the starting point in the calculation of the response times.

### 3.2.2 Designing and running the syntactic questionnaires

In the design of our questionnaires, we had to consider a number of constraints related to the status of the varieties under analysis, the specific differences among the syntactic phenomena considered in our study, and the type of population that we were targeting.

The first issue we faced in the design of our questionnaire was the fact that Italo-Romance languages are not standardized and, as such, many of them do not have an orthography and are mainly spoken. This is not the case for all of them: some of the varieties under investigation have a long written tradition and therefore a standardized spelling convention, but even then their written systems show microvariation, mirroring the actual linguistic microvariation found across the Italo-Romance domain. Once again, presenting the speakers with a slightly

---

<sup>19</sup>Despite our efforts to comply with the test requirements and with the non lab-based nature of our data collection, this set-up still does not meet the HALA guidelines. Ideally, the test should be performed in one language a day, and after having started the conversation in that specific language, so that the informant is in the “right” language mode. Clearly, this option was not available in our case.

<sup>20</sup>An anonymous reviewer points out that “In addition to assessing reaction time, etc., it seems careful attention needs to be paid to whether they are producing words in the ‘right’ variety – the regional variety whose proficiency you want to assess. This means KNOWING whether words borrowed from Italian are part of the vocabulary of the regional variety.” We do not see this as an issue at all. While fieldworkers were not native speakers of all varieties, they did know the various words in the different Italo-Romance varieties. Furthermore, the data were double-checked by native speakers after the return from fieldwork.

The reviewer also observes: “It’s also not clear to me why you could not follow the “one language a day” protocol. At the least, it’d be good to make sure the participant has been speaking in the relevant lg. for some time just before the test.” This first fieldwork session was mainly focused on checking the language profile of the speakers, rather than on eliciting data: thus, the fieldworkers did not have enough days in one location to follow the “one language a day” rule. Furthermore, after an hour of telling stories in their native Italo-Romance varieties, very often with other family members present, the speakers did not show particular problems with using the language.

different spelling system for their variety could have resulted in slight unfamiliarity with the stimuli, similar to what we might expect for varieties without conventionalized spelling mentioned above. Furthermore, the choice of one standard written variety over another might have triggered unwanted judgments on the spelling, rather than on the stimuli themselves. Another point that we had to consider is that most G1 speakers may have vision problems as well as issues with reading due to their age or to illiteracy.

Having ruled out the possibility of a written questionnaire, we were left with two possible options for an oral questionnaire: to lead the interview personally, or to use pre-recorded stimuli. Given that every interviewer had to test speakers of all varieties involved in our study, it would have been difficult if not impossible for the fieldworker to perform the interviews in all target varieties; an attempt to do so would have led to biased data. Therefore, we decided to have native speakers of each target Italo-Romance variety pre-record a set of stimuli in their variety and present our informants with those auditory stimuli. Nonetheless, when possible and whenever the interviewer and the informant spoke the same variety, that specific language was used throughout the whole interview.

Moreover, in New York, the language of interaction for the interviews was adapted to the speakers' relative confidence with the languages in question. G1 speakers who learnt English after their adolescence preferred being spoken to in Italian, while the remaining G1 speakers and all HSs preferred English. Whenever possible, the interviewer also used the relevant Italo-Romance variety to encourage the speaker not to switch to English or Italian. However, this strategy was not always successful, as Haller (1987: 394) also reports: "even though the interviews were conducted by Italian-Americans accepted in the community [... w]hen asked to switch to dialect, the informants generally continued to speak their high variety [(dialectal) Italian] after uttering a few dialect words, even if the interviewer was somewhat fluent in the specific dialect".

Some issues related to the phenomena under analysis further influenced our choice of tests. For subject clitics, a two-alternative forced choice task was the best way of identifying the agreement-like or pronominal behavior. Participants had to choose between two proposed sentences: one with a well-behaved agreement-like subject clitic and one without the clitic or with a clitic displaying anomalous behavior. This is shown, for instance, by the context of coordination:

(1) Friulian

- a. Al mangje e al bêf  
he.SCL eat-PRS.3SG and he.SCL drink-PRS.3SG

- b. Al mangje e bēf  
 he.scl eat-PRS.3SG and drink-PRS.3SG  
 'He is eating and drinking.'

The sentence in (1a) shows that the subject clitic is repeated in both conjuncts in a coordinated structure; this is expected, as subject clitics are obligatory agreement markers realized every time a finite verb appears. The sentence in (1b) shows that the marker is realized only in the first conjunct, which is taken to constitute pronominal behavior. The pronominal and agreement-like behavior were presented in a random order. Speakers heard the two stimuli one after the other and in random order, and had to choose which one they preferred.

The forced choice task proved successful for subject clitics in most cases: informants understood the task correctly. However, the spontaneous production task provided crucial support. Not only did it help to confirm – or otherwise – the results that we obtained through the questionnaire, but it allowed us to observe further aspects of the distribution of subject clitics that would otherwise have been left unnoticed. The most relevant example in this respect is the tendency to realize more overt pronominal subjects in heritage northern varieties in comparison to heritage southern varieties.

With regard to DOM, the forced choice task targeted the following range of direct objects, to determine whether they would trigger DOM:

- (2) 1<sup>st</sup> person pronoun > 3<sup>rd</sup> person pronouns [+human] > kinship >  
 [+human][+animate][-definite] > [-human][+animate][+definite] >  
 [-human][+animate][-definite]

This order reflects Silverstein's (1976) animacy scale, since the general understanding of DOM in Italo-Romance varieties is that the higher the object is on the scale, the more likely it is to be marked.

These objects were tested both in situ and in fronted topic position (Rizzi 1997). Speakers of the southern and northern groups had two slightly different questionnaires. Informants of the southern varieties had 13 sentences testing DOM plus fillers, for a total of 24 sentences. Speakers of northern varieties were given 9 sentences testing DOM plus fillers for a total of 23 sentences. We made this decision because we were not expecting production of DOM on a wide range of arguments by speakers of northern varieties, as the equivalent varieties spoken within Italy are not typically considered to have DOM.

The informants were asked to choose between a sentence including DOM and one without: these stimuli were presented in random order. Although speakers

needed guidance when taking the questionnaire (e.g. sometimes they had difficulty understanding certain lexical items in the stimulus due to microvariation in the lexical entry), and a translation had to be provided, the test worked in most cases and revealed differences in the use of DOM with respect to the homeland varieties.<sup>21</sup> In some cases, when informants deemed the first sentence correct, they confirmed it before listening to the second sentence. In these cases we had to ask informants to wait until they heard both sentences before deciding between the two options.

For deixis, we decided to avoid grammaticality judgments, sentence completion and elicited imitation, as demonstratives heavily rely on the context in which the conversation takes place. In fact, demonstratives are always grammatical, but they carry semantic differences that make them more or less suitable for a given context: different forms are used in different contexts, and this choice may depend on other indexical properties of the sentence as well. In grammaticality judgments, it is rather difficult to recreate such a context.

Although sentence completion and elicited imitation are typically not bound to any context, they raise other issues for investigating deixis. In both these task types, the target form can show a mismatch with the elicited form because of the switch in the deictic center at the conversation turn. For instance, in the case of elicited imitation, the informant might switch the deictic center when repeating the sentence, e.g. *I am here > You are there*. While both sentences are equally grammatical, they change in their interpretive content, but this is not tested (or indeed testable) in an elicited imitation task.

To circumvent these issues, we selected a picture-sentence matching task and a semi-spontaneous production task. For the former, we presented our informants with some pictures of dog owners and their dogs; one of the dog owners was marked as the speaker with the help of a balloon.

Our informants had to identify themselves with the speaking character and refer to the dog present in the context of the picture (Figure 2: a, b, or c) by choosing one of either two or three (depending on the system in the target variety) recorded audio stimuli associated with each picture. For instance, given Figure 2(a) with a dog owner holding their dog and another person (the hearer) on the other side of the picture, and given the dialectal audio stimuli for ‘This (close to me) is my dog’, ‘That (close to you) is my dog’ (if available in the target variety), and ‘That (far from us) is my dog’, the target item would have been ‘This

---

<sup>21</sup>An anonymous reviewer asks: “What if they prefer the non-DOM form because it’s more like std. Italian?”. In northern varieties, the choice of the non-DOM option is to be attributed to the absence of the phenomena in the dialect rather than to the influence of Standard Italian. We did not find a consistent preference for the non-DOM option in southern varieties.

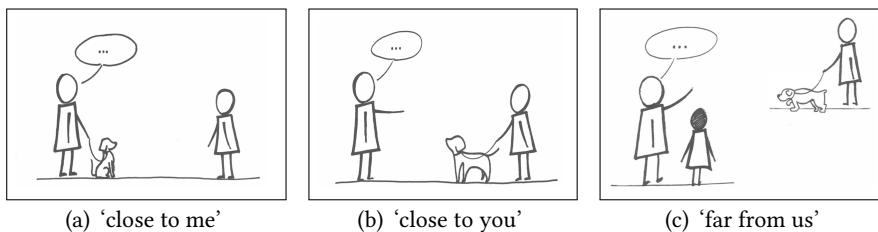


Figure 2: Picture-sentence matching task.

(close to me) is my dog', i.e. the proximal demonstrative *this*.<sup>22</sup> The stimuli were also presented in random order for this task.

This set-up was not without problems: most importantly, some of our informants found it particularly hard to identify themselves with the speaker in the picture; similarly, some participants found it difficult to understand that the speaker actually had an interlocutor inside the picture itself. Instead, some informants selected one of the audio stimuli on the basis of where the dog was in relation to them: given that the stimuli were presented on a laptop screen and that the screen was within their arm reach, they tended to point at or touch the dog and identify it as 'this', in any context, even the distal one. Moreover, in consideration of possible vision difficulties, the main characters on the picture were very large, which resulted in the picture itself being quite cramped and the distance between the characters to be overall too reduced: specifically, the 'close to you' space could easily be reduced to the 'close to me' one, as the speaker and the hearer were only a small distance apart. The informants sometimes explained their answer by saying: "it's still close, if it were *that* it would be something else". These size considerations, together with the identification problem, led to an overall higher rate of proximal forms even in non-proximal contexts. However, responses changed substantially when real-life situations were investigated. One such method used to elicit the (actual) distal forms was the question "Would you still use *this* if the dog that you see was on the other side of the street?". Still, no specific protocol for these cases was agreed before the fieldwork, so the data collection was, in this respect, not uniform, and the results not completely trustworthy.

Semi-guided production proved a better test for deixis: in this case, we used three pictures of cats of different colors: black, orange, and white. These pictures

<sup>22</sup>This was the target sentence for the pronominal context. Other syntactic contexts tested were adnominal (e.g. 'This dog is mine'), and demonstrative-reinforcer ('This here is my dog').

were placed either near the informant (the speaker), near the interviewer (the hearer), or far from both. Our informants were then asked where each cat was in the context, to which they had to reply with a demonstrative form or with a spatial adverb. We judged that the actual contrast within the context would make this task easier to perform for our speakers: they effectively needed to choose different demonstratives to make us understand which cat they meant. However, this method was also far from perfect: the most significant issue that we encountered was how to elicit the demonstrative or spatial adverb, rather than a description of the image or of its location with respect to other objects in the room (e.g. ‘the one on the chair’, rather than ‘that one’). To try to elicit the target response, we sometimes suggested the whole set of answers in the contact language to help the informants to understand, without priming the language-specific demonstrative system.

One last issue that arose in the preparation of the deixis questionnaire was the clear difference between the tasks. For SCLs and DOM the task was comparable and we could use the sentences targeting SCLs as fillers for those targeting DOM and vice versa; this kept the questionnaire to a minimum length so as not to tire our elderly informants, but still ensured the quality of our investigation. It was impossible, however, to run a comparable task for the deixis part of the questionnaire. Yet, it would have been ideal to show some sets of filler pictures targeting other phenomena alongside the images in Figure 2, which would also have had the benefit of making the task less repetitive. While designing the task, we thought that the addition of fillers would have been an online confounding factor (the informants would have had to correctly interpret multiple scenes) and would also have been time-consuming, particularly given that we were trying to design a questionnaire that targeted all phenomena at once, while still being of a manageable size. However, upon testing, we realized that the absence of variation in the referent (always a dog, although in different positions in the picture and in different syntactic contexts: pronominal, adnominal, demonstrative-reinforcer construction) made the test extremely repetitive, which resulted in complaints from the participants, who thought that they were being asked the same question over and over again. Variation in the referent could have been beneficial to the task.

### **3.3 General issues concerning experimental design and statistics**

In an ideal world, all our participant groups would have had an equal number of participants, who would all have spoken the exact same local varieties, and all

possible variables would have been perfectly controlled for. Moreover, all participants would carry out exactly the same task in exactly the same way. However, due to the scarcity and the heterogeneous make-up of our target populations, as well as problems that arose during the fieldwork, this was not possible. While we must accept that no research is perfect, it is important to be aware of the possible consequences of these issues for the interpretation of the results.<sup>23</sup>

Regarding the characteristics of the participants, it is clear from the description presented in Section 2 that they were not evenly distributed across the different varieties, host countries and generations. This must be taken into account when analyzing the results, particularly for the purposes of statistical analysis. For instance, there were eight speakers of Abruzzese in Argentina, only two in Canada and none in Brazil. Two speakers is too small a number to be able to perform any statistical analyses, so for this variety, we were only able to statistically model the linguistic behavior of the speakers in Argentina. In addition, of these eight speakers of Abruzzese in Argentina, two were G1, and six were second-generation HSs. Again, given that two speakers cannot really constitute a separate subgroup, it was impossible to take “generation” into account as a variable in the statistical analysis. All 8 speakers were therefore treated as belonging to the same group, whereas in fact there was an important difference, namely that some of them were immigrants and others were born in the host country.

Moreover, as mentioned above, there were differences between communities in terms of literacy, education level, exposure to other languages, etc. While it is impossible to completely control for these variables in this type of study, it is important to keep their impact in mind when analyzing the results. For instance, we found certain differences between the use of SCLs by speakers of heritage Friulian in Argentina and Brazil. While an initial interpretation of a difference of this sort might be that there is an effect of the contact language (Spanish and Brazilian Portuguese differ in terms of their configuration of the pro-drop parameter), there were other differences between the communities. First, as mentioned, the communities in Brazil tend to be more isolated and the HLs therefore tend to be better preserved. Moreover, HSs in Argentina were mostly second-generation speakers while those in Brazil were almost exclusively third generation.

The design and the execution of our tasks was less than ideal from the perspective of experimental validity. The materials, i.e. the specific sentences for each of the phenomena, were selected with specific research questions in mind. In order to reduce the length of the questionnaire, in most cases, only one sentence (pair)

---

<sup>23</sup>Recall, however, that this article is not concerned with these questions themselves, but is primarily reporting on the first fieldwork exercise, which had a mainly descriptive aim.

per condition (sentence type) was used, with the understanding that we would return to carry out more extensive and targeted second fieldwork. Another issue that should be taken into account is that for some of the phenomena, all the sentences were presented together, without filler/distractor items. This may have made the participants aware of the topic of investigation, which could in theory have led participants to use specific answer strategies (for instance: always picking the sentence with DOM). We chose this set-up to avoid having to stop the questionnaire half way through due to speaker tiredness.

Finally, as mentioned above, some of the interviewers had to improvise, either because the informants did not understand the task, or because they did not have enough time to perform the complete questionnaire. This affected the uniformity of the study in various ways. For instance, not all participants answered an equal number of questions for each of the phenomena, reducing comparability across participants and/or groups. Another issue is that some of the researchers carried out the experiment in the dialects, whereas others did so in the contact language or in standard Italian. It has been noted (Aalberse & Muysken 2013) that the specific language spoken by the researcher may affect the respondents' linguistic behavior. The task type was also sometimes adapted on-the-go by the researcher. For instance, for those respondents who did not understand the forced choice task, it was sometimes (orally) adapted to a translation task. Similarly, in the guided production task used for deixis, some of the researchers chose to present the participants with the full set of options for demonstratives in the contact language, which may have led to a higher number of target responses for those participants.

### **3.3.1 Interviewing the elderly**

The main target speakers of this project are first-generation emigrants, who are quite elderly. The average age of G1 speakers was around 75. This brings with it additional issues that we considered before fieldwork; however, we found that it had a larger impact on the results than expected. Advanced age brings a number of common issues, such as partial or complete loss of hearing and sight, which we tried to take into account when designing the questionnaire, while still respecting the constraints imposed on us by the different phenomena.

A further problem is the difficulty of retaining long sentences: we therefore tried to keep the stimuli as short as possible. Furthermore, while this was true of many younger speakers too, many elderly speakers had clear difficulties with the very concept of choosing between two options: rather, they would approve of the very first stimulus out loud, regardless of its grammaticality and without

listening to the second one.<sup>24</sup> When it was impossible to provide more instructions that would help them to complete the task as originally planned, they were given sentences in the contact language, and they were asked to translate them into the dialect.

### 3.3.2 Tips and warnings

Here are some tips to design a good questionnaire for heritage speakers:

- Make sure the stimuli are culturally appropriate.

Verbs like *kissing* (a man or a woman) or *liking someone* may create uneasiness, especially among elderly speakers. Moreover, it is not a good idea to record sentences that imply any act of *killing*, as the speakers could be afraid that a recording containing such a statement could be taken to reflect reality.

- Make sure you have some spontaneous data.

It is always a good idea to compare the questionnaire responses to some spontaneous data. If that is not possible, e.g. because the speakers are not comfortable speaking in their non-dominant language without a predefined topic, some (controlled) production tasks can help (and can make the collected data more comparable). Remember that spontaneous speech is very useful in assessing proficiency, too. If the elicited data contradict the spontaneously produced data, they should be excluded (at least, we chose to exclude them).

- Use the target variety when possible, to minimize the interference of other languages.

- If the questionnaire is too long and the researchers wish to test at least two phenomena, use the questions for the other phenomenon as fillers in your questionnaire.

- Agree in advance what should be done if the task does not go as expected. If you are running several parallel fieldwork sessions, it is not always possible to exchange experiences with your colleagues and solve unexpected situations in a uniform way. It is therefore advisable to think about the

---

<sup>24</sup>An anonymous reviewer points out that heritage speakers have difficulty in giving acceptability judgments. We are aware of that, and that is precisely why we went for forced choice and sentence completion tasks rather than the classical generative method of “is this sentence acceptable for you?” which we knew would not provide results.

main issues that may arise beforehand (e.g.: a participant does not understand the task) and to devise a protocol on how to proceed in those cases, in order to limit the degree of unwanted variation.

- Carry out a pilot study when possible.

Before starting your fieldwork, it is a good idea to perform a pilot version of your questionnaire/experiment with speakers who are comparable in age and other sociolinguistic factors to those of your target population. This might highlight some issues that can be improved upon before the actual fieldwork.

- Try to avoid priming.

While this is true for all speakers, elderly informants seem to be more prone to just repeating what they have heard last (or to listen to just one stimulus) or to what the fieldworker suggested as an example. It is therefore important to be careful to avoid priming wherever possible while explaining the task.

- Pay extra attention to the design of your stimuli if you are planning to interview elderly people.

Elderly people may present some challenges that are linked to their age: hearing and sight issues, longer processing times and more expensive processing overall, weaker short-term memory, lower attention span, etc. You should keep these factors in mind when designing your questionnaire, and specifically: use short questions, both for the short-term memory and to limit exposure time; make sure that your stimuli are fully accessible (the volume of audio stimuli must be loud enough; if written stimuli are chosen, the font size should be fairly large). It is a good idea to split a long questionnaire into two parts and test them separately.

- Be ready to get more involved with the community, especially when testing elderly speakers.

As elderly speakers can be suspicious, especially when using modern technology such as recording equipment, make sure there is always a relative present, if possible, or another member of the community who can assist the person and reassure them that you are not doing anything inappropriate. Also, be ready to spend more time with your informants than you were planning to: some of them are lonely and really enjoy company, and they especially appreciate the opportunity to speak to younger people from their home country. Having an hour-long recording of spontaneous speech

is great for your research, but may be problematic if you have scheduled another interview shortly after the current one.

## 4 Closing remarks

In this article, we have tried to highlight all the information we collected and all the things we learned when setting up and carrying out fieldwork relating to heritage Italo-Romance speakers in North and South America and Europe. While many of these tips and much of this information can be found in general manuals or fieldwork reports, some are specific to the Italo-Romance community.<sup>25</sup> Furthermore, we provided a description of the status of these varieties, many of which had not been documented since the 1960s. When we did have some documentation of previous stages of the languages, we compared that to what we found, and showed that the situation has changed considerably. With the exception of the northern Italian-speaking community, most Italo-Romance heritage varieties in America are close to extinction: for this reason, documenting these languages is now all the more important. While this paper only draws some practical conclusions, there is much more to the study of heritage Italo-Romance in the Americas and we hope that our remarks will be helpful to researchers willing to undertake similar investigations.

## Abbreviations

PRS present      SCL subject clitic      SG singular

## Acknowledgements

We would like to express our gratitude to the informants who participated in our study and all our contacts in the Americas and Belgium who helped us reach the communities and supported the fieldworkers. We would also like to thank two anonymous reviewers for useful comments and suggestions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement: CoG 681959\_MicroContact).

---

<sup>25</sup>For other considerations specific to Italo-Romance varieties, leaving aside the different field-work setting with respect to our study, see Cornips & Poletto 2005.

## References

- Aalberse, Suzanne, Ad Backus & Pieter Muysken. 2019. *Heritage languages: A language contact approach*. Amsterdam: John Benjamins Publishing.
- Aalberse, Suzanne & Pieter Muysken. 2013. Perspectives on heritage languages. Position paper for the Leeuwenhorst Workshop on Heritage Languages, January 23-25th 2013, Nijmegen: Radboud University.
- Andriani, Luigi. To appear. Differential Object Marking in Barese: Old and new challenges. In Monica-Alexandrina Irimia & Alexandru Mardale (eds.), *Differential Object Marking in Romance: Towards microvariation*. Amsterdam: John Benjamins Publishing.
- Andriani, Luigi, Roberta D'Alessandro, Alberto Frasson, Brechje van Osch, Luana Sorgini & Silvia Terenghi. 2020. *The ranking of internal and external factors in heritage language syntactic change*. Paper presented at the 43rd Generative Linguistics in the Old World (GLOW) conference, HU Berlin/online, April 2020. DOI: <https://osf.io/es53m/>.
- Bagna, Carla. 2011. America Latina. In Massimo Vedovelli (ed.), *Storia linguistica dell'emigrazione italiana nel mondo*, 305–357. Roma: Carocci.
- Belletti, Adriana. 2018. On a-marking of object topics in the Italian left periphery. In Roberto Petrosino, Pietro Cerrone & Harry van der Hulst (eds.), *From Sounds to Structures: Beyond the Veil of Maya*, 445–466. Berlin/Boston: De Gruyter. DOI: 10.1515/9781501506734-016.
- Benincà, Paola & Cecilia Poletto. 2004. Topic, Focus and V2: Defining the CP sublayers. In Luigi Rizzi (ed.), *The structure of CP and IP*, 52–75. Oxford/New York: Oxford University Press.
- Bettoni, Camilla. 1990. Italian language attrition in Sydney: The role of dialect. In Michael Halliday, John Gibbons & Nicholas Howard (eds.), *Learning, keeping and using language*, vol. 2, 75–89. Amsterdam: John Benjamins Publishing.
- Bettoni, Camilla. 1991. Language shift and morphological attrition among second generation Italo-Australians. *Rivista di linguistica* 3(2). 369–387.
- Bossong, Georg. 1985. *Differentielle Objektmarkierung in den Neuiranischen Sprachen*. Tübingen: Narr Verlag.
- Bossong, Georg. 1991. Differential Object Marking in Romance and beyond. In Dieter Wanner & Douglas A. Kibbee (eds.), *New analyses in romance linguistics: Selected papers from the XVIII Linguistic Symposium on Romance Languages, Urbana-Champaign (April 7–9 1988)*, 143–170. Amsterdam: John Benjamins Publishing.

## 2 Documenting Italo-Romance minority languages in the Americas

- Brandi, Luciana & Patrizia Cordin. 1989. Two Italian dialects and the Null Subject Parameter. In Osvaldo Jaeggli & Kenneth J. Safir (eds.), *The Null Subject Parameter*, 111–142. Dordrecht: Springer.
- Casalicchio, Jan & Alberto Frasson. 2019. Cambiamenti nell’uso dei soggetti clitici veneti: Il ruolo del contatto con l’italiano. In Annamaria Chilà & Alessandro De Angelis (eds.), *Capitoli di morfosintassi delle varietà romanze d’Italia: Teoria e dati empirici*, 193–258. Palermo: Centro di studi filologici e linguistici siciliani.
- Cavaioli, Frank J. 2008. Patterns of Italian immigration to the United States. *The Catholic Social Science Review* 13. 213–229.
- Cenni, Franco. 2003. *Italianos no Brasil: “andiamo in ‘Merica”*. São Paulo: EdUSP.
- Cornips, Leonie & Cecilia Poletto. 2005. On standardising syntactic elicitation techniques. Part I. *Lingua* 115. 939–957.
- D’Alessandro, Roberta. 2017. When you have too many features: Auxiliaries, agreement and clitics in Italian varieties. *Glossa: A Journal of General Linguistics* 2(1). 50. DOI: <http://doi.org/10.5334/gjgl.102>.
- D’Alessandro, Roberta. 2021. Syntactic change in contact: Romance. *Annual Review of Linguistics* 7(1). 309–328. DOI: [10.1146/annurev-linguistics-011619-030311](https://doi.org/10.1146/annurev-linguistics-011619-030311).
- De Fina, Anna. 2014. Language and identities in US communities of Italian origin. *Forum Italicum* 48(2). 253–267.
- De Fina, Anna & Luciana Fellin. 2010. Italian in the USA. In Kim Potowski (ed.), *Language diversity in the USA*, 195–205. Cambridge: Cambridge University Press.
- Diez, Friedrich Christian. 1874. *Grammaire des langues romanes*. Vol. II, III. Paris: Franck.
- Ferrari, Andrea. 2008. *Aspetti socio-culturali dell’emigrazione italiana in Argentina: Il caso di Santa Fe*. University of Turin. (MA thesis).
- Frasson, Alberto. In press. Clitics are not enough: On agreement and null subjects in Brazilian Venetan. *Linguistic Analysis* 43(1–2).
- Frasson, Alberto, Roberta D’Alessandro & Brechje van Osch. In press. Subject clitics in microcontact. A case study from heritage Friulian in Argentina and Brazil. *Heritage Language Journal*.
- Guardiano, Cristina. 2000. Note sull’oggetto diretto preposizionale nel siciliano. *L’Italia Dialettale* 51. 1–35.
- Guardiano, Cristina. 2010. L’oggetto diretto preposizionale in siciliano. Una breve rassegna e qualche domanda. In Jacopo Garzonio (ed.), *Quaderni di lavoro ASIt (Studi sui dialetti della Sicilia)*, vol. 11, 83–101.
- Haller, Hermann W. 1987. Italian speech varieties in the United States and the Italian-American lingua franca. *Italica* 64(3). 393–409.

- Haller, Hermann W. 1991. Atteggiamenti linguistici nelle comunità italo-americane. *Rivista di Linguistica* 3(2). 389–405.
- Haller, Hermann W. 1993. *Una lingua perduta e ritrovata: L’italiano degli italo-americani*. Florence/Rome: La Nuova Italia.
- Haller, Hermann W. 1997a. Dialects abroad. In Martin Maiden & Mair Parry (eds.), *The dialects of Italy*, 401–411. London: Routledge.
- Haller, Hermann W. 1997b. Italian. In Joshua A. Fishman & Ofelia García (eds.), *The Multilingual Apple. Languages in New York City*, 119–142. Berlin/New York: De Gruyter Mouton.
- Haller, Hermann W. 2002. I dialetti italiani negli USA. In Manlio Cortelazzo, Giannrenzo Clivio, Nicola De Blasi & Carla Marcato (eds.), *Dialetti italiani: Storia, struttura, usi*, 1083–1086. Turin: UTET.
- Ledgeway, Adam. 2009. *Grammatica diacronica del napoletano*. Tübingen: Max Niemeyer Verlag.
- Ledgeway, Adam. 2015. Varieties in Italy. In Konstanze Jungbluth & Federica Da Milano (eds.), *Manual of deixis in Romance languages*, 146–162. Berlin: De Gruyter.
- Ledgeway, Adam, Norma Schifano & Giuseppina Silvestri. 2019. Differential object marking and the properties of D in the dialects of the extreme south of Italy. *Glossa: A journal of general linguistics* 4(1). 51. 1–25.
- Ledgeway, Adam & John Charles Smith. 2016. Deixis. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 879–896. Oxford: Oxford University Press.
- Leivada, Evelina, Roberta D’Alessandro & Kleanthes Grohmann. 2019. Eliciting big data from small, young, or non-standard languages: 10 experimental challenges. *Frontiers in Psychology* 10. 313. DOI: 10.3389/fpsyg.2019.00313.
- Leonetti, Manuel. 2004. Specificity and Differential Object Marking in Spanish. *Catalan Journal of Linguistics* 3. 75–114.
- Manzini, Maria Rita & Leonardo Maria Savoia. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell’Orso.
- Maurizio, Roxana. 2008. Migración y desarrollo: El caso de Argentina. In Andrés Solimano (ed.), *Migraciones internacionales en América Latina. Booms, crisis y desarrollo*, 75–184. Fondo de Cultura Económica.
- Meo Zilio, Giovanni. 1990. *Estudios hispanoamericanos*. Vol. 1: Temas lingüísticos. Roma: Bulzoni.
- Meyer-Lübke, Wilhelm. 1895–1900. *Grammaire des langues romanes*. Paris: Welter.
- Meyer-Lübke, Wilhelm. 1890–1902. *Grammatik der romanischen Sprachen*. Leipzig: Fues.

- Moravcsik, Edith A. 1978. On the case marking of objects. In Joseph H. Greenberg (ed.), *Universals of human language*, vol. 4. Syntax, 249–289. Stanford: Stanford University Press.
- O’Grady, William, Amy J. Schafer, Jawee Perla, On-Soon Lee & Julia Wieting. 2009. A psycholinguistic tool for the assessment of language loss: The HALA Project. *Language Documentation & Conservation* 3. 100–112.
- Pascual y Cabo, Diego. 2013. *Agreement reflexes of emerging optionality in heritage speaker Spanish*. University of Florida. (Doctoral dissertation).
- Poletto, Cecilia. 1993. *La sintassi del soggetto nei dialetti italiani settentrionali*. Padova: Unipress.
- Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. Oxford: Oxford University Press.
- Polinsky, Maria. 2018. *Heritage languages and their speakers*. Cambridge: Cambridge University Press.
- Reinke, Kristin. 2014. Language contact in a multilingual setting. The attractive force of Italo-Romance dialects on Italian in Montreal. In Juliane Besters-Dilger, Cynthia Dermarkar, Stefan Pfänder & Achim Rabus (eds.), *Congruence in contact-induced language change*, 149–167. Berlin: De Gruyter.
- Rizzi, Luigi. 1986. On the status of subject clitics in Romance. In Osvaldo Jaeggli & Carmen Silva-Corvalán (eds.), *Studies in Romance linguistics*, 391–420. Dordrecht: Foris.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar. Handbook in generative syntax*, 281–337. Dordrecht: Springer.
- Saab, Andres. 2018. Case conflicts and A-movement: DOM, doubling and clitics in Rioplatense Spanish. Lecture series held at Utrecht University (December 2018).
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert W. M. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Sorgini, Luana. 2020. *Differential object marking in heritage Romance varieties*. Paper presented at the 1st Heritage Language Syntax workshop, Utrecht/online, 12–16 October 2020. DOI: <https://osf.io/g7k2n/>.
- Stadthagen-González, Hans, Luis López, M. Carmen Parafita Couto & C. Alejandro Párraga. 2018. Using two-alternative forced choice tasks and Thurstone’s law of comparative judgments for code-switching research. *Linguistic Approaches to Bilingualism* 8. 67–97. DOI: 10.1075/lab.16030.sta.

- Terenghi, Silvia. 2020. *Indexicality is not affected by contact. Ternary demonstrative systems in Italo-Romance HLs*. Paper presented at the 1st Heritage Language Syntax workshop, Utrecht/online, 12–16 October 2020. DOI: 10.17605/osf.io/5tusx.
- Villata, Bruno. 2010. *L’italianese – L’italiano comune parlato a Montreal*. Montreal: Lòsna & Tron.

# Chapter 3

## Spanish-Nahuatl bilingualism in Indigenous communities in Mexico: Variation in language proficiency and use

Justyna Olko<sup>a</sup>, Szymon Gruda<sup>a</sup>, Joanna Maryniak<sup>a</sup>, Elwira Dexter-Sobkowiak<sup>a</sup>, Humberto Iglesias Tepec<sup>b</sup>, Eduardo de la Cruz<sup>a,c</sup> & Beatriz Cuahutle Bautista

<sup>a</sup>University of Warsaw <sup>b</sup>Instituto de Educación Media Superior de la Ciudad de México <sup>c</sup>Instituto de Docencia e Investigación Etnológica de Zacatecas

The focus of this paper is bilingualism in Spanish and Nahuatl from the sixteenth century until the present day, with an exploration of its scope, functions and stability. We include a historical perspective to provide the necessary background for the contemporary context, which is approached with both qualitative and quantitative data acquired during fieldwork carried out in four different regions where Nahuatl and Spanish bilingualism is present today. Of special importance for the present study is the analysis of the results of proficiency assessment in both languages, performed with the participation of members of selected Nahua communities, which represent different degrees of assimilation to Mexican identity and shift to Spanish. We conclude that due to power differentials, economic, sociopolitical and cultural pressures and discriminatory language policies, contemporary Spanish-Indigenous bilingualism at the community level is unstable and transitional.

### 1 Introduction: Goals of the present study

The origins of Nahuatl-Spanish bilingualism go back to first encounters between Europeans and Indigenous people living in the area controlled by the Aztec Empire in 1519. At first largely limited to individual bilingual specialised skills, the

Justyna Olko, Szymon Gruda, Joanna Maryniak, Elwira Dexter-Sobkowiak, Humberto Iglesias Tepec, Eduardo de la Cruz & Beatriz Cuahutle Bautista. 2022. Spanish-Nahuatl bilingualism in Indigenous communities in Mexico: Variation in language proficiency and use. In Matt Coler & Andrew Nevins (eds.), *Contemporary research in minoritized and diaspora languages of Europe*, 57–78. Berlin: Language Science Press.  
DOI: 10.5281/zenodo.7446959



contact between the two languages and the growing pressures within the colonial system and then under the independent Mexican state (1821) gradually led to the appearance of societal bilingualism. This was particularly the case in multi-ethnic urban contexts during this time. In the twentieth and twenty-first century, it was also attested in rural areas. While we give importance to the historical perspective underlying current developments, the main focus of this paper is the bilingual situation of four different regions where unstable Nahuatl and Spanish bilingualism is present today. The study is based on the analysis of qualitative and quantitative data acquired during fieldwork, supplemented by historical sources.

By unstable bilingualism we mean the situation of parallel acquisition and use of the heritage language and Spanish, not exceeding two-three generations and leading, eventually, to language shift. This can be contrasted with the notion of stable bilingualism, when two languages are used – perhaps in a complementary manner and not necessarily at the same level of proficiency – for an extended period of time without either of the languages displacing the other one; such a situation has been observed e.g. in Quebec (French and English), Belgium (Dutch and French, especially in Flanders), Paraguay (Spanish and Guarani) or in many states of India (regional languages and Hindi, and, to a lesser extent, English). Furthermore, we use the term “individual bilingualism” to refer to the capacity of particular individuals, such as translators, friars, Indigenous notaries and other members of local communities, to speak two languages proficiently. Consequently, we use the term “societal bilingualism” to refer to a widespread use of two languages by significant parts of speech communities for whom this kind of linguistic practice is part of everyday interaction and not a specialised skill.

## **2 Historical context: The Colonial Period**

The Spanish colonisation of Mesoamerica, initiated by the landing of Hernán Cortés and his people on the shore of the Gulf of Mexico in the spring of 1519, created the urgent need for the development of bilingual and multilingual skills in Spanish and local languages. During the first stage of contact, the availability of Spanish-Indigenous translators was the greatest need, but with the ongoing colonisation and organisation of the European rule, the demand for individual Indigenous-Spanish bilingualism grew on both sides. Among all of the Indigenous languages spoken in sixteenth-century Mesoamerica, it was Nahuatl, belonging to the Uto-Aztec family, that was most frequently used between local populations and Spaniards and in bilingual arrangements with Spanish speakers,

e.g. in administration, courts or in efforts at Christianisation. It was also the language most often used by other non-Nahua ethnic groups, no doubt due to its central role in the Aztec Empire and other powerful states of earlier Pre-Hispanic Mesoamerica.

The Aztec imperial infrastructure collapsed and disintegrated rapidly after the arrival of the Spaniards, but the local organisation of Nahua states, called *altepetl*, proved resistant to conquest and colonisation. During initial attempts to introduce their rule, the Spanish had to rely on local Indigenous organisation structures, which meant dealing directly with particular *altepetl*. Such interactions contributed to the survival of preexisting entities and political-territorial units, ensuring their continued importance in the early colonial period (Gibson 1964: 63–74, Lockhart 1992: 28–29). In large urbanised zones, such as the capital city of México-Tenochtitlán, an organisational duality was introduced, with parallel Indigenous and Spanish municipal structures and organisations. Newly founded centers for European populations, such as the town of Puebla de los Angeles, tended to replicate Spanish structures more closely, despite housing significant portions of the Indigenous population.

This sociopolitical and sociolinguistic situation contributed to the Spaniards' reliance on Nahuatl as the language of administration and religious instruction. Moreover, beyond the core area of New Spain, Nahuatl was widely used as a vehicular language by other Indigenous groups, including in southern and northern Mesoamerica. The linguistic landscape and associated power relationships were complex and varied between regions, depending on the numbers and kinds of languages spoken in the areas and the extent of local multilingualism. From the sixteenth century on, different forms of polyglossia existed in New Spain: some were brought over from Spain – including Latin, high Spanish and low Spanish – and some originated in the Indigenous world. Thus, Latin and Spanish were high varieties in New Spain, while Nahuatl occupied a higher position with regard to other Indigenous languages (Parodi 2010: 308–310). Nahuatl therefore served different purposes, including that of an intermediary language in translation between Spanish and other local languages, as well as being, to a certain degree, the language of direct communication between Spaniards and local Indigenous populations (Nesvig 2012, Schwaller 2012). As a result, the uses of Nahuatl in colonial New Spain were by no means limited to members of Nahua communities (Yannanakis 2012: 669–670, Nesvig 2012: 739–758). For example, a large number of mestizos and creoles learned the language due to continuous daily contact with the speakers in households where Indigenous servants and workers were employed (Parodi 2010: 334).

While it seems unquestionable that extensive multilingualism in local languages existed in the early and mid-colonial period, the scale of societal bilingualism with Spanish and its geographical extent is a much more contentious issue. The combined evidence from Spanish and Indigenous sources points to what could have been, in the first phase of the colonial period, a partial and elitist bilingualism present among Native nobles and Spanish friars and clerics, accompanied by incipient and growing general bilingualism in the second part of the colonial period between Indigenous populations in specific regions and higher social groups (Zimmermann 2010: 945, Tab. 15). More debatable are other phenomena proposed for the colonial period (see Zimmermann 2010: 945), such as a notable increase of Indigenous people monolingual in Spanish. However, the available evidence suggests that this situation varied greatly, depending on the particular setting. Most notably it would have applied to large urban contexts, where societal bilingualism does not seem to have been particularly stable, thus giving way to Spanish monolingualism across several generations. This scenario seems to be supported by the decreasing number of documents in Nahuatl in the late seventeenth and eighteenth centuries in cities such as México-Tenochtitlán, Toluca or the even more ethnically homogeneous Tlaxcala, where Indigenous legal matters were increasingly conducted in Spanish. However, monolingual speakers of Nahuatl were still present in numerous communities in Tlaxcala until the second half of the twentieth century, including in some of the communities located close to highly urbanised areas, so widespread bilingualism toward the end of the colonial period in this area seems rather improbable.

The areas that most favored bilingualism and multilingualism in Indigenous languages and Spanish were large towns with already existing local populations, or those attracting immigrant labor. Although the legal divisions of “mixed” towns into Spanish and Indigenous municipalities created a jurisdictional and administrative separation between these ethnic groups, these divisions were not impermeable, and the process of mixing and contact between Natives and Spaniards contributed to the development of different levels of bilingualism and/or multilingualism. Interesting pieces of evidence on inter-ethnic contact and relationships come from the capital city of México-Tenochtitlán. By 1612, approximately 80,000 Indigenous persons reportedly lived in this city, as well as some 50,000 persons of African and mixed African-Indigenous origin, and about 15,000 “Spaniards” (including creoles) (Nutini & Isaac 2009: 34). In Zacatecas in the north and Puebla de los Ángeles to the south of the Valley of Mexico, both of which were formally established as “Spanish” towns serving the purposes of colonial settlers and businesses, the cities’ migrant and locally born populations of Indigenous people and Africans outnumbered their Spanish counterparts. Already in the

1570s, some 3,000 Indigenous people, 500 Africans and 500 Spaniards, as well as many mulattos, are reported for the entire city of Puebla (Nutini & Isaac 2009: 34–36). The situation changed dramatically by the time of the 1777 census, however, when the town's population reached over 56,000 inhabitants, with ca. 31.8% of Spaniards, 21.4% of Indigenous people, 16.1% of mestizos, 4.6% of mulattos, and the rest constituting “other castes” (Nutini & Isaac 2009: 48). In Zacatecas, the center of the silver mining industry, by 1572 the approximately 1,500 Natives and 500 slaves of African descent outnumbered the resident population of 300 Spaniards (Velasco Murillo & Sierra Silva 2012: 109–117). The situation in Native towns, and especially in rural areas, was distinct because in many cases a very limited, mainly individual bilingualism survived until the second half of the twentieth century, and even more recently for those in more secluded, peripheral locations. Nonetheless, while varying between regions, depending on their degree of accessibility, the influx of Spanish-speaking settlers into Indigenous areas grew steadily during the colonial period, eventually causing language pressure, along with a growing pressure on Native land.

### **3 Modern and contemporary Mexico**

Fostering transitional Spanish-Indigenous societal bilingualism was an important goal of the Mexican state right from its creation. It abolished the legal category of *indios* at the expense of “citizens” and, in the associated rhetoric of “progress”, Indigenous tongues were deprived of their importance as essential components of ethnic identity. Instead they were reduced to symbols of backwardness and obstacles to modernisation, as well as to successful integration into society (Heath 1972: 62–64, Estrada Fernández & Grageda Bustamante 2010: 582–583). In terms of diglossia, Spanish, as an official and national language, was the only prestigious and high “variety”, while all Indigenous languages became low varieties (Zimmermann 2010: 911). As a consequence, Spanish gained more and more presence in different social domains of daily life and in communicative interaction between Indigenous and non-Indigenous people, contributing to the growth of societal bilingualism and the shift to the dominant language (Villavicencio Zarza 2010: 723–730, González Luna 2012: 92–93).

While the level of implementation of the state’s linguistic policy varied, significant changes came about in the second half of the twentieth century. In 1948 the National Indigenist Institute (INI) of Mexico was established by the President, Miguel Alemán, with the aim of exploring problems affecting the Indigenous population and seeking ways to improve their living conditions, e.g. by sending special educators (*promotores culturales*) to regional centers. Envisioned as

“agents of change” within local communities, they were Natives from the same region and knew both Spanish and the Indigenous tongue of the area (Heath 1972: 135–138). In the following decades, the Hispanisation of Native children, achieved via the direct method of using Spanish as the only language of school instruction, became the most widely applied educational model, even though “bilingual education” was one of the state’s official goals (Heath 1972: 162–163). In view of its failure, this program gave way in 1981 to another initiative called *educación bilingüe-bicultural* – the most recent myth serving as a disguise for the imposition of Spanish. The aim of this approach is officially to develop literacy in a Native language before teaching Spanish, yet, ultimately, the role of local languages is reduced to a medium of instruction of the target language (Flores Farfán 1999: 40–41).

The most widespread assimilationist education model, based exclusively on Spanish, has remained the most common practice in local communities until the present day. It has often been combined with a more or less official prohibition of the use of Indigenous tongues at school and the consequent stigmatisation of children who do not speak Spanish. As language shift has deepened, attitudes of internal racism have surfaced in Indigenous communities. They have been directed toward those community members, including children, who have been less successful in achieving Hispanisation. The Mexican school system and its teachers have been instrumental in such cases. The widespread shift is reflected in the census data, even if this is treated with extreme caution. These data show a steady and rapid decrease in the numbers of Indigenous monolinguals and a subsequent increase in bilinguals and monolingual Spanish speakers. This is confirmed by ethnographic and linguistic surveys: for example, in the Tlaxcala Pueblan Valley at the end of the nineteenth century, more than 70% of the population was Nahua, living in traditional, monolingual communities almost untouched by secularisation. However, it is estimated that only about 2% of the valley’s population could still be considered “Indigenous” in the year 2000, with rapidly fading Indigenous-mestizo differences (Nutini & Isaac 2009: 194). Transitional societal bilingualism and an accelerating shift to Spanish has come to be the dominant situation for Nahua communities, albeit occurring on different timescales for different communities.

In terms of currently dominant language ideologies and associated power relationships, members of Native communities usually situate Nahuatl (and other local languages) at the very bottom of the language hierarchy. Spanish is in the middle as a national language and that of the dominant “modern” society, and most recently, English has claimed its place at the very top as a symbol of upward social mobility and opportunities, associated with technology, business, youth

and popular culture. For communities with high rates of migration to the US, it is also the language of remote opportunities and a symbol of a better life. Spanish remains linked to all basic dimensions of social life as the unique language of education, politics, work, and legal and public services. In comparison, Nahuatl's typical (and often only) domains include household, family and agriculture. It is regarded as a lower-status tongue of *campesinos* (peasants), who are situated in a much less advantageous societal position than Spanish-speaking professionals (Sandoval Arenas 2017). Decisions to favor the unmarked choice of Spanish are often behind a community-level shift to this national language, in accordance with the strong discourse of *salir adelante*, "forging ahead" and improving one's socioeconomic position (Messing 2007: 569–572).

## **4 Research in four Nahua-speaking regions: contexts, study and participants**

While language ideologies and attitudes shed important light on the nature of contemporary Spanish-Nahuatl bilingualism, important insights also come from quantitative research. The research we report on here is part of a team project that included four Nahuatl-speaking regions of Mexico: the town of Atliaca in the municipality of Tixtla in the state of Guerrero; rural communities in the municipality of Chicontepec (Huasteca Veracruzana, in the state of Veracruz); Xilitla and other municipalities in Huasteca Potosina (the state of San Luis Potosí); and the municipality of Contla de Juan Cuamatzi in the state of Tlaxcala. These regions represent complex cultural traditions dating back to pre-Hispanic Mesoamerica. While they share a general cultural background and history that is typical of the broad Mesoamerican cultural area, the members of these communities speak distinct variants of Nahuatl, which are nevertheless mutually intelligible to a high degree. They also differ in terms of Indigenous language retention and strength of Indigenous identity.

The most traditional are rural communities located in the municipality of Chicontepec, where, according to the 2010 census, 67% of the population spoke Nahuatl, including 51% of children (INEGI 2010). Community members continue many core elements of traditional religion and corn-based agriculture, sharing the strong identity of *macehualmeh* or Indigenous people. In Atliaca, a small town in the municipality of Tixtla in Guerrero, some 80% of inhabitants knew Nahuatl, according to the 2010 census. However, Spanish is becoming increasingly dominant, especially in central sectors of the town and among the younger generations. Inhabitants live on traditional agriculture, brick production and other

specialised professions. In the third locality, Xilitla, some 40% of residents were reported to be speakers of Nahuatl in 2010. Cultural assimilation (*mestizaje*) is quite strong here, with traditional agriculture increasingly being eroded; many inhabitants of the region rely on state support and small-scale wage work. Even more culturally assimilated, and most urbanised, are communities in the municipality of Contla in Tlaxcala, where the shift to Spanish is the most advanced. According to the 2010 census data, only ca. 15.5% inhabitants identified themselves as speakers of Nahuatl; among children under 14, less than 3% were reported to speak the language. While local communities continue some forms of traditional religious organisation and corporate government, the economy is mainly based on wage labor, local industries (such as textile production) and other small businesses. All four regions share a history of discrimination and stigmatisation of Nahuatl-speaking children at school. Almost all speakers of Nahuatl also speak Spanish (with the oldest generations displaying differing levels of proficiency), while the youngest often exhibit reduced or passive skills in Nahuatl.

The survey for this project, carried out in 2018 and 2019, was based on an extensive panel questionnaire in Spanish, which was conducted mainly in person by local Nahuatl-speaking project members and collaborators; in the case of respondents whose preferred language of communication was Nahuatl (in Chicontepec and Atlíaca), the interviews were conducted in this language and questions were translated into Nahuatl. Some of the younger participants, mainly in the region of Huasteca Potosina and in Tlaxcala, completed the questionnaire online. In total, the survey reached 552 respondents, whose mean age ( $M_{age}$ ) was 37.9 ( $SD = 18.3$ ). 55.4% of the sample were women ( $n = 306$ ). Samples in the four regions varied from 108 to 156: Atlíaca ( $n = 152$ ;  $M_{age} = 33.3$ ,  $SD = 18.26$ ; 75 women), Chicontepec ( $n = 108$ ;  $M_{age} = 59.17$ ,  $SD = 17.91$ ; 62 women), Xilitla and neighboring municipalities<sup>1</sup> ( $n = 136$ ;  $M_{age} = 30.79$ ,  $SD = 17.77$ ; 65 women), and Contla ( $n = 156$ ;  $M_{age} = 40.25$ ,  $SD = 14.62$ ; 104 women). Language use was assessed with a set of 14 items relating to narrow subdomains of everyday communicative situations, including different functional and social network-related domains. This scale aimed to reflect a broad range of domains of language use, taking into account the use of both the minority and dominant languages within family circles, immediate social networks, with friends, in schools, institutions, services, public events, and on social media (Table 3). Previous scales of this kind include those by Landry & Allard (1994), Ehala & Zabrodskaja (2014), and in the EuLaViBar Project (Åkermark et al. 2013). In contrast with the previous tools, we addressed frequent patterns of interrupted intergenerational language transmission, in which lan-

---

<sup>1</sup>Matlapa, Axtla de Terrazas, Tampacán, Tamazunchale and Coxcatlán.

guage teaching skips the parents' generation, with the oldest family members transmitting the language to the youngest generations.

## 5 Results

Preliminary qualitative analysis of the frequency of Nahuatl and Spanish language use across ethnic groups was assessed in the 14 subdomains of everyday communication with: parents, grandparents, children, friends, neighbors, doctors, attendees of cultural activities, people on social media, municipal authorities, community authorities and healers, as well as during participation in family meetings, ceremonies, and church services. A Likert scale of 1-7 was used to rate language use across different domains, where steps 1-3 indicated prevalent use of the Spanish language (over the Nahuatl language), step 4 represented an equal use of Spanish and Nahuatl, and steps 5-7 indicated prevalent use of the Nahuatl language (over Spanish). These frequencies are reported in Table 1 and Figures Figure 1 and Figure 2. The results of the survey confirm and further reveal significant differences between the four regions. The highest retention of Nahuatl was found in Chicontepec in Veracruz, followed by Atliaca in Guerrero and Xilitla region in San Luis Potosi; the lowest use of Nahuatl and the most widespread expansion of Spanish to all domains of life were found in the region of Contla in Tlaxcala. These results confirm preliminary observations and conclusions drawn from qualitative data acquired in fieldwork, but at the same time they show measurable differences across regions and domains, revealing aspects of life where Spanish has almost completely taken over spaces previously reserved for Indigenous languages. The outcomes of the quantitative survey also illustrate the strong dominance of Spanish in new spheres of usage, such as the Internet, social media and health services. Nahuatl's strongest bastion is the family domain and, in particular, communication with grandparents and parents. However, that drops abruptly, even in Chicontepec and Atliaca, in the case of communication with children. This pattern bespeaks widespread ruptures in the intergenerational transmission of the heritage language, and an ongoing and rapid shift to Spanish. This accelerated process can be described as a generational turn from transitional Spanish-Nahuatl bilingualism to monolingualism in the national language. In Chicontepec and Atliaca, communication in Nahuatl outside of the family domain is the strongest with healers, and remains strong with neighbors, friends, community authorities, and during traditional ceremonies, with the averages showing an equal use of Spanish and Nahuatl, or a slightly more prevalent usage of Nahuatl. Diagram 2 illustrates the expansion

of Spanish into different domains of life. In the case of the Contla region it is almost exclusively the only language used in every context of life, except for communication with grandparents and parents, where there is still some retention of Nahuatl. An ANOVA test was run to compare mean level differences in Nahuatl language use across the four groups. The results, presented in Table 1, revealed statistically significant differences between communities in mean levels of all variables regarding the relative use of Nahuatl and Spanish in various domains of life. The highest use of Nahuatl was in Chicontepec with grandparents (6.14), parents (5.74) and healers (5.72), followed by the communication with grandparents in Atliaca (5.52). The lowest values across all 14 domains are found invariably in Contla.

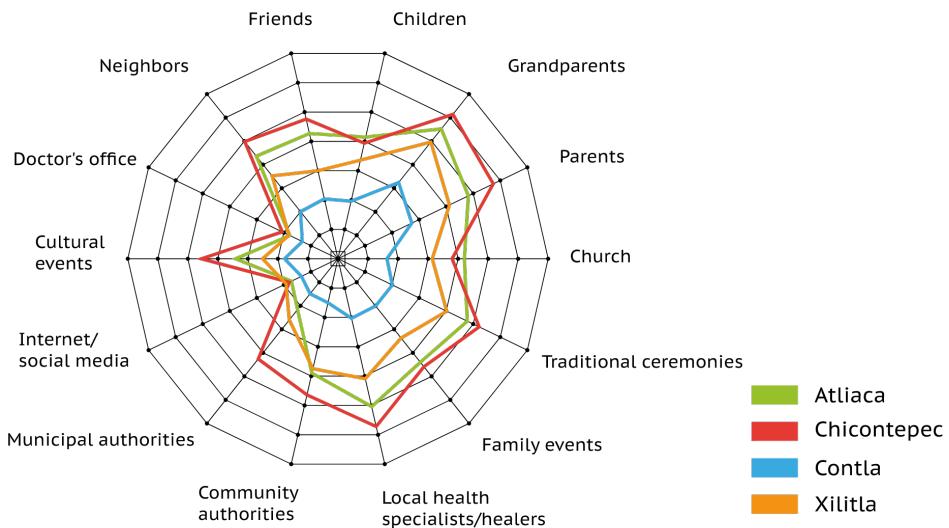


Figure 1: The use of Nahuatl across different domains in the four regions

During the survey, participants were also asked to self-assess their oral and writing skills in Nahuatl and Spanish, indicating how well they speak and write according to the following 6-item Likert scale: 1 not at all (I can't speak it or I can't write it), 2 hardly any, 3 a little bit, 4 moderately (neither good nor bad), 5 well, 6 very well. They were also asked to assess the degree of difficulty or ease with which they speak both languages, using a 6-item Likert scale where 1 indicated a lack of knowledge of the language in question, 2 very difficult, 3 difficult, 4 moderate (neither difficult nor easy), 5 easy, 6 very easy. The results are presented in Table 2. In general, it is clear that in all the regions except Chicontepec,

3 *Spanish-Nahuatl bilingualism in Indigenous communities in Mexico*

Table 1: Averages regarding the use of Nahuatl and Spanish in different domains of life and social networks. \*\*\*  $p < 0.001$

	Merged sample (n=235)	SD	Atliaca (n=63)	SD	Chicon-tepec (n=30)	SD	Contla (n=99)	SD	Xilitla (n=43)	SD	F (df)
Parents	4.23	2.328	4.82	2.044	5.74	1.787	2.73	2.019	4.12	2.33	48.531*** (519)
Grandparents	4.85	2.296	5.52	1.769	6.14	1.634	3.24	2.231	4.96	2.333	46.48*** (491)
Children	3.17	1.92	4.15	1.691	3.94	1.91	1.96	1.344	3.36	1.938	38.893*** (355)
Friends	3.42	2.013	4.27	1.94	4.76	1.776	2.04	1.455	3	1.65	68.16*** (523)
Neighbors	3.6	2.121	4.36	2.057	4.98	1.75	2	1.456	3.52	1.889	68.399*** (521)
Doctor's office	1.71	1.299	1.78	1.368	2.06	1.866	1.32	0.713	1.81	1.116	7.663*** (520)
Cultural events	2.93	2.004	3.41	2.147	4.55	1.899	1.78	1.287	2.51	1.571	55.569*** (518)
Internet/ social media	1.66	1.031	1.72	1.066	1.8	1.181	1.35	0.77	1.92	1.125	7.279*** (427)
Municipal authorities	2.46	1.826	2.16	1.466	4.26	2.114	1.49	0.876	2.6	1.803	64.486*** (514)
Community authorities	3.29	2.245	3.9	2.438	4.64	1.916	1.5	0.947	3.75	2.038	66.722*** (510)
Local healers	4.09	2.39	5.04	2.157	5.72	1.631	2.03	1.65	4.09	2.186	89.661*** (509)
Family events	3.5	2.018	4.4	1.839	4.59	1.75	2.01	1.402	3.37	1.916	64.853*** (522)
Traditional ceremonies	3.88	2.22	4.78	2.052	5.22	1.784	2.01	1.426	4.02	2.012	82.958*** (517)
Church	3.13	2.084	4.28	2.099	3.8	2.138	1.63	1.089	3.14	1.855	57.471*** (499)

Table 2: Frequencies (n and percentage in brackets) of individuals using Spanish, Náhuatl or both languages equally in 14 everyday domains in four ethnic groups. \*  $p < 0,05$ ; \*\*  $p < 0,010$ ; \*\*\*  $p < 0,001$ .

Language type across ethnic groups															
Atliaca (n=152)				Chiapas (n = 108)				Contla (n = 156)				Xilitla (n = 136)			
Subdomains	S	Equal N&S	N	S	Equal N&S	N	S	Equal N&S	N	S	Equal N&S	N	S	χ²	
Parents	39 (26)	28 (18)	80 (53)	11 (10)	21 (20)	69 (64)	105 (67)	17 (11)	28 (18)	48 (35)	23 (17)	51 (38)	116.071***		
Grandparents	20 (13)	23 (15)	104 (68)	8 (7)	72 (67)	81 (52)	17 (11)	41 (26)	33 (24)	12 (9)	73 (54)	115.862***			
Children	24 (16)	30 (20)	32 (21)	29 (27)	33 (31)	22 (20)	114 (73)	12 (8)	5 (3)	31 (23)	12 (9)	12 (9)	185.922***		
Friends	46 (30)	41 (27)	62 (41)	19 (18)	37 (34)	44 (41)	122 (78)	17 (11)	13 (8)	73 (54)	37 (27)	13 (10)	149.319***		
Neighbors	48 (32)	31 (20)	71 (47)	14 (13)	32 (30)	53 (49)	124 (80)	18 (12)	10 (6)	53 (39)	35 (26)	33 (24)	161.292***		
Doctor's office	138 (91)	2 (1)	8 (5)	83 (77)	2 (2)	13 (12)	152 (97)	0 (0)	1 (1)	115 (85)	5 (4)	2 (2)	46.105***		
Cultural events	79 (52)	29 (19)	40 (26)	24 (22)	39 (36)	34 (32)	135 (87)	9 (6)	9 (6)	91 (67)	22 (16)	8 (6)	140.051***		
Internet/ social media	121 (80)	10 (7)	3 (2)	35 (32)	3 (3)	2 (2)	135 (87)	3 (2)	1 (1)	104 (76)	10 (7)	1 (1)	136.874***		
Municipal authorities	123 (81)	12 (8)	10 (7)	35 (32)	22 (20)	39 (36)	148 (95)	3 (2)	1 (1)	91 (67)	14 (10)	17 (13)	143.783***		
Community authorities	71 (47)	13 (9)	65 (43)	21 (19)	27 (25)	42 (39)	141 (90)	7 (5)	2 (1)	54 (40)	30 (22)	38 (28)	181.422***		
Local healers	34 (22)	19 (13)	95 (63)	8 (7)	18 (17)	71 (66)	121 (78)	9 (6)	16 (10)	41 (30)	27 (20)	51 (38)	195.588***		
Family events	34 (22)	58 (38)	58 (38)	15 (14)	50 (46)	34 (32)	124 (80)	22 (14)	6 (4)	58 (43)	38 (28)	26 (19)	173.132***		
Traditional ceremonies	34 (22)	36 (24)	77 (51)	13 (12)	30 (28)	57 (53)	122 (78)	19 (12)	10 (6)	41 (30)	35 (26)	44 (32)	177.013***		
Church	46 (30)	35 (23)	57 (38)	36 (33)	31 (29)	27 (25)	138 (89)	9 (6)	3 (2)	63 (46)	35 (26)	20 (15)	147.756***		

### 3 Spanish-Nahuatl bilingualism in Indigenous communities in Mexico

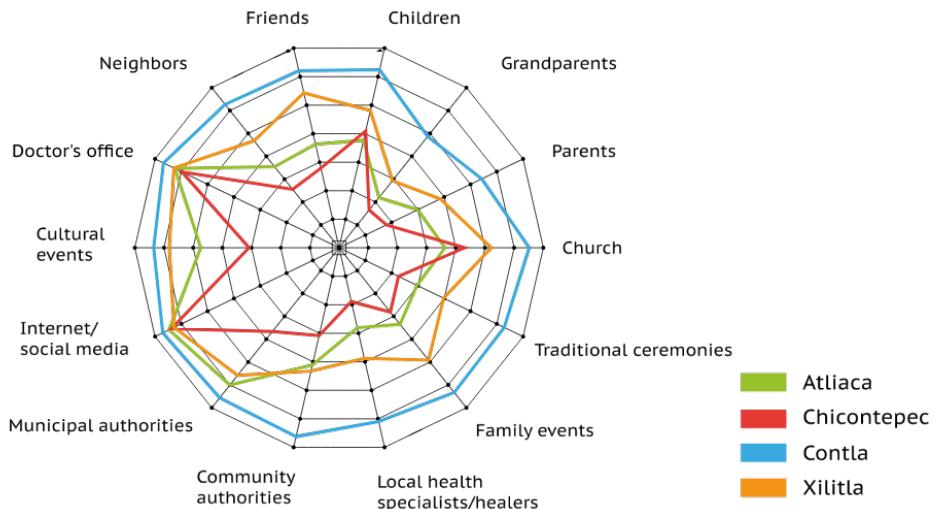


Figure 2: The use of Spanish across different domains in the four regions

respondents declared a higher spoken proficiency in Spanish than in Nahuatl. In Chicontepec the average oral skills in Nahuatl are slightly higher than in Spanish (4.80 to 4.63); in Atliaca and Xilitla the average proficiency in Spanish is only slightly higher than self-assessed proficiency in Nahuatl (4.65 to 4.22 and 5.24 to 4.22 respectively). The difference is most striking in Tlaxcala (4.99 to 2.78).

The same pattern is seen in responses to the question regarding the difficulty of expression in both languages, with only respondents from Chicontepec self-declaring more difficulty speaking Spanish than Nahuatl, although the difference is relatively small (5.04 to 4.69). With regard to writing skills, participants in all regions declared a much higher writing proficiency in Spanish (4.80 to 2.44 in the overall sample). The highest Nahuatl writing skills were recorded in the region of Xilitla, which is explained by the participation of students at a local university where some courses are given in Nahuatl (this also accounts for the highest average score for Spanish literacy in this sample). In Chicontepec and Atliaca, despite a generally high oral proficiency in Nahuatl, written competence was low, which attests to the role of Nahuatl as a predominantly oral language, absent from written spaces and school education. This very limited presence of the Indigenous language in written media and the limited literary culture among its speakers makes it more difficult to expand its use to less traditional domains of life associated with technology, education or administration. As a matter of fact, this is a significant difference with regard to the colonial period, when written Nahuatl

was widely present and used in administrative, legal, religious, economic, educational and even private or personal spheres of life. An ANOVA test was run to compare mean level differences in Nahuatl and Spanish self-assessed proficiency across the four groups. The results, presented in Table 3, revealed statistically significant differences between communities in mean levels of all variables regarding language proficiency. Thus, summing up, the highest self-assessed oral skills in Nahuatl were found among respondents from Chicontepec, followed by Atlíaca and Xilitla with the same average value. The same pattern is confirmed in the self-assessment of the feeling of ease while speaking Nahuatl. The highest self-reported writing skills in Nahuatl were found among Xilitla respondents, whereas Spanish skills were ranked highest in Xilitla and lowest in Chicontepec.

The results discussed above are fully congruent with Pearson's correlations (all assumptions of Pearson's correlations hold) between analyzed variables based on the overall sample from the four regions. Table 4 presents statistically significant correlations between Nahuatl use in the family domain (including with family members, neighbors and during family gatherings), the use of Nahuatl across different domains, and self-assessed oral skills in Nahuatl and Spanish. It is not surprising that a high proficiency in spoken Nahuatl is strongly and positively correlated with its use in the household and immediate neighborhood, as well as with its usage in different aspects of life. However, proficiency in Spanish is negatively correlated with the usage of Nahuatl. Moreover, it is negatively correlated with oral proficiency in Indigenous languages, which confirms that Spanish-Nahuatl bilingualism is highly unstable, competitive and transitional toward the national language.

These data are additionally explained by outcomes of a complementary survey of proficiencies in Spanish and Nahuatl using visual elicitation tools. 74 participants from the four regions (mean age: 46.02; 37 men, 37 women) were interviewed and recorded using purely visual elicitation tools: a series of pictures embracing both traditional and non-traditional objects (including some unusual ones, e.g. hybrid animals, included in order to assess language skills and lexical creativity, such as the ad-hoc creation of neologisms), and two movies, one showing some traditional daily activities of an Indigenous family and another presenting a short story featuring agricultural work and children's activities. The participants were asked to name each object and describe the movies as they watched them, using either Nahuatl and Spanish in a randomised order (changing the language after each whole sequence of elicitation).

Table 5 shows the vocabulary density and the frequency of the usage of loanwords in both languages for the recorded elicitation sample. The vocabulary density, or the ratio of the number of unique words (types) to the number of all words

Table 3: Averages regarding language competences. \*\*\*  $p < 0.001$ .

	Merged sample (n=516)		Atliaca (n=150)		Chicontepec (n=100)		Contla (n=132)		Xilitla (n=114)		F (df)
	M	SD	M	SD	M	SD	M	SD	M	SD	
I speak Nahuatl.	3.91	1.478	4.22	.874	4.80	1.123	2.78	1.529	4.22	1.423	63.083*** (526)
I speak Spanish.	4.88	.990	4.65	.803	4.63	1.084	4.99	1.003	5.24	.983	11.327*** (526)
I write Nahuatl.	2.44	1.486	2.21	1.166	2.13	1.376	1.92	1.277	3.68	1.512	46.583*** (525)
I write Spanish.	4.80	1.178	4.53	1.079	4.50	1.324	4.88	1.170	5.28	1.006	12.550*** (522)
When I speak Nahuatl, it is...	4.37	1.317	4.64	.861	5.04	1.044	3.46	1.465	4.65	1.187	46.620*** (522)
When I speak Spanish, it is...	5.04	.958	4.94	.771	4.69	1.134	5.11	.987	5.37	.855	10.587*** (516)

Table 4: Pearson's correlations between Nahuatl use in family, Nahuatl use across different domains and self-assessed proficiencies in Nahuatl and Spanish; \*\* $p < .01$  (n=525)

	Nahuatl general use	Nahuatl oral proficiency	Spanish oral proficiency
Nahuatl use in family	.920**	.762**	-.388**
Nahuatl general use		.738**	-.422**
Nahuatl oral proficiency			-.190**

(tokens) in the utterance of a specific person, serves as a proxy for the complexity of the utterance. Where an utterance includes, on average, many tokens of the same type, the density is lower, whereas a higher ratio indicates a richer repertoire of word types used. In all the regions studied, Spanish elicitations had a lower density than Nahuatl ones; however, the smallest difference between the two languages was observed in the two regions with the most advanced shift to Spanish: Contla and Xilitla. The differences between the mean Nahuatl and Spanish results obtained via an ANOVA test are presented in Table 6. They were found to be statistically significant for all three measures discussed (i.e. ratio of borrowed words to all words, ratio of borrowed types to all types and vocabulary density) in all four communities, with the exception of the difference in vocabulary density in Atlíaca. The lack of statistical significance might be explained by a smaller number of elicitations in that community.

While it is hard to draw any far-reaching conclusions from such a comparison between utterances in two languages with profound differences in morphosyntax, it is clear that Nahuatl spoken in communities with the highest vitality of this language – Chicontepec and Atlíaca – reveals a higher density, i.e. richer vocabulary in utterances, than among speakers living in the more assimilated and linguistically endangered regions, Xilitla and Contla. This is also fully consistent with the data previously discussed, in that the shift to Spanish is ongoing and widespread especially in these two latter regions. In addition, the shrinking proficiency and reduced semantic functions of the Indigenous language are notable.

The analysis of the frequency of loanwords in Spanish and Nahuatl elicitations is also quite revealing. In the overall sample, the percentage of Spanish loanwords into Nahuatl is relatively high: from 13% in Xilitla to 25% in Atlíaca. The rate of borrowing in Contla is not much different – 23% – even though the Nahuatl used in this region was characterised in the past (Hill & Hill 1986) as a “syncretic language”, drawing heavily on the Spanish lexicon. This categorisa-

### 3 Spanish-Nahuatl bilingualism in Indigenous communities in Mexico

Table 5: Quantitative results of the visual elicitation assessment of proficiency in Nahuatl and Spanish

	Vocabulary density		Borrowed words : all words ratio		Borrowed types : all types ratio	
	Nahuatl	Spanish	Nahuatl	Spanish	Nahuatl	Spanish
Atliaca	0.55	0.45	0.22	0.01	0.25	0.02
Chicontepec	0.47	0.33	0.16	0.01	0.20	0.02
Contla	0.37	0.31	0.18	0.01	0.23	0.02
Xilitla	0.39	0.33	0.11	0.01	0.13	0.01
Merged sample	0.45	0.36	0.17	0.01	0.20	0.02

Table 6: *F*-tests and degrees of freedom of the differences between the quantitative results of proficiency assessment in Spanish and in Nahuatl: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

	Atliaca	Chicontepec	Contla	Xilitla
Borrowed words : all words ratio	6.855* (12)	72.037*** (45)	94.312*** (59)	32.259*** (39)
Borrowed types : all types ratio	7.781* (12)	79.173*** (45)	96.209*** (59)	38.627*** (39)
Vocabulary density	1.328 (12)	39.638*** (45)	12.831** (59)	7.255* (39)

tion is not confirmed in the documented elicitations of older proficient speakers of Nahuatl from the region, whose borrowing rate is lower than in the less assimilated Atliaca region, and not much higher than in Chicontepec, where language transmission still occurs, and where Nahuatl-Spanish bilingualism is more widespread. The average rate of usage of Spanish loanwords in the four regions is 17%, and this rises to 20% when overall word types and borrowed word types are compared.

What is even more striking, however, is an almost complete absence of Nahuatl loanwords in the Spanish utterances of the participants of our survey: regardless of the region, the rate is always  $\leq 1\%$  and  $\leq 2\%$  for all tokens and types respectively. Moreover, the few loanwords from Nahuatl which are attested in the Spanish utterances are essentially limited to those commonly used in Mexican Spanish, such as *aguacate* ‘avocado’, *chiquihuite* ‘basket’, *comal* ‘type of griddle traditional

in Mesoamerica' and *zacate* 'forage'. The lack or avoidance of Nahuatl loanwords in the speech of persons for whom, in the majority of cases, Nahuatl was the first language<sup>2</sup> suggests that Spanish was probably learned largely at school or outside the community as a more "standardised" language devoid of easily perceptible (i.e. lexical) Indigenous impact. Perhaps Nahuatl loanwords were avoided because of their association with a stigmatised identity. This finding is even more striking when compared to the Spanish language used by non-Indigenous members of the colonial society of New Spain, including Spaniards and creoles (as attested in numerous genres of colonial written documents), where Nahuatl loanwords were quite common, especially for local objects, plants, animals and even concepts that went on to become part of the general culture and lexicon. This strong asymmetry in the results of language contact between Spanish and Nahuatl is yet further salient evidence of very unstable and transitional bilingualism.

## 6 Discussion and conclusions

Widely shared and popularised views about generalised Spanish-Indigenous societal bilingualism and *mestizaje* that developed during the colonial era find little support in the available data, nor in the most recent linguistic trajectories of Native communities. This kind of general bilingualism was not very common among the Indigenous population, even if it was increasing in large urbanised zones, particularly during the seventeenth and eighteenth centuries. Given the widespread presence of Nahuatl in New Spain, its official recognition and strong economic and sociopolitical potential, it was also quite common for non-Indigenous members of the colonial society to learn this local language for practical purposes. Undeniably, some communities, due to a number of factors, underwent assimilation and experienced a more or less complete shift to Spanish by the latter part of the colonial period. In the nineteenth and twentieth centuries, the marginalisation and discrimination of Indigenous communities that were largely monolingual in Nahuatl and/or bi/multilingual in local Indigenous languages deepened, and many chose the path of quick assimilation toward a *mestizo* status and the use of Spanish. While most of the Nahua communities were exposed to differing degrees to the Spanish language and culture from the first phase of colonisation, this did not constitute a threat to the heritage language used by Indigenous groups, which at this time was still characterised by high

---

<sup>2</sup>59% declared Nahuatl to be their first language, 27% Nahuatl and Spanish, 9% Spanish, and 5% did not specify.

ethnolinguistic vitality. The pressure of Spanish became much stronger after Independence, altering the nature of cultural and linguistic contact, which became more aggressive and displacive (Olko 2018). Although Mexican bilingualism has been seen as “as a long-term historical process” (Flores Farfán 2003: 332), it was, in fact, limited and ephemeral in local communities during the colonial period. In large urban contexts, the scale of societal bilingualism increased over time, gradually becoming transitional, and eventually unstable and transitory during later (post-colonial) times, triggering accelerated assimilation processes toward the national culture.

In contemporary “bilingual” Nahua communities this dynamic process is characterised by differing proficiencies in the two languages. Until recently, many speakers of Nahuatl, especially elderly ones, had limited proficiency in Spanish (this is still attested in regions such as Chicontepec). Now, however, it is more common to see highly varying proficiencies in the heritage language, with many non-fluent and/or non-active speakers among the younger generations (see Dorian 1981, 1986, Grinevald 1998). Even in communities where Nahuatl is still spoken by the majority of people alongside Spanish, it is not uncommon to find families in which the grandparent and parent generations are fully proficient in Nahuatl, where the oldest, usually adolescent, children can speak with differing levels of competence, while their younger siblings are passive speakers and the very youngest are non-speakers. Such patterns strongly influence the dynamics and patterns of bilingual communication within specific households and across the whole community. Thus, among Nahua communities today we find a broad continuum of proficiency in the ancestral language, strictly related to the mode and circumstances of its transmission and the degree of socialisation in it (Olko 2018, Flores Farfán & Olko 2021). The results of the quantitative large-scale survey in four different regions where Nahuatl is still spoken, complemented by the assessment of proficiencies in Nahuatl and Spanish, allow us to draw a data-driven and coherent picture of current bilingual arrangements. The sociolinguistic situation can be described as unstable, asymmetrical Spanish-Nahuatl bilingualism leading to shift to the national language. Depending on the region, this may occur as quickly as within two to four generations due to strong power differentials between the two languages, as well as related economic, sociopolitical, cultural and educational pressures.

## Acknowledgements

We have no conflicts of interests to disclose. The Project “Language as a cure: Linguistic vitality as a tool for psychological well-being, health and economic sus-

tainability” is carried out within the Team program of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund. We are very grateful to the members of the Nahua communities for participating in the survey and in other parts of our research, to the anonymous reviewers for their comments regarding the article, and to Ellen Foote and Mary Chambers for the stylistic revision of the text.

## Author contributions

Justyna Olko played a lead role in Conceptualisation, Funding acquisition, Investigation, Methodology, Supervision, Field research, Writing – original draft, Writing – review & editing and she played a supporting role in Data curation and Project administration. Szymon Gruda played a supporting role in Writing – review & editing. Joanna Maryniak played a lead role in Data curation and a supporting role in Field research, Writing – review & editing. Justyna Olko, Szymon Gruda and Joanna Maryniak played equal roles in the data analysis. Elwira Dexter-Sobkowiak, Humberto Iglesias Tepec, Eduardo de la Cruz, and Beatriz Cuahutle Bautista played a lead role in Field research and supporting role in Conceptualisation.

## References

- Åkermark, Sia Spiliopoulou, Johanna Laakso, Anneli Sarhima & Reetta Toivanen. 2013. *EuLaViBar toolkit: Practical guide to the EuLaViBar tool, with reference to the ELDIA comparative report*. [https://fedora.phaidra.univie.ac.at/fedora/objects/o:301101/methods/bdef:Container/get?param0=ELDIA\\_EuLaViBar-Toolkit.pdf](https://fedora.phaidra.univie.ac.at/fedora/objects/o:301101/methods/bdef:Container/get?param0=ELDIA_EuLaViBar-Toolkit.pdf).
- Dorian, Nancy. 1981. *Language death: The life cycle of a Scottish Gaelic dialect*. Philadelphia: University of Pennsylvania Press.
- Dorian, Nancy. 1986. Abrupt transmission failure in obsolescing languages: How sudden they “tip” to the dominant language communities and families. In Vasiliki Nikiforidu, Mary Van Clay, Mary Niepokuj & Deborah Feder (eds.), *Proceedings of the twelfth annual meeting of the Berkeley Linguistics Society*, 72–83. Berkeley: Berkeley Linguistics Society.
- Ehala, Martin & Anastassia Zabrodskaja. 2014. Hot and cold ethnicities in the Baltic states. *Journal of Multilingual and Multicultural Development* 35(1). 76–95.

### *3 Spanish-Nahuatl bilingualism in Indigenous communities in Mexico*

- Estrada Fernández, Zarina & Arón Grageda Bustamante. 2010. Colonización y política del lenguaje: El norte de México. In Rebeca Barriga Villanueva & Pedro Martín Butragueño (eds.), *Historia sociolingüística de México*, vol. 1, 545–604. México: Colegio de México.
- Flores Farfán, José Antonio. 1999. *Cuatreros somos y toindioma hablamos: Contactos y conflictos entre el náhuatl y el español en el sur de México*. México: Centro de Investigaciones y Estudios Superiores en Antropología Social.
- Flores Farfán, José Antonio. 2003. Nahuatl purism: Between language innovation, maintenance and shift. In Joseph Brincat, Winfried Boeder & Thomas Stolz (eds.), *Purism in the age of globalization*, 281–313. Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Flores Farfán, José Antonio & Justyna Olko. 2021. Types of communities and speakers. In Justyna Olko & Julia Sallabank (eds.), *Revitalizing endangered languages: A practical guide*, 75–87. Cambridge: Cambridge University Press.
- Gibson, Charles. 1964. *The Aztecs under Spanish rule. A history of the Indians of the Valley of México 1519–1810*. Stanford: Stanford University Press.
- González Luna, Ana María. 2012. La política lingüística en México entre independencia y revolución (1810–1910). In Patrizia Botta & Sara Pastor (eds.), *Actas del Congreso de la Asociación Internacional de Hispanistas*, vol. VIII, Lengua, 91–101. Rome: Bagatto Libri.
- Grinevald, Colette. 1998. Language contact and language degeneration. In Florian Coulmas (ed.), *The handbook of sociolinguistics (Blackwell reference online)*, 176–184. Oxford: Blackwell Publishing.
- Heath, Shirley Brice. 1972. *Telling tongues: Language policy in México, colony to nation*. New York: Teachers College Press.
- Hill, Jane H. & Kenneth C. Hill. 1986. *Speaking mexicano: Dynamics of syncretic language in Central México*. Tucson: University of Arizona Press.
- Instituto Nacional de Estadística, Geografía e Informática. 2010. *Censo de población y vivienda 2010. Características culturales de la población, población hablante de lengua indígena de 5 y más años por principales lenguas*.
- Landry, Rodrigue & Real Allard. 1994. Diglossia, ethnolinguistic vitality, and language-behavior. *International Journal of the Sociology of Language* 108. 15–42.
- Lockhart, James. 1992. *The Nahuas after the conquest: A social and cultural history of the Indians of central México, sixteenth through eighteenth centuries*. Stanford: Stanford University Press.
- Messing, Jacqueline. 2007. Multiple ideologies and competing discourses: Language shift in Tlaxcala, México. *Language in Society* 36(4). 555–577.

- Nesvig, Martin. 2012. Spanish men, Indigenous language, and informal interpreters in postcontact México. *Ethnohistory* 59(4). 739–776.
- Nutini, Hugo H. & Barry L. Isaac. 2009. *Social stratification in Central México*. Austin: University of Texas Press.
- Olko, Justyna. 2018. Unbalanced language contact and the struggle for survival: Bridging diachronic and synchronic perspectives on Nahuatl. *European Review* 26(1). 207–228.
- Parodi, Claudia. 2010. Tensión lingüística en la colonia: Diglosia y bilingüismo. In Rebeca Barriga Villanueva & Pedro Martín Butragueño (eds.), *Historia sociolingüística de México*, vol. 2, 287–345. México: Colegio de México.
- Sandoval Arenas, Carlos O. 2017. Displacement and revitalization of the Nahuatl language in the high mountains of Veracruz, México. *Arts & Humanities in Higher Education* 16(1). 66–81.
- Schwaller, John. 2012. The expansion of Nahuatl as a lingua franca among priests in sixteenth-century México. *Ethnohistory* 59(4). 675–676.
- Velasco Murillo, Dana & Pablo Miguel Sierra Silva. 2012. Mine workers and weavers: Afro-Indigenous labor arrangements and interactions in Puebla and Zacatecas, 1600–1700. In Dana Velasco Murillo, Mark Lentz & Margarita R. Ochoa (eds.), *City Indians in Spain's American empire. Urban Indigenous society in colonial Mesoamerica and Andean South America, 1530–1810*, 104–127. Brighton & Eastbourne: Sussex Academic Press.
- Villavicencio Zarza, Frida. 2010. Entre una realidad plurilingüe y un anhelo de nación: Apuntes para un estudio sociolingüístico del siglo XIX. In Rebeca Barriga Villanueva & Pedro Martín Butragueño (eds.), *Historia sociolingüística de México*, vol. 2, 713–793. México: Colegio de México.
- Yannanakis, Yanna. 2012. Introduction: How did they talk to one another? Language use and communication in multilingual New Spain. *Ethnohistory* 59(4). 667–674.
- Zimmermann, Klaus. 2010. Diglosia y otros usos diferenciados de lenguas y variedades en el México del siglo XX: Entre el desplazamiento y la revitalización de las lenguas indomexicanas. In Rebeca Barriga Villanueva & Pedro Martín Butragueño (eds.), *Historia sociolingüística de México*, vol. 2, 881–955. México: Colegio de México.

# Chapter 4

## Trilingual modality: Towards an analysis of mood and modality in Aymara, Quechua and Castellano Andino as a joint systematic concept

Philipp Dankel<sup>a</sup>, Mario Soto Rodríguez<sup>b</sup>, Matt Coler<sup>c</sup> & Edwin Banegas-Flores<sup>d</sup>

<sup>a</sup>University of Basel <sup>b</sup>University of Freiburg <sup>c</sup>University of Groningen – Campus Fryslân <sup>d</sup>Peruvian Ministry of Education

This contribution examines how indigenous minority languages impact majority European ones, by considering the case of Quechua and Aymara, on the one side, and Castellano Andino (CA) on the other. We extend particular focus to how Aymara and Quechua have impacted CA's grammatical(ized) modality. We show that regional varieties of CA reflect Aymara and Quechua mood, even in the speech those who do not speak either indigenous language by illustrating the emerging strategies used to express the modal values in Aymara and Quechua grammar on different structural levels. We also elaborate on how contact induced change arises from multiple impulses.

### 1 Introduction

In this contribution we examine how long-term language contact with the Amerindian languages Aymara and Quechua have impacted mood and modality in the diaspora language Andean Spanish ("Castellano Andino", hereafter: CA). Historically, all Latin American Spanish varieties developed in diasporic communities. In their recently published volume, Drinka & Chappell (2021) brought up the



socio-historical relevance of the fact that all native Spanish speakers outside of the Iberian Peninsula "...are the linguistic heirs of the Spanish language diaspora that began in 1492 - with the expulsion of Spanish-speaking (Sephardic) Jews, and with the arrival of the Spanish language in the Americas" (Lipski 2010: 550). Currently, Spanish is the majority language in Latin America. However, as we will illustrate, varieties like CA can be considered diaspora languages from another perspective: CA retains the cultural identity of Andean speakers, although standard Spanish, still heavily influenced by the peninsular norm, is clearly the politically dominant code.

We show that CA reflects Aymara and Quechua mood and modality, even in the speech of those who do not speak either indigenous language. Examining mood and modality in CA is interesting because modal marking is an integral part of Andean discourse and is part of the grammar of both indigenous languages. We show how minoritized languages impact majority languages and illustrate how contact-induced change arises from multiple impulses, with particular focus on:

1. speakers' need to express semantic notions important for cultural and communicative routines,
2. the structural and functional potentials of established grammatical constructions in CA, and
3. decreased standardization.<sup>1</sup>

Our use of the terms *mood* and *modality* follows that of Palmer (2001: 1), who defines modality as "the status of the proposition that describes the event". Our analysis encompasses both of Palmer's (2001) distinctive forms that signal modality: mood, on the one hand, and, on the other, what may be referred to as the *modal systems* of the languages. This seems appropriate, considering that modality is rarely coded in a neat, symmetrical system and that "in fact, it may be impossible to come up with a succinct characterization" (Bybee et al. 1994: 176) of its notional domain. This account is crucial especially when performing an analysis of a dynamic contact situation where incipient grammaticalization processes are taken into account.

It is useful to provide a sketch of the mood and modality phenomena in the languages under investigation. We ask the readers to keep in mind that the descriptions below are, by necessity, brief and so are not meant to provide finer-grained

---

<sup>1</sup>This topic falls outside the scope of this chapter. We briefly address it in the conclusion.

analyses of phenomena in each language. To begin, let us consider Peninsular Spanish (hereafter: PS), broadly speaking, so that we may better appreciate how CA has come to differ as a result of sustained contact with Aymara and Quechua.

PS is typically understood as having four moods: indicative (unmarked form used for making assertions and statements), subjunctive (inflected for tense, occurs mostly in subordinate clauses to express nonfactual states like evaluations, possibilities, necessities, emotional states, and intentions), conditional<sup>2</sup> (used to express epistemic modality) and the imperative (Bosque 2012). Additionally, some verbal tenses have modal uses (e.g. the future tense can be used to express conjecture as in *Serán las tres de la tarde* ‘It is probably three in the afternoon’) and there is a set of modal periphrasis that are systematically used to express modal distinctions (e.g. *tener que* ‘holding to’, *deber (de)* ‘owing to’, *poder* ‘can’, *haber de* ‘having to’ all of which are inflected for person and followed by the infinitive (Real Academia Española 2009: ch. 25, ch. 28.6).

The Aymara suffixes which are classified as part of mood and modality fall into three categories (Coler 2014):

1. the past and present counterfactual paradigm (members of which express event modality, i.e. deontic and dynamic modality),
2. the evidential suffixes which express evidential modality, and
3. the imperative paradigm.

We do not include those suffixes whose primary grammatical function is not modal. For example, the nominalizer-*ña* can express deontic modality, but since it is not part of the inflectional paradigms for mood and modality, it is not discussed here.

The mood system for Quechua (specifically Southern Quechua, the variety investigated here) is similar to that of Aymara. The three categories of suffixes which are classified as part of mood and modality include: 1) The tense-aspect system with includes suffixes that express experienced and non-experienced events, 2) the evidential suffixes which express evidential-epistemic modality, and 3) the potential suffix which expresses potential and counterfactual events.

This text is structured as follows: In the following sections we outline previous research (Section 2), our theoretical framework (Section 3), and describe the corpora we used (Section 4). Then, we provide analyses of tense/evidentiality

---

<sup>2</sup>In contemporary PS grammars, conditional verb constructions are part of the past subjunctive paradigm. However, for comparative purposes, and as we focus on mood and modality, we consider them as conditional mood.

(Section 5.1), hearsay and quotatives (Section 5.2), inferential/conjecturals (Section 5.3) and the potential and counterfactual (Section 5.4). There is also a final analysis dedicated to outstanding issues (Section 5.5) in which we outline other issues worthy of deeper reflection. The following section is dedicated to a discussion. In Section 7 we conclude and suggest future research.

## **2 Previous research**

Quechuan and Aymaran languages have been studied by many researchers. Key work for Quechua includes Adelaar & Muysken (2004), Albó (1970), Cerrón-Palomino (1987), Howard (2007), Mannheim (1991), Taylor (2006), and Urton & Llanos (1997). For Aymara, language descriptions are provided in Briggs (1976), Cerrón-Palomino (2000), Coler (2014), Hardman (1973), and Hardman (2001).

One theme of research on Quechua and Aymara pertains to the historically-founded structural convergence between the two languages, which were in close relationship over centuries. While a thorough overview is outside the scope of this chapter, the work of Adelaar (2012) and Cerrón-Palomino (2008) analyzes the convergence and parallel structures of the two languages. Adelaar with Muysken (2004) and Crevels & Muysken (2009) provide an excellent overview of the indigenous languages in the Andean area.

CA has also received some scholarly attention, especially in the last decades. Readers are referred to the overview of Escobar (2011) and the studies of Pfänder (2009), Merma-Molina (2007), Haboud (2008), and Mendoza (1991). All these authors attest to systematic changes and innovations in CA grammatical structure, attributed to the influence of indigenous languages. Previous work on CA has described the effect of Aymara (e.g. Hardman 1982, Quartararo 2017) and Quechua (e.g. Lee 1997, Quartararo 2017) on different grammatical levels. However, little work has been dedicated to mood and modality apart from reportative evidentiality and the tense-evidentiality interface. For Aymara, the tense-evidentiality interface has been described by e.g. Coler (2015) and Martínez Vera (2020). For Quechua, Faller (2002), Howard (2018), and Manley (2015) can be named, among others. The systematic influence on the use of CA tense has, to the best of our knowledge, first been described for Aymara-dominant contact varieties (Schumacher de Peña 1980, Hardman 1982, Stratford 1991) and has also been described by various authors for Quechua contact zones (e.g. Andrade Ciudad 2005, 2020, Dankel & Soto Rodríguez 2012, Escobar 1997, Escobar & Crespo del Río 2021, García Tesoro 2015, Klee & Ocampo 1995, Palacios & Pfänder 2018, Pfänder & Palacios 2013, Sánchez 2004). Most of the studies on Aymara and Quechua also anal-

yse reportative evidentiality. For Quechua, Kalt (2021) in a recent study specifically focuses on this category. Reportative evidentiality in different CA speaking zones has been studied in Andrade (2007), Babel (2009), Chang (2018), Dankel (2015), Feke (2004), Olbertz (2005), among others. The conceptions, categorizations, and associations proposed by these authors slightly diverge and can still be considered open for discussion in certain aspects. For example, a recent extensive overview of the different approaches to the evidential tenses is provided by Andrade Ciudad (2020: 79–93). Both the research on hearsay/reportative evidentiality as well as the tense-evidentiality interface are as fascinating as they are complex. However, as this topic is outside the scope of our contribution we do not detail it further. Our focus is on a trilingual comparison of the domain of mood and modality based on empirical spoken data, an approach that is unique in this configuration.

### 3 Theoretical framework

As the convergence of mood and modality in Aymara, Quechua and CA is an ongoing process, we do not measure contact induced change only in terms of direct lexical and structural borrowing. Instead, we understand it as a process of indirect and covert adjustment on different formal and functional levels based on the existing potentialites of the (sub-)systems of the respective languages. This perspective is based on Johanson's (2008) code-copying framework, and has found support in many studies on language contact over the last decade. For example, Babel & Pfänder (2014) show in a case study on the use of the past perfect (*había* + past participle) in CA, how “[t]he effects of language contact are the accumulation of communicative routines or habits, which speakers play on as they engage in creative language use” (Babel & Pfänder 2014: 254).

Johanson's (2008) framework models how parts of languages can be combined or copied selectively. He distinguishes four types of copies:

1. Combinatorial: Typical cases of combinatorial copies are loan translations or syntactic calques in which a structure or pattern of the target language is partially rearranged to fit into a scheme from the model language. For example, whereas (S)OV word order would be considered highly marked in PS grammar, speakers of Aymara- or Quechua-influenced CA frequently make use of it, particularly in emphatic contexts. This occurs with such frequency that its usage is loosing its marked status, reflecting the unmarked SOV pattern of Aymara and Quechua (Pfänder 2009: 102–108).

2. Material: Material copies include loanwords (which are frequent in Aymara, Quechua and CA) and phonological or morphological copies, though grammatical morphemes are only very rarely copied (see Pfänder 2009: 242).
3. Semantic: Semantic copies overlay the notional content of one language with the semantics of the other, as when speakers of CA, adapting the function of Quechua subordination suffixes, use the CA gerund construction mostly for adverbial subordination (Pfänder 2009: 139–147, Soto Rodríguez & Dankel in print).
4. Frequency based: Frequency copies adopt the usage of a feature from the model code in the corresponding feature of the target code: a well-known example is the higher percentage of explicit subject pronouns in the Spanish spoken in the US because of language contact with English (Silva-Corvalán 1994). Johanson (2008: 62) stresses the fact that copies cannot, by definition, be identical to their models. Most typically, the semantic functions of copies have not reached the same stage of grammaticalization as their models and their use is often pragmatically determined (2008: 70).

This framework helps explain how Aymara and Quechua mood and modality find their way into CA in notional transfer as a result of intense cultural and language contact. That is, it explains how languages can influence each other by borrowing conceptual notions (but not necessarily forms) that must be expressed in a speech community. Johanson's (2008: 62) framework is based in an extensive empirical-observational work on a well-established contact variety. It also receives support by comparable observations from other sub-fields. For example, consider the research of Jarvis & Pavlenko's (2008) on L2 acquisition processes. They find that the speakers' assumptions and perceptions about the L2 (in other words, their beliefs about the congruence between languages) fuel language transfer, also between typologically distant languages. From a more cognitive viewpoint, by focusing first on the dynamics in language processing mechanisms, Slobin (2016) explains that speakers who switch between languages frequently conceptualize the world in one language while speaking in another. This leads to contact-induced changes when speakers accommodate their "thinking for speaking" from the source language to the target language as an adjustment of language processing efficiency.

Accordingly, speakers' communicative routines, which work on a cognitive level but have been shaped culturally, affect their understanding of how target

languages work and give rise to linguistic outcomes in such a way that they are contextually and socially adequate. Speakers creatively operationalize the potential of the available linguistic forms to convey their semantic and pragmatic needs in context-dependent ways. Here we show that mood distinctions and other systemic possibilities to express modality in CA reflect the creative operationalization (that is, reinterpretation) of CA linguistic forms to convey a communicative routine that is fully grammaticalized in Aymara and Quechua.

## 4 Corpora

This section describes the source of the data for the analyses in Section 5. Aymara data are from a variety of the language spoken in Southern Peru. All language data come from Coler (2014) and directly from Edwin Banegas-Flores, a co-author on this contribution and a native speaker both of this Aymara variety and of CA.

The Quechua data alongside CA spoken in Quechua regions comes from the Corpus Quechua Español (CQE) gathered by Soto Rodríguez, a co-author on this publication. The CQE comprises approximately 9 hours of recordings made in different locations in the Cochabamba region of Bolivia in 2009 and 2011. It contains material from daily conversations, interviews, and radio and television broadcasts in both languages. Note that Soto Rodríguez himself is a native speaker both of Quechua and CA.

Additional CA data come from the Corpus Heroina gathered by Dankel and colleagues and published in Dankel & Pagel (2012). For additional information on this data, see also Dankel (2015). Other sources are cited when relevant.

## 5 Analyses

This section provides an analysis of mood and modality in Aymara, Quechua and CA from our empirical data. Since mood and modality are complex, and as we want to illustrate phenomena at different grammatical levels across different parts of the language systems, we have divided the analysis as follows: Section 5.1 tense/evidentiality, Section 5.2 hearsay and quotatives, Section 5.3 inferentials/conjecturals, Section 5.4 potential and counterfactuals, and Section 5.5 outstanding issues, which deals with other observations that do not fit neatly into the preceding sections.

Each of these sections includes a list of elicited examples that shows which markers in the three languages express the respective modal function. In the Aymara and Quechua examples in this section, we always provide both a CA and PS

translation of the indigenous language. The CA translations were provided by the indigenous Peruvian native-speaking authors. The PS translations were provided by asking Spain-born native speakers to translate from the English translations into PS.

### **5.1 Tense/evidentiality**

Contrary to PS, where the tense paradigm has precise temporal functions, grammatical tense in Aymara and Quechua is more modal-evidential than temporal. Both indigenous languages have four mood-tense distinctions. While their scopes differ somewhat, they can be described in roughly the same terms:

1. Unmarked: A non-future (present and immediate past) directly experienced by the speaker. This tense is not glossed in Quechua. In Aymara it is glossed as “simple tense” (SIM)
2. Experienced past: A past directly experienced by the speaker. We gloss members of this paradigm as “experienced past” (EXP.PAST)
3. Non-experienced past: A past not experienced by the speaker. Note that even events which occurred in the recent past, but for which the speaker was e.g. intoxicated or sleepwalking are inflected into this paradigm – thus demonstrating that the tense distinction is evidential. We gloss members of this paradigm as “non-experienced past” (NEXP.PAST)
4. Future: A future not experienced by the speaker, glossed as FUT.

We will show that the three Aymara and Quechua non-future tenses map to CA perfect and pluperfect temporal morphology in a way that indicates contact-induced change.<sup>3</sup> This change can be considered a selective semantic copy in Johanson’s (2008) framework and therefore is barely noticeable on the structural surface but becomes evident in language use. In examples (1)–(6) we provide constructed examples to illustrate parallel usage.

---

<sup>3</sup>Note that Aymara and Quechua non-future tenses map to other CA tenses, not just the perfect and pluperfect, as CA of course makes use of other tenses. This mapping falls outside the scope of this paper.

**Personal knowledge:**

- (1) Experienced past in Aymara:  
uma-na  
drink-3.SUBJ.EXP.PAST  
'he drank' (the speaker witnessed it)
- (2) Experienced past in Quechua:  
uyja-rqa  
drink-3.SUBJ.EXP.PAST  
'he drank' (the speaker witnessed it)
- (3) Perfect tense in Castellano Andino:  
ha       tomado  
have.3 drink.PTCP  
'he drank' (the speaker witnessed it)

**Non-personal knowledge:**

- (4) Non-experienced past in Aymara:  
uma-tayna  
drink-3.SUBJ.NEXP.PAST  
'he drank' (the speaker did not witness it)
- (5) Non-experienced past in Quechua:  
uyja-sqa  
drink-3.SUBJ.NEXP.PAST  
'he drank' (the speaker did not witness it)
- (6) Pluperfect in Castellano Andino:  
había               tomado  
have.3.PAST.IPFV drink.PTCP  
'he drank' (the speaker did not witness it)

### 5.1.1 Aymara tense/evidentiality

Beginning with Aymara, personal knowledge can be expressed with the simple tense (as in (7)) or the experienced past tense (as in (8)). In both of these tenses, the speaker asserts first-hand direct knowledge over the event.

The non-experienced past, in contrast, not only refers to a remote and mythological past, but also a past which was not witnessed by the speaker. That is, an event can be marked in the non-experienced past even if it occurred during the speaker's lifetime if they lack personal knowledge of it. This is evident in (9) which occurred moments before the speech-act. However, as the speaker did not witness the event, it does not receive either the simple or the experienced personal knowledge tenses. Likewise, events which were witnessed by the speaker, but which they forgot about (owing to amnesia, intoxication, senility, or similar) are in the non-experienced past.

- (9) jupa-w qullq-∅ chura-taytam-x  
     he-DECL money-ACC give-3.SUBJ.2OBJ.NEXP.PAST-TOP  
     ‘He gave you the money.’ (Coler 2014: 421)  
     CA: ‘Él te **había** dado la plata.’  
     PS: ‘(Él) Te dio el dinero.’

The non-experienced past is also used to express mirative meanings (e.g. Martínez Vera 2020: 83). Here, the term “mirativity” is taken as the “linguistic marking of an utterance as conveying information which is new or unexpected to the speaker” (DeLancey 2001). Compare (10) with (11), the former is in the non-experienced past and the latter is in the experienced past.

- (10) uñja-taysta  
see-2.SUBJ.1.OBJ.NEXP.PAST  
'You saw me! (without my knowledge)' (Coler 2014: 422)  
CA: '**¡Me habías visto!**'  
PS: '**¡(Tú) me viste!** (no me di cuenta)'

- (11) *uñj-ista*  
 see-2.SUBJ.1.OBJ.EXP.PAST  
 ‘You saw me.’  
 CA: ‘Me **has** visto.’  
 PS: ‘Me viste.’

### 5.1.2 Quechua tense/evidentiality

This distinction in Quechua is largely parallel to that described for Aymara. Examples (12) and (13) show the distinction between experienced and non-experienced past. The speaker in these two examples is speaking about a break-in that happened in her village which she did not personally witness. Although she has personal knowledge of her own whereabouts the moment the break-in happened, the knowledge of the victim crying for help is indirect in that the speaker did not witness it herself. (In this, and all following examples, CA loanwords are italicized.)

- (12) *miércoles tarde nuqa ma ka-rqa-ni-chu chay-pi*  
 Wednesday evening I no be-EXP.PAST-1SG-NEG this-LOC  
 ‘I was not there on Wednesday evening.’  
 CA: ‘El miércoles en la tarde yo no estuve ahí’ (CQE)  
 PS: ‘No estuve allí el miércoles por la tarde.’
- (13) *eh ajina encapuchado ka-xti-n auxilio maña-sqa*  
 eh like.this hooded be-SR-3SG help ask-NEXP.PAST  
 ‘When the hooded guys were there, she asked for help (I did not witness it).’ (Dankel & Soto Rodríguez 2012: 96)  
 CA: ‘Cuando ha visto que estaban encapuchados había pedido auxilio.’ (CQE)  
 PS: Cuando los hombres encapuchados llegaron, ella pidió ayuda (yo no lo presencié).

As shown in (14), personal knowledge marking is especially relevant in situations of social or even legal accountability. In these cases its function could be described as providing testimonial specification about a state of affairs, which explains its use with negation markers. The example starts with the telling of a piece of common, undisputed knowledge with the unmarked non-future tense: they found someone dead.

- (14) wañu-sqa-lla-ta taripa-nku. Madre Obrera chay chay-s-itu-pi...  
 die-PART-LIM-ACC find-3SG Madre Obrera there there-EU-DIM-LOC  
 ‘They found him already dead. Near the Madre Obrera (hospital)...’  
 (Soto Rodríguez 2002: 165)  
 CA: ‘Muerto lo han encontrado. Cerca del Madre Obrera...’  
 PS: ‘(Se) lo encontraron ya muerto. Cerca del Hospital Madre Obrera...’

The speaker, who knows the deceased, then continues with the reporting of the circumstances where she carefully marks the details in a way she cannot be held accountable for the state of affairs.

- (15) qayna      *tarde*      lluxsi-pu-sqa,      ari.      Calle-pi-chus  
 yesterday afternoon go.out-REFL-NEXP.PAST MOD.INT street-LOC-CONJ  
 toma-rqa,      mana nuqa yacha-rqa-ni-chu.  
 drink-EXP.PAST no      I      know-EXP.PAST-1-NEG  
 ‘He went out yesterday afternoon (I did not witness it). Maybe he drank in the street, I didn’t know it.’ (Soto Rodríguez 2002: 165)  
 CA: ‘Había salido ayer en la tarde, en la calle ha debido tomar, yo no sabía.’  
 PS: ‘Salió ayer por la tarde (no lo ví). Bebía, quizás en la calle, yo no lo sabía.’

With the non-experienced past suffix she signals that she did not know of his absence. Interestingly, she marks her supposition that he drank in the street and her assurance that she did not know that he drank in the street with an experienced past suffix. In both cases, this emphasizes her accountability. Such examples demonstrate how these suffixes are more modal than temporal.

As in Aymara, the Quechua non-experienced past can express the mirative.<sup>4</sup>

- (16) Erma      qhawa-yku-sa-sqa      kay lluqallu  
 Erma.ACC look-DIR-PROG-NEXP.PAST this boy  
 ‘Oh! This guy is interested in Erma!’  
 CA: ‘A la Erma se había estado mirando este chico.’ (CQE)  
 PS: ‘¡Oh! ¡Este chico está interesado en Erma!’

---

<sup>4</sup>While the Quechua data in this contribution is from Southern Quechua, note that the Central Peruvian variety of Tarma Quechua has a grammatical paradigm that exclusively conveys mirative meaning (Adelaar 2013).

### 5.1.3 Castellano Andino tense/evidentiality

We show that speakers of CA map the modal-evidential distinction to the Spanish perfect and pluperfect temporal morphology. The speaker in example (17) recounts a break-in at her house. As she did not personally witness the burglars entering into her house, she uses the pluperfect form (*habían entrado*).<sup>5</sup> However, as she did experience directly that her audio system was missing when she returned, for this part of her telling she uses the perfect form (*han llevado*).

- (17) Los ladrones **habían entrado** y se **han llevado** equipos de  
the burglars had entered and they have carried equipments of  
sonido, ¿no?  
sound no  
'The burglars entered (I did not witness it) and they took the audio  
system with them, right?' (Dankel & Pagel 2012)  
PS: 'Los ladrones entraron (no lo presencié) y se llevaron el sistema de  
música, ¿sabes?'

In (18) the use of perfect tense and pluperfect tense forms according to the experience of the speaker can be observed. Furthermore, the pluperfect also gets a mirative reading through its sequential position in this context. The speaker tells about how he learned by arriving at his destination (direct experience marked with the perfect tense form), that the house he was looking for was not where he thought it was (newly revealed knowledge, not formerly experienced, marked with the pluperfect tense form).

- (18) Y fuimos, hemos llegado. No había sido su casa en el centro,  
and we.went we.have arrived no had been his house in the center  
sino había sido más alejadito de Quillacollo, como es, como en aquí,  
but had been more far.away from Quillacollo as is as in here  
como Achocalla así, un pueblito así alejadito.  
as Achocalla thus, a little.town like far.away  
'And we went, we arrived. Her house was not in the town, it was quite a  
bit further away from Quillacollo, like is, like here, like Achocalla, a small  
town, quite a bit outside.' (Quelca Huanca 2006: 206)

---

<sup>5</sup>Interestingly, also in this CA example, speaker accountability plays a role for why the non-experienced past marker is chosen, although there might be some directly experienced evidence of how and where the burglars entered. By distancing herself from knowing anything about how the burglars could enter into the house, the speaker cannot be held accountable for any overlooked security measures.

PS: ‘Y fuimos, llegamos. Su casa no estaba en la ciudad (centro), estaba bastante lejos de Quillacollo, como es, como aquí, como Achocalla, una ciudad pequeña, bastante a las afueras.’

## 5.2 Hearsay and quotatives

Both in Aymara and Quechua, hearsay marking and quotatives are a grammaticalized part of a culturally-relevant evidential subsystem. Both languages have no grammatical mechanism to express indirect speech and show certain parallelism in the grammatical marking of evidentiality. In PS, on the other hand, hearsay and quotatives are expressed only selectively by lexical and discourse-pragmatic ways (e.g. *se dice que* ‘it is said that’). There are no comparable systematically used markers. Nevertheless, even with reduced possibilities for the direct transfer of grammatical morphology due to typological incompatibility, CA shows contact induced incipient grammaticalization of hearsay and quotative markers as a result of a step-by-step development based on the potential of constructions with say-verbs that express these two functions in the appropriate contexts: *dice / dice que / dizque* are used as reportative particles to express hearsay, whereas *diciendo* ‘saying’ is used for as a quotative, often also postponed to the quoted proposition. Again we observe a selective copy of a semantic notion crucial in Aymara and Quechua. There is also evidence of the reverse influence of CA *dice / dice que* and *diciendo* on some varieties of Aymara and Quechua, where new reportative strategies emerge modeled on CA – a combinatorial copy, so to speak, in Johanson’s (2008) framework. This can be seen as a strong indication that speakers do not distinguish between different systems for each language, but have developed diverse linguistic resources for one system. This is summarized in the constructed examples provided in (19)–(24).

### Hearsay:

- (19) Verb ‘to say’ (subordinated) in Aymara:

uma-nt-i-w	sa.s
drink-iw-3.SUBJ.TOP	say.SUBR
‘She drinks, it is said.’	

- (20) Dedicated suffix or verb ‘to say’ (inflected) in Quechua:

ujya-n	ni-n
drink-3.SUBJ	say-3.SUBJ
‘She drinks, it is said.’	

- (21) Dedicated verbs (*dice que*, *dicen que*, *dizque*) in Castellano Andino:  
 Toma,            dice.  
 drink.3.SUBJ say.3.SUBJ  
 ‘She drinks, it is said.’

**Quotative:**

- (22) Verb ‘to say’ (inflected) in Aymara:  
 Uma-nt-i-x            s-i-w  
 drink-iw-3.SUBJ-TOP say-3.SUBJ-DECL  
 “She drinks”, he says.
- (23) Verb ‘to say’ (inflected) in Quechua:  
 Ujya-n            ni-spa    n-in  
 drink-3.SUBJ say-GER say-3.SUBJ  
 “She drinks”, he says.
- (24) Dedicated verbs (*dice*, *diciendo*) in Castellano Andino:  
 “Toma”            diciendo dice.  
 drink.3.SUBJj say.PROG say.3.SUBJ  
 “She drinks”, he says.

Within the subsystems for hearsay and quotative marking there is a lot of variation regarding how the hearsay and quotative are conveyed across language varieties. This dynamic instability holds, interestingly enough, for Aymara, Quechua, and CA. However, the presence and use of these markers are highly consistent.

### 5.2.1 Aymara hearsay and quotatives

In the Intermediate Aymara of Southern Peru, there is a distinction between *s-i-w* ‘s/he says’, which became lexicalized and can be seen as a marker meaning contextually, ‘it is said’ or ‘one says’ (referring to common knowledge) and *sas* which is used as a quotative. Examples follow in (25) and (26).

- (25) uka usu-x    wali *phiyu*-∅-tayna-w  
 that bear-TOP very bad-COP.VBZ-3.SUBJ.NEXP.PAST-DECL  
 s-i-w  
 say-3.SUBJ.SIM-DECL  
 ‘That bear was very bad, they say.’

CA: ‘Muy malo había sido el oso, dice.

PS: ‘El oso era muy malo / malvado, decían.’

- (26) uka-t timpranu-t sara-tan sa.s imilla-nak  
 that-ABL early-ABL go-1.INCL.SUBJ.SIM say girl-PL  
 ‘After, “We go early” the girls say.’  
 CA:‘Después, temprano nos vamos, dicen las chicas.’  
 PS: ‘Después, “(nos) vamos temprano”, dicen las chicas.’

### 5.2.2 Quechua hearsay and quotatives

Both the variety of Quechua investigated here and the variety spoken in Cusco have a reportative suffix *-si* which marks hearsay, as in (27).

- (27) wakin-si maqa-mu-nku  
 some-REP hit-CIS-3.PL  
 ‘Some hit him, it is said.’ (Faller 2002: 22)  
 CA: ‘Algunos le pegaron, dice.’  
 PS: ‘Dicen que alguno (de ellos) / alguien le pegó.’

However, in corpus data from natural conversations in Quechua, the use of lexicalized *nin* (say-3) as a particle with the same hearsay function, replaces the suffix. This new reportative strategy seems to have emerged as a case of reverse influence of CA *dice* (Olbertz 2005, Dankel & Soto Rodríguez 2012) on these varieties of Quechua.

- (28) askha dolares-ni-n ka-sqa ni-n  
 many dollars-EU-3 be-NEXP.PAST say-3  
 ‘It was a lot of dollars, it is said.’ (Dankel & Soto Rodríguez 2012: 96)  
 CA: ‘Había tenido muchos dólares, dice.’  
 PS: ‘Dicen que fueron muchos dólares.’

The Quechua quotative also emerged from a say-verb construction, in this case *diciendo* ‘saying’. As we can see in (29), it usually appears together with a finite say-verb form.

- (29) ni-tax chaya-chi-mu-wa-n-chu ni-spa ni-ri-sa-n  
 no-CONT arrive-CAUS-CIS-1.OBJ-3-NEG say-GER say-INCH-PROG-3  
 mama-yki-qa  
 mother-2-TOP  
 ‘He didn’t bring (his girlfriend) home either (that’s what) your mother is saying. (CQE)

CA: ‘Ni a la casa no le ha traído está diciendo tu madre.’

PS: ‘Tampoco la trajo [a su novia] a casa, [eso es lo que] está diciendo / dice tu madre’

### 5.2.3 Castellano Andino hearsay and quotatives

The situation for CA is very similar to what is attested in Aymara and Quechua. The hearsay marker emerged from the same model. Though for CA it is often lexicalized in combination with the relative pronoun *que*, as in *dice que*:

- (30) Villa Pagador **dice que** es inmenso eso.  
 Villa Pagador say that is enormous that  
 ‘Villa Pagador is said to be enormous.’  
 PS: ‘Parece que Villa Pagador es enorme.’

Note that *dice que* as an incipient grammaticalization structurally still behaves as a verb + complement construction. Functionally, however, the hearsay marking is unambiguous (see Dankel 2015).<sup>6</sup>

The quotative marker is based on the gerund form of *decir*, i.e. *diciendo*. Also in CA, it frequently is constructed with a say-verb that introduces direct speech and *diciendo* to mark the quoted utterance, as in (31).

- (31) Yo le he avisado a doña Simona pue... estoy saliendo doña Simona  
 I her have told to Mrs. Simona well I.am leaving Mrs. Simona  
 un rato **diciendo**.  
 a while saying  
 ‘I told Mrs. Simona “I am leaving a while”, I said.’ (CQE)  
 PS: ‘(Yo) le dije a la Sra. Simona: “salgo un ratito”’

However, also for the quotative we can speak of a situation of incipient grammaticalization, which means that there is still a lot of paradigmatic and syntagmatic variability regarding the say-verbs that occur with *diciendo* and at best incipient fixation (Lehmann 1982), whereas in Quechua and Aymara the say-verb and the quotative tend to occurs phrase-final and in a fixed order.

The fact that there is not only a clear hearsay marking device as well as a quotative marking device in all three codes, but also that these markers emerge

<sup>6</sup>For a more complete picture, it is important to emphasize the incipient grammaticalization status of the CA hearsay markers. This means that there is still a highly variable use of forms (*dice*, *dicen*, *dice que*), also in terms of their functional proximity to *decir* ‘say’. We also find a fully fused form *dizque*. However in the CA spoken in the areas investigated, *dizque* is used with an additional meaning of gossip or doubt, whereas *dice que* can also be used in contexts where hearsay information is used to convey epistemic authority of the distant source.

from the grammaticalization of say-verb constructions in all three varieties is a strong indication for the three codes converging into one system.

### 5.3 Inference/Conjecture

Inference/conjecture is integral to Aymara and Quechua morphology. The use of these markers in naturally occurring speech remains under-researched for both languages. The research on inference/conjecture in CA is also minimal. This is likely because the changes in CA caused by a need for inference/conjecture marking are often unnoticed. Hardman (1982), for example, mentions some unusual uses of *seguramente* in the CA of La Paz. In our own work we also found *seguramente* in marked syntactic contexts where speakers refer to inferential knowledge (see Dankel & Soto Rodríguez 2012). As will be illustrated presently, there are additional techniques in CA, reflecting Aymara and Quechua markers for inference and conjecture. Our data shows that there are differences in the manifestation of the inferencial/conjectural in Aymara and Quechua. This is elaborated in the constructed examples provided in (32a)–(34b)

#### Inferential:

- (32) *-spha*, *-pacha* or counterfactual paradigm (intrinsically deduced) in Aymara:
- a. Uma-nta-spha-wa  
drink-IW-3.SUBJ.3.OBJ.INFR-DECL  
'He must have drunk it.'
  - b. Uma-nt-irki-wa  
drink-IW-3.SUBJ.3.OBJ.CF-DECL  
'Maybe he could/should have drunk it.'

#### Conjectural:

- (33) *-chi* (extrinsically deduced) in Aymara:  
Uma-nt-ch-i-xall  
drink-IW-CNJ-3.SUBJ.3.OBJ.SIM-CNJ  
'Surely he must have drunk'

**Inferential/Conjectural:**

- (34) *-cha, -sina, -chus, icha, ichas* in Quechua:

- a. ujya-n-cha  
drink-3.SUBJ-INF  
'He must have drunk.'
- b. **icha** ujya-n  
maybe drink-3.SUBJ  
'Maybe he drunk.'

- (35) *Seguramente, seguro (que), deber, haiga, creer*, among others in CA.:

- a. **Seguramente** habrá tomado.  
surely have.3.FUT drink.PTCP  
'Surely he must have drunk.'
- b. **Debe** haber tomado.  
must.3 have.INF drink.PTCP  
'He must have drunk.'
- c. **Haiga** tomado.  
have.3.SBJV drink.PTCP  
'He must have drunk.'

### 5.3.1 Aymara inference/conjecture

Aymara seems to make a fine-grained distinction between extrinsically (relying on direct evidence) and intrinsically (relying on logical reasoning) deduced knowledge. The former are treated in (36–39) and the latter in (40–41).

- (36) kha-n-x trucha-x ut.ja-spha-w  
yonder-LOC-TOP trout-TOP exist-3.SUBJ.INFR-DECL  
'There must be trout yonder.' (Coler 2014: 291)  
CA: 'Allá **debe de** haber trucha.'  
PS: 'Debe haber truchas allí / allá.'

Here, the speaker's claim is based on direct, extrinsically-deduced knowledge of the world. He knows trouts tend to hide in the specific location in that river bend. Types of inference like these, are expressed in Aymara either with a dedicated suffix (*-spha* or *-pacha*, depending on the region), as in (37). In this example, the speaker speculates about a relative who she has not seen in a long time.

- (37) lik'i-∅-nta-spha-w  
fat-COP.VBZ-IW-3.SUBJ.INFR-DECL  
'She must have fattened up.' (Coler 2014: 439)  
CA: 'Seguro que ha engordado.'  
PS: 'Debe haber engordado.'

Next, members of the counterfactual paradigms overlap with inferential/conjectural marking in Aymara and can convey a myriad of meanings. A paradigmatic one appears in (38) which is premised on the evidence that the accused trickster has revealed himself to be disingenuous and so is liable to deceive again. Consider also (39) in which a member of the present counterfactual paradigm is used to make a question-like expression without the use of interrogative morphology. Observe that the phrase in the latter is not translated with *seguro*, though the meaning is less inferential than the one in the former. Indeed, there is significant contextual variation in how this morpheme is translated.

- (38) jicha-x wasita-mp *inkaña-jw-itasma-x*  
now-TOP again-COM deceive-BFR-2.SUBJ.1.OBJ.PRES.CF-TOP  
'Now again you may deceive me.' (Coler 2014: 434)  
CA: 'Ahora, seguro que de nuevo me puedes engañar'  
PS: 'Ahora, puedes engañarme de nuevo'
- (39) juma-x thuqu-sma-x  
you-TOP dance-2.SUBJ.PRES.CF-TOP  
'Might you dance?' (Coler 2014: 257)  
CA: '¿Tú por qué no bailas?'  
PS: '¿Bailarías? / ¿Querrías bailar? / ¿Te gustaría bailar?'

All of these notions are extrinsic deductions based on direct experiences. These are differentiated from intrinsically deduced knowledge for which the speaker relies on knowledge arrived at through deductive reasoning based on learned experience of the world. This is expressed by the conjectural suffix, as in (40) in which the speaker knows that someone caught some trout in this spot recently, but did not see trout there himself. However, given what he knows about the world, there are probably more .

- (40) kha-n        *trucha-x ut.ja-s-ch-i-s*  
yonder-LOC trout-TOP exist-REFL-CNJ-3.SUBJ.SIM-AD  
'Perhaps there are trout yonder.'  
CA: 'Tal vez todavía allá hay trucha.' (Coler 2014: 290)  
PS: 'Tal vez haya truchas allí / allá.'

Finally, in some varieties of Aymara, like those spoken in the Southern Peruvian highlands, there is a phrase-final suffix *-jalla* which can also express the conjectural. This suffix typically co-occurs with the conjectural *-chi*.<sup>7</sup> Consider the minimal pair in (41) and (42) and note how the CA (and PS and English) translations are identical, even though the meanings expressed in Aymara are different, again on the basis of the type of reasoning used.

(41) may be uttered when a speaker notes that a colleague did not arrive at work (i.e. intrinsic reasoning). (42) may be uttered when a speaker observes a caregiver leaving the referent's house looking distraught (i.e. extrinsic reasoning).

- (41) usu-ta-∅-s-ch-i-xall  
sick-RE-COP.VBZ-PROG-CNJ-3.SUBJ.SIM-CNJ  
'He must be sick.' (Banegas-Flores, p.c.)  
CA: 'Seguro que está enfermo.'  
PS: 'Debe estar enfermo (no ha venido a trabajar).'
- (42) usu-ta-∅-spha-w  
sick-RE-COP.VBZ-3.SUBJ.INFR-DECL  
'He must be sick.' (Coler 2014: 564)  
CA: 'Seguro que está enfermo.'  
PS: 'Debe estar enfermo (la veo muy preocupada por él).'

### 5.3.2 Quechua inference/conjecture

Inference/conjecture marking is also very common in Quechua, though it does not make a fine-grained distinction between the expression of intrinsically (based on logical reasoning) and extrinsically (based on direct evidence) deduced knowledge on the morphological level. Both are expressed with *-cha*. While Quechua *-cha* isn't identical to the Aymara *-chi*, in some contexts their roles are very similar. Compare example (43) with (44).

- (43) Uta-p-∅-x                    alxa-wj-chi-ni-xall  
house-3.POSS-ACC-TOP sell-BFR-CNJ-3.SUBJ.3.OBJ.FUT-CNJ  
'Probably she will sell her house.'  
CA: 'Seguramente va a vender su casa.'  
PS: 'Probablemente venderá su casa.'

---

<sup>7</sup>The semantics and distribution of this suffix are described in detail in Coler (2014: 560).

- (44) *vende-nqa-cha wasi-n-ta á*  
sell-FUT-INF house-3-ACC MOD.INT  
'Probably she will sell her house.'  
CA: 'Seguramente va a vender su casa.'  
PS: 'Probablemente venderá su casa.'

The particle *icha* (or *ichas*) is used in similar contexts as the co-occurrence of the counterfactual and the conjectural in Aymara (as in (54)). *icha* expresses conjecture in the sense of an expected event and can be considered to be at the boundary between inferentials/conjecturals and potentials/counterfactuals (see the analysis on the potential and counterfactual mood in Section 5.4)

- (45) *ichas chaya-chi-mu-wan-pis ni-n pero á*  
maybe arrive-CAUS-CIS-3.SUBJ.1.OBJ-CONC say-3 but MOD.INT  
'Yet, maybe he brought here, she says, alright.' (CQE)  
CA: 'Puede que tal vez le ha traído, dice pero pues.'  
PS: 'Aunque tal vez la trajo, dice, ¿sabes?'

### 5.3.3 Castellano Andino inferentials and conjecturals

The aforementioned inference/conjecture functions are present in CA, yet are often unnoticed, as they are still at the beginning of being grammaticalized. That is, their individual lexico-syntactic appearance may not be seen as new in comparison to other Spanish varieties. Nevertheless, in their overall systemic-functional configuration and frequency they are unattested in other Spanish varieties.

Constructions with *seguramente* (*que*) or *seguro* (*que*) are used in CA to express knowledge arrived at through intrinsic and extrinsic deduction. (46), based on the latter, is produced jokingly in response to another speaker's anecdote about looking for gold at an excavation site.

- (46) *Seguro cría de los españoles eres.*  
surely descendent of the Spanish you.are  
'Then presumably you are a descendent of Spaniards.' (CQE).  
PS: 'Entonces parece que eres descendiente de españoles.'

In (47), the deduction process is internal logical reasoning. Clearly, the line between direct evidence-based reasoning and reasoning based on already established world knowledge is not always clear cut.

- (47) Después, siempre hemos sido aficionados a la música porque mi after always we.have been amateurs to the music because my papá tocaba mandolina, tocaba charango, nunca lo he visto tocar father played mandolin played charango never it have seen to.play guitarra, pero **seguramente que** sabía.  
guitar but surely that knew  
‘Also, we always have been music enthusiasts, because my father played mandolin, played charango, I never saw him play the guitar, but he surely knew.’ (Dankel & Pagel 2012: 107)  
PS: ‘Además, siempre hemos sido aficionados/as a la música, porque mi padre tocaba la mandolina, tocaba el charango, nunca lo he visto tocar la guitarra, pero seguro que sabía.’

The narrator explains that he and his siblings became music enthusiasts because of his father’s skill with stringed instruments. Interestingly, he signals a difference in his mode of access to this knowledge between “directly witnessed” for the mandolin and charango and inferred (introduced by the contrast *nunca lo he visto ...pero seguramente que*) for the guitar. For non-Andean speakers, such precise evidential marking would only occur if explicitly accounted for by the interlocutor or the context of their telling, but is produced systematically by CA speakers. This is a clear case for a change in the way of speaking because of a different way of thinking, in other words, Slobins (2016) “thinking for speaking”.

CA constructions with *deber* can also express inferential/conjectural meanings, as in 48. In PS the last two lines of this example would be *¿Qué edad tenías?* and *Probablemente tendría unos 12 años*.

- (48) A: Mi hermano estaba tirado en la cama quemado aquí con una ampolla.  
E: ¿Cuántos años tenías?  
A: Más o menos unos 12 años **yo he debido tener.**  
A: My brother was laying on the bed, burned here, with a blister.  
E: How old were you?  
A: I **would probably have been** around 12 years old.  
(Dankel & Pagel 2012: 195)

Particularly in the area where Aymara is spoken with CA, speakers use an archaic form of the perfect subjunctive construction, *haiga* + participle, as an inferential construction. This is evident in the following testimony of a confrontation during a social protest, where the speaker got shot. The speaker infers that the military started to shoot because they ran out of tear gas.

- (49) A lo así nomás, sus gases se haiga acabado, qué será.  
to what thus just their gases themselves have run.out what will.be  
Ahí nos han baleado con armas de guerra.  
there us have shot with weapons of war  
'Suddenly, surely they ran out of tear gas, I don't know. They started to shoot with their war guns.'  
PS: 'De repente, seguramente se les acabó el gas lacrimógeno, no lo sé.  
Empezaron a disparar con sus armas de guerra.'

The conjectural reading of the subjunctive in (49) can be considered in a state of grammaticalization, as other varieties of Spanish, particularly Peninsular Spanish, do not allow this reading without the co-occurrence of additional conjectural adverbs. This example is even more interesting because the construction is followed by the question-like *qué será* ('what will be'), as a dubitative, similar to what we see in Quechua with the dubitative *-chus*. These co-occurrence of inferentials/conjecturals with certainty markers seem to appear in Aymara, Quechua and especially CA.<sup>8</sup> We will briefly discuss *-chus* and other certainty markers in Section 5.5. Similar to the potential/counterfactual, they are closely intertwined with inference and conjecture and allow for highly nuanced elaborations of knowledge states and responsibilities. They also deserve more attention in future research.

## 5.4 Potential and counterfactual mood

As alluded to earlier, potential/counterfactual mood is closely intertwined with inference and conjecture (treated in Section 5.3). Aymara uses the counterfactual mood paradigm to refer to possible events and conditional constructions and Quechua has a potential suffix.

### 5.4.1 Aymara counterfactual mood

Aymara does not have a form that is glossed as 'potential', though the conjectural, inferential and counterfactual forms can be employed to express potential meanings. The counterfactual is the only marked mood in Aymara (Coler 2014: 428). It is expressed with recourse to a pair of paradigms that inflect for simple and past tense. Usually the counterfactual is translated into CA with the modal verb *deber* 'should/must'.

---

<sup>8</sup>Especially for the dynamic situation in CA, it is not surprising that speakers combine a variety of strategies to get as close as possible to their culturally shaped communicative goals.

- (50) *jut-irki-w*  
 come-3.SUBJ.SIM.CF-DECL  
 ‘He should/must come.’  
 ‘Debe venir.’  
 Evidence: He left something behind, he will certainly return for it.  
 CA: ‘Debe venir.’  
 PS: ‘Debería venir. / Tiene que venir.’

The following minimal triplet facilitates comparisons of the inferential evidential and counterfactual and conjectural suffixes (both of which were treated in Section 5.3) and the counterfactual, both in Aymara and CA.

- (51) *kha-n-x trucha-x ut.ja-spha-w*  
 yonder-LOC-TOP trout-TOP exist-3.SUBJ.INFR-DECL  
 ‘There must be trout yonder.’ (Coler 2014: 456)  
 CA: ‘Allá debe de haber trucha.’  
 PS: ‘Debe haber truchas allí / allá.’  
 Evidence: Speaker infers this from his knowledge in what specific location in the river bend trout tend to hide.
- (52) *kha-n trucha-x ut.ja-s-ch-i-s*  
 yonder-LOC trout-TOP exist-REFL-CNJ-3.SUBJ.SIM-AD  
 ‘Perhaps there are trout yonder.’ (Coler 2014: 456)  
 CA: ‘Tal vez todavía allá hay trucha.’  
 PS: ‘Tal vez haya truchas allí / allá.’  
 Evidence: Speaker knows that someone caught some trout in this spot recently so there are probably more there.
- (53) *kha-n-x trucha-x ut.ja-s.k-irki-w*  
 yonder-LOC-TOP trout-TOP exist-PROG-3.SUBJ.PRES.CF-DECL  
 ‘There may be trout yonder.’ (Coler 2014: 456)  
 CA: ‘Truchas podrían estar viviendo allí.’  
 PS: ‘Puede que haya truchas allí / allá.’

The conjectural suffix *-jalla* can co-occur with a member from the counterfactual paradigm, as in (54).

- (54) *Salur-t'a-s-irki-pun-jall*  
 salute-M-REFL-3.SUBJ.PRES.CF-EM-CNJ  
 ‘She would really say hello.’(Coler 2014: 517)  
 CA: ‘Me podría saludar.’  
 PS: ‘Ella realmente saludaría.’

#### 5.4.2 Quechua potential mood

The Quechua potential is expressed through the suffix *-man* and its variant *-wax* added to an inflected verb form thus transforming it into a non-finite verbal constructions that can also convey counterfactuality. The suffix can be used for notions such as ability, possibility, enabling conditions, and permissions (as in (55), in which a speaker is calling in to a radio show to request a song). In conversational interaction, these notions also allow for a directive use of the form.

- (55) Takiy-s-itu-ta      maña-ri-ku-yki-**man**-chu Doña Rosmery  
song-EUF-DIM-ACC ask-INCH-REFL-2-POT-INT Mrs. Rosmery  
'May I request a song, Mrs Rosmery?' (CQE)  
CA: '¿Puedo pedir una canción, Doña Rosmery?'  
PS: '¿Puedo pedir una canción, Sra. Rosmery?'

Particularly with directive functions, the potential suffix is used in counterfactual constructions. The pragmatic use of *-man* lets speakers refer to events that have occurred which they would have wished to be different. Also, the same suffix in combination with the auxiliary *kay* 'be' in the past allows the speaker to express explicit counterfactual events, as in (56), where the speaker tells of his participation in an armed confrontation and reflects on what may have happened.

- (56) ima-ta      may *tiempo-pis* wañu-y-**man**-pis ka-rqa á  
what-ACC how time-CONC die-1-POT-CONC be-EXP MOD.INT  
'You know, I would be dead a long time ago.' (CQE)  
CA: 'Sabes qué, hace tiempo tal vez ya hubiera muerto, pues.'  
PS: "Sabes que yo (ya) estaría muerto/a hace mucho tiempo." / 'Yo estaría muerto/a hace mucho tiempo, ¿sabes?"'

#### 5.4.3 Castellano Andino counterfactual and potential mood

CA mechanisms to refer to potential events reflects Aymara and Quechua indirectly. The most prominent strategies are the use of *poder* 'can' (as in (57)) to express hypothetical events (instead of its canonic dynamic modal function to express ability and willingness or deontic uses),<sup>9</sup> constructions with modal parti-

<sup>9</sup>The use of periphrastic constructions with *poder* 'can' in the Andean zone, on the other hand, is causing analytical processes in modal constructions of Quechua (see Haimovich 2016). It is common to find the use of periphrastic modal constructions with the auxiliary *atiy* 'can' such as *imatapis rantikuyta atinki* 'you would/could buy anything' (CQE) instead of *imatapis*

cles like *tal vez* and *capaz*<sup>10</sup> (in (58) and (59), respectively) that clearly differ from Peninsular Spanish patterns, and the use of the pluperfect subjunctive in single utterances with an implicit second component instead of a bipartite subordinate conditional structure<sup>11</sup> (see (60) in which the speaker complains about the father of her son and ex-partner, and his indifference about her situation when they met after a long time).

- (57) Todos eran pobres. Pobres, **podían** robarnos, pero no,  
all they.were poor poor they.can rob.us but no  
considerados eran.  
considerate they.were  
'Everybody was poor. Poor, they could have robbed us, but no, they were considerate.' (Dankel & Pagel 2012: 47)  
PS: 'Todos eran pobres. Pobres, podrían habernos robado, pero no, tuvieron consideración.'
- (58) O sea el me ha dicho siempre los colores **tal vez** estoy equivocado.  
it be the me has said always the colors maybe I.am wrong  
'I mean, he said it clearly, about colors I could be wrong.' (CQE)  
PS: 'Quiero decir, lo dijo claramente, sobre los colores yo podría estar equivocado.'
- (59) **Capaz** me dé con el palo así.  
where you you.go able me give with the stick so  
'She would possibly hit me with the stick or something.' (CQE)  
PS: 'Probablemente, me iba a pegar con un palo o algo (así)'.

---

*rantikuwax*. In some of those cases, *atiy* 'can' converges with the potential Quechua *-man*. For example, the continuity of a dialogue is in danger: *atinman p'akikuyta* 'it would break out' (CQE). If we consider uses of *poder* with these notions as an effect of the contact with Quechua, and, conversely, the use of new periphrastic modal constructions with *poder* in Quechua, we can observe a relationship of round-trip effects in the processes of influx that we consider as a boomerang process.

<sup>10</sup>The use of *capaz* with modal value has been reported in other varieties in America, too (Narrog 2012, Grández Ávila 2010, Yelin & Czerwionka 2017, among others).

<sup>11</sup>Again, a detailed account of conditional constructions is outside the scope of this chapter. Nonetheless, note that the patterns in CA do not follow the combinatorial grammatical restrictions that are obligatory in Peninsular Spanish, and, furthermore, orient to patterns which closely resemble Quechua syntax: non-finite subordinate structures (e.g. *sabiendo esto me hubiese ido*) as a conditional or constructions with auxiliaries as matrix clause (parallel to the use of the Quechua auxiliary in example (56)) in exhortative contexts (e.g. *Era que vengas más temprano*) as a contrafactual, and which go beyond a discursive or stylistic phenomenon.

- (60) ¿Tú lo mantienes? ¿Estás con trabajo? ¿Tienes plata? Por  
you him support you.are with work you.have money for  
lo menos por ahí hubiera empezado.  
at.least for there would.have started  
'Who? Do you support him? Do you have a job? Do you have money? At  
least so he would have started.' (CQE)  
PS: '¿Tú lo mantienes? ¿Tienes un trabajo? ¿Tienes dinero? Al menos así  
(él) habría empezado.'

CA speakers reflect the meanings which are possible with dedicated Aymara and Quechua suffixes by using strategies entirely (or nearly) absent in Peninsular Spanish. This is a first step on the linguistic surface for establishing a semantic copy (Johanson 2008) or for Slobin's (2016) concept of "thinking for speaking" as the cognitive impulse for contact-induced change.

## 5.5 Outstanding issues

We have only scratched the surface of the contact-induced processes lying beneath the parallels we described. In this section we mention some noteworthy observations which future studies may take into account: certainty markers<sup>12</sup>, the expression of the future, the progressive and the habitual.

Closely intertwined with evidential marking, the languages under investigation also have many ways to express epistemic modality, which is hardly researched. Among these, the following certainty-markers are the most relevant as they overlap with inferentials/conjecturals, especially in Quechua and CA. The Quechua suffix *-sina* expresses uncertain conjecture, for example, when speculating about the age of a family member, as in (61). In CA, as is typical for situations of incipient grammaticalization, this is reflected in the use of *creo que* ('I believe that'). This is illustrated in (62), which comes from the testimony of a confrontation during a protest. In this example, *creo que* is neither located at the beginning of enunciative sequences as in its function as a discourse maker, nor is it used with an explicit subject, as in manifestations of points of view.

- (61) *cincuenta y tres-ni-yux-sina*  
fifty and three-EUF-COM-DUB  
'I guess she's fifty three.'

<sup>12</sup>We use "certainty markers" as a cover term for these suffixes, because they all reflect the speakers stance on the veracity of the state of affairs. However, this is a working definition, as more research is necessary to define the precise functions of these suffixes.

CA: ‘Creo que tiene cincuenta y tres.’ (CQE)

PS: ‘Creo / supongo que tiene cincuenta y tres años.’

- (62) Pero ya ni bien estábamos defendiendo y todo, regresé  
 but already or well we.were defending and all I.returned  
 de nuevo hacia adelante para poder ayudar a mis compañeros y  
 again toward ahead for be.able to.help to my colleagues and  
 creo que me vieron. Todo eso directo un tiro ¡pum!  
 I.believe that me they.saw all this direct a shot bang  
 ‘As soon as we were defending ourselves or so, I went ahead again to help  
 my colleagues and, I guess they saw me – suddenly a shot: bang!’  
 Defensoría del Pueblo (2020: 97)  
 PS: ‘Cuando nos estábamos defendiendo, yo volví de nuevo para ayudar a  
 mis compañeros/as y creo que me vieron – de repente un disparo: ¡pam!'

Furthermore, there are question-like conjectures with the Quechua suffix *-chus*. However, function-wise, speakers use *-chus* to express a doubt, as in (63). The suffix *-chus* is often reflected in CA with the postponed tag question *qué será*, which is frequently attested in our CA data.

- (63) imayna-chus á mana yacha-ni-chu pi-chus ka-n-pis  
 how-CONJ MOD.INT no know-1-NEG who-CONJ be-3-CONC  
 ‘No idea, I don’t know, I don’t know even who she is.’  
 CA: ‘Cómo será pues, no sé quién será.’ (CQE)  
 PS: “Ni idea, no lo sé, ni siquiera sé quién es ella.”

Finally, there are clear systematic parallels in the use of the Quechua suffix *-puni*, the Aymara suffix *-puni* and the use of CA adverbial *siempre* ‘always’, which are used as validators (Cerrón-Palomino 2008, Plaza Martínez 2009, Coler 2014).

- (64) Uk-jama-puni-tayna-w  
 that-CP-EMP-3.SUBJ.3.OBJ.NEXP.PAST-DECL  
 ‘That’s how it went.’ (Velasco Rojas 2015: 148)  
 CA: ‘Así siempre había sido.’  
 PS: ‘Así fue.’
- (65) Anchay, mama-y wasi-n-puni  
 that mother-1 house-3-CERT  
 ‘That house really belongs to my mother.’ (Peralta Zurita 2006: 170)  
 CA: ‘Esa casa de mi mamá siempre es.’  
 PS: ‘Esa casa realmente es de mi madre.’

Certainty markers in all three varieties often co-occur with evidential inferentials. This leads to epistemically highly complex phrases, especially in CA where one finds sentences like *Alicia seguramente le insultaría, qué sería, no sé* ‘Alicia must have insulted her, how would it be, I don’t know’. Observe the three different epistemic markers (in bold) interacting within one sentence. Speakers recombine a variety of Spanish strategies to get as close as possible to their culturally shaped communicative functions, expressed with suffixes in Aymara and Quechua. A further discussion of these nuances would be interesting, even if it falls outside the scope of this chapter. More research is required.

As regards the future tense, in CA the analytic future is used as a tense, whereas the synthetic future imparts a modal function (asking for permission/confirmation): *le he dicho iremos a comer jarwi uchu* ‘Let’s go to eat a jarwi uchu’ (where *iremos* ‘we should go’ is used in CA instead of *vamos* ‘we will go’, as in PS). This is also what one finds in the indigenous languages discussed where the distinction between the interpretation of the future tenses may be made contextually and/or by adding temporal adverbs. Observe the synthetic future form in the CA translation of (66) in Quechua.

- (66) Elvira t’anta-ta ruwa-pu-sqayki-chu? Ichas qam  
Elvira bread-ACC make-BEN-1.SUBJ.2OBJ.FUT-INT perhaps you  
extraña-yku-chka-nki.  
miss-DIR-PROG-2SG  
'Elvira, do you want me to bake bread for you? Perhaps you're missing it  
(the bread).'  
'Elvira te lo haré **pancito**? Tal vez vos estás extrañando (el pan).'  
(Peralta Zurita 2006: 204)  
PS: 'Elvira, ¿quieres que te hornee pan? Quizás lo echas / eches de menos  
(el pan).'

At first glance, the progressive aspect has little to do with mood and modality. According to Quechua grammars, verbal morphology obligatorily marks the immediate as opposed to non-immediate character of an event (Adelaar & Muysken 2004, Zariquey & Córdova 2008, Cerrón-Palomino 2008, Cole 1982: 231). This is done with the suffix *-sa*. Consider the parallel ways in which Quechua and CA use progressive constructions in the following examples. In both, a speaker makes reference to an ongoing activity. This is achieved in (67) with the progressive morpheme and in (68) with a periphrastic gerund construction.

- (67) chara-ta qarpa-x ri-sa-ni  
 crop-ACC water-SUB go-PROG-1SG  
 ‘Right now I’m going to water the crop.’ (Cayetano Choque 2003: 267)  
 CA: ‘Estoy yendo a regar el terreno.’  
 PS: ‘Ahora mismo voy a regar la cosecha.’
- (68) y hacen de todo, como está viendo pantalones, camisas, de todo  
 and they.make of all as are seeing pants shirts of all  
 ‘And they make everything, as you see, pants, shirts, everything.’ (CQE)  
 PS: ‘Y hacen de todo, como ves, pantalones, camisas, de todo.’

Progressives go beyond an aspectual function. They show “the speaker’s attitude and perspective of the situation; and, in so doing, [convey] her epistemic stance at a particular moment in the context of utterance” (Wright 1995: 157). The experiential character of the progressive therefore expresses a modal notion. A further modal notion that has been observed is a volitional reading to express deliberate intentions that can also be realized in a non-immediate future, as in a phrase like *Estoy saliendo de la casa. Llego en un rato* ‘I am [already] leaving the house. I’ll arrive in a bit.’ Upon closer examination of the CA and Quechua progressive constructions for current events, it becomes clear that their development is a diachronic path, where both Quechua and Spanish are dynamically intertwined with patterns being exchanged back and forth in both directions at various points in time (Soto Rodríguez 2016). Future research should also take into account the progressive in Aymara.

Habitual actions or states in Aymara and Quechua are expressed with agentive constructions. In Aymara, they are formed with the agentive nominalizer *-iri*. A verb nominalized with the agentive connotes a habitual action, in the sense that *um-iri* ‘drinker’ can refer to one who drinks habitually. Consequently, *jupa-x um-iri-w* ‘she is a drinker’ (of alcohol) may also be translated as ‘she often drinks’ (*ella sabe tomar*) likewise *jupa-x chur-ir-itu-w* ‘she usually gives it to me’ (*ella me sabe dar*) can be understood roughly as ‘she was a giver to me’. This mechanism works similarly in Quechua where the agentive suffix *-x* can be used to nominalize verbs such as *ruwa* ‘do’ to become *ruwa-x*, ‘the one who makes/the one who does habitually’.

Whereas habitual marking is central to Quechua and Aymara the habitual in PS is not a central systematic category and is formed with a dedicated verb *soler*. CA speakers did not promote the *soler* construction but seemingly chose another mode of marking the habitual based on *saber* ‘to know’ + infinitive: *las*

*toallas saben estar colgadas en allá* ‘the towels usually are hung there’. This strategy already existed rather peripherically in old Spanish spoken varieties (Pfänder 2009), but is a cross-linguistically frequent grammaticalization path (Kuteva et al. 2019) because of its pragmatically and metonymically/metaphorically plausible developmental potential.<sup>13</sup>

## 6 Discussion

We selected four case studies to illustrate how mood and modality in Aymara, Quechua and CA converge. For more than a century, Aymara and Quechua were minority languages and were hardly integrated. Nonetheless, their conceptual notions highly influenced CA. The three languages developed many similarities and show ongoing exchange of notions and patterns.

In the analysis of the tense/evidentiality distinction, CA is shown to have developed on the structural and the notional levels. CA speakers do not just map the modal-evidential (non-)experienced past distinction, rooted in Quechua and Aymara morphology, onto the Spanish perfect and pluperfect temporal morphology; they also transfer the mirative interpretation, which is associated with the Aymaran and Quechuan non-experienced past tense, to the newly established CA counterpart (the pluperfect morphology).

In the discussion of hearsay and quotative marking, we showed how two culturally highly-relevant practices which became grammaticalized in both Aymara and Quechua, lead to the formation of grammatical markers for the same category in CA, formerly not systematically present in other varieties of Spanish. We also pointed to a reverse influence of the new CA marker onto some varieties of

---

<sup>13</sup>In Aymara and Quechua we also find a development in the opposite direction. The modal verb ‘want’ (*munay* in Quechua, *muna* in Aymara) is used aspectually for the description of incipient natural events. This use is reflected in CA, where *querer* can be used in a comparable way.

- (i) Iqra-cha-ku-y-ta-ña                    muna-chka-sqa.  
wing-CAUS-REFL-INF-ACC-DISC want-PROG-NEXP.PAST  
‘Wings were about to grow up [from a larva].’ (Lit. ‘it wanted to grow up wings’)  
(Mamani Lopez 2018: 75)
  
- (ii) Calamina quiere levantar, quiere llevar.  
metal.roof want to.lift.up want to.take.off  
‘It [the strong wind] is about to lift up metal roofs, it is about to take them up’  
(Lit. It wants to lift up metal roofs, it wants to take off the roofs), (CQE).

Aymara and Quechua, which developed a counterpart modeled on the CA strategy.<sup>14</sup> This seems counter-intuitive at first sight, however it is an even stronger indication that speakers primarily orient to expressing culturally relevant notions and use diverse linguistic resources to do so. The main aim of speakers is communicational success and processing efficiency (“thinking for speaking” as modelled in Slobin 2016). This leads to a step-by-step structural convergence, as we can observe here, where the distinction between different systems and their structural symmetries yield to the emergence of one joint system.

The description of inference/conjecture does not show a strict one-to one mapping of grammatical categories between the investigated languages, as Aymara shows a fine grained distinction for the source of the deduced knowledge, which is neither present in Quechua, nor transferred to CA. This might raise questions of comparability regarding the extent to which it is possible to support assertions of parallel structures and convergence. Yet, upon closer examination, the context of use for inferentials/conjecturals in all three varieties shows systematic parallelism within the larger notional category. The dynamic situation in CA, with more than just one strategy to match the Aymara and Quechua morphemes, confirms the premises set out in our theoretical framework. Speakers engage in creative language use to express their communicative needs (Babel & Pfänder 2014) and establish competing patterns to carry over their semantic copy (in this case inference/conjecture) at the stage of incipient grammaticalization. Only in another step do they accommodate their “thinking for speaking” (Slobin 2016) on the linguistic surface: the accumulation of the most successful communicative routine carves out a new potential grammatical maker. Or in terms of Johanson’s (2008) observation: the semantic functions of copies typically have not reached the same stage of grammaticalization as their models.

In our description of potential and counterfactual marking, the mapping between languages again is not neat and uniform. Both Aymara and Quechua use dedicated suffixes for this category. However, whereas Aymara uses a paradigm with past and present person-marked suffixes, Quechua has just one suffix that can be attached to the inflected verb form. And CA speakers try to reflect these with newly-formed regular constructional patterns that are hardly (or not at all) present in Peninsular Spanish and may compete with each other at this stage of incipient grammaticalization. This also affects how conditional constructions are formed. This is an interesting topic for future research.

---

<sup>14</sup>This reverse influence is not exclusive to the reportative and hearsay makers, as we pointed out in our analysis. Accordingly, we want to re-emphasize that contact induced change, and in this vein, convergence, is not a one way street. It works simultaneously in both directions.

In comparison, we find a quite homogeneous category mapping for the analysis of tense/evidentiality (Section 5.1) and hearsay and quotatives (Section 5.2) relative to the more dynamic picture for inference/conjecture (Section 5.3) and potential and counterfactual forms (Section 5.4). This dynamicity also holds for certainty markers, addressed in outstanding issues (Section 5.5). This may be due to an early grammaticalization stage in this dynamic contact situation. But we may add that the fact that these categories are closely intertwined for being used in highly nuanced stance-taking regarding the certainty of a state of affairs. The speakers' needs for a gradual tuning in communicative interaction therefore is a complex issue where the accumulation of the most successful communicative routines in CA are less straight-forward. This type of graduality is less relevant for the (non-)experienced past distinction from Section 5.1 and hearsay and quotative from Section 5.2, as their functional opposition is clearly binary (either one hears/experiences something or not) and they are nuanced in discourse in a different way. The emergence of new grammatical strategies for these categories might therefore be quicker.<sup>15</sup> This needs to be confirmed by a broader and more thorough analysis.

Finally, we showed that a more detailed elaboration on the uses of the future tenses, progressive and habitual constructions in future studies, as described in Section 5.5, could contribute to more complete understanding.

In sum, our approach to extend our perspective beyond a narrow grammatical focus lead us to a complex and highly diverse set of linguistic structures relevant for mood and modality in Aymara, Quechua and CA. This diversity and variation do not invalidate our claim to speak of mood and modality as a joint systematic concept in all three languages. To the contrary, it confirms the crucial role of mood and modality in Aymara and Quechua and underscores how speakers repurpose CA to fit their needs.

## 7 Conclusion and outlook

We compared Aymara, Quechua and CA mood and modality. We wanted to show that we find contact induced change in these three languages and that the ongoing long-term contact situation profoundly influences the language systems in the long run and leads, step-by-step and in different speeds for different domains,

---

<sup>15</sup>We are aware of the multicausality of grammaticalization processes and do not want to exclude other reasons for these differences in the stage of development. Yet this may be a relevant factor that has not received much attention.

to a conceptual conflation into a joint system of categories beneath the structural surface.

We compared mood and modality because it plays a key role in Andean discourse. The modal dimension can serve as a good example for how categories that we find in Aymara and Quechua, despite the typological distance and without direct grammatical borrowing, get overwhelmingly reflected in the majority language CA. This also holds in the speech of those who do not speak either indigenous language, the minority status of those two languages notwithstanding. In this regard, we hope to have illustrated that CA can be seen as a diaspora language, because this variety heavily retains the cultural identity of Andean Spanish speakers although the Standard Spanish norm politically clearly dominates.

We based our analyses on the culturally-shaped communicative routines of the speakers and their ingenuity to operationalize the potential of the available linguistic forms in each of the three codes to convey their semantic and pragmatic needs. For this, we took a procedural perspective on mood and modality, an approach that proved successful and brought to light a level of convergence that goes far beyond superficial contact influence and which could be missed when referring to static grammatical descriptions only.

There remains much work to be done to better understand these processes. Contact induced change is complex and has multiple causes. One of these aspects, which seems crucial for the success of speakers' creative language use in operationalizing a different language for their culturally shaped communicative routines, is decreased standardization as a sociopolitical factor that promotes it. The lack of consequent efforts to establish school programs for a systematic second language education in Spanish, especially in the rural areas, where pre-school children grow up monolingual in Quechua or Aymara leaves the learning of Spanish as a practice of assimilation by submersion and improvisation (see Soto Rodríguez 2006 for Bolivia). The same goes for Quechua and Aymara which are still not adequately represented in primary school and beyond. From the perspective of contact research this is an important factor, which does not receive sufficient attention, also in other contact situations, as language education might contribute considerably to the openness of a language system for contact induced change<sup>16</sup> and therefore might explain how the minoritized languages Aymara and Quechua remain highly influential.

---

<sup>16</sup>The question of language education and standardization are of course also highly relevant from a sociolinguistic perspective, regarding language policy, language preservation, and equal opportunities in access to (higher) education for rural populations.

## References

- Adelaar, Willem F. H. 2012. Modeling convergence: Towards a reconstruction of the history of Quechuan–Aymaran interaction. *Lingua* 122(5). 461–469.
- Adelaar, Willem F. H. 2013. A Quechuan mirative? In Alexandra Y. Aikhenvald & Anne Storch (eds.), *Perception and cognition in language and culture* (Brill's Studies in Language, Cognition and Culture), chap. 4, 95–109. Leiden: Brill.
- Adelaar, Willem F. H. & Pieter Muysken. 2004. *The languages of the Andes*. Cambridge: Cambridge University Press.
- Albó, Xavier. 1970. *Social constraints on Cochabamba Quechua* (Latin American Studies Program, Dissertation Series). Ithaca, New York: Cornell University Press.
- Andrade, Luis. 2007. *Usos de dice en el castellano andino: Estrategias evidenciales y narrativas en contacto con el quechua*. Pontificia Universidad Católica del Perú. (MA thesis).
- Andrade Ciudad, Luis. 2005. *Aguas turbias, aguas cristalitas: El mundo de los sueños en los Andes surcentrales*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.
- Andrade Ciudad, Luis. 2020. Evidentiality and epistemic modality in the Andean Spanish verb. In Sandro Sessarego, Juan J. Colomina-Almiñana & Adrián Rodríguez-Riccelli (eds.), *Variation and evolution. Aspects of language contact and contrast across the Spanish-speaking world*, 75–103. Amsterdam: John Benjamins Publishing.
- Babel, Anna & Stefan Pfänder. 2014. Doing copying: Why typology doesn't matter to language speakers. In Juliane Besters-Dilger, Cynthia Dermarkar, Stefan Pfänder & Achim Rabus (eds.), *Congruence in contact-induced language change*, 239–257. Berlin: de Gruyter.
- Babel, Anna M. 2009. Dizque, evidentiality, and stance in Valley Spanish. *Language in Society* 38(4). 487–511.
- Bosque, Ignacio. 2012. Mood. indicative vs. subjunctive. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *Handbook of Hispanic linguistics*, 373–394. Oxford: Blackwell.
- Briggs, Lucy T. 1976. *Dialectal variation in Aymara*. University of Florida: Center for Latin American Studies.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Cayetano Choque, Miriam. 2003. *La causalidad en la lengua quechua*. La Paz: Artes Gráficas Latinas.

- Cerrón-Palomino, Rodolfo. 1987. *Lingüística quechua*. Cuzco: Centro de Estudios Rurales Andinos Bartolome de las Casas.
- Cerrón-Palomino, Rodolfo. 2000. El origen centroandino del aimara. *Boletín de Arqueología PUCP* (4). 131–142.
- Cerrón-Palomino, Rodolfo. 2008. *Quechumara: Estructuras paralelas del quechua y del aimara*. La Paz: Plural editores.
- Chang, Lidia. 2018. “Dizque eran compadres el zorro y el quirquincho”: El verbo decir como marcador reportativo y metadiscursivo en el español del noroeste argentino. *Lexis* 42(1). 155–176.
- Cole, Peter. 1982. *Imbabura Quechua* (Lingua Descriptive Studies 5). Amsterdam: North-Holland.
- Coler, Matt. 2014. *A grammar of Muylaq' Aymara: Aymara as spoken in Southern Peru*. Leiden: Brill.
- Coler, Matt. 2015. Aymara inflection. In Matthew Baerman (ed.), *The Oxford handbook of inflection*. Oxford: Oxford University Press.
- Crevels, Emily I. & Pieter C. Muysken. 2009. *Lenguas de Bolivia*, vol. 1 *ámbito andino*. La Paz: Plural Editores.
- Dankel, Philipp. 2015. *Strategien unter der Oberfläche. Die Emergenz von Evidentialität im Sprachkontakt Spanisch–Quechua*. Freiburg / Berlin / Wien: Rombach.
- Dankel, Philipp & Steve Pagel (eds.). 2012. *Más que una heroína* (New Ideas in Human Interaction /Stories). Freiburg: HPSL (coord. Carlos Coello & Stefan Pfänder).
- Dankel, Philipp & Mario Soto Rodríguez. 2012. Convergencias en el área andina: La testimonialidad y la marcación de la evidencialidad en el español andino y en el quechua. *Neue Romania* 41. 89–120.
- Defensoría del Pueblo. 2020. *Informe defensorial: Crisis de estado, violación de los derechos humanos en Bolivia octubre–diciembre 2019*. La Paz.
- DeLancey, Scott. 2001. The mirative and evidentiality. *Journal of Pragmatics* 33(3). 369–382.
- Drinka, Bridget & Whitney Chappell. 2021. Spanish socio-historical linguistics. *Spanish Socio-Historical Linguistics*. 1–241.
- Escobar, Anna María. 1997. From time to modality in Spanish in contact with Quechua. *Hispanic Linguistics* 9(1). 64–99.
- Escobar, Anna María. 2011. Spanish in contact with Quechua. In Manuel Díaz-Campos (ed.), *The handbook of hispanic sociolinguistics*, 323–352. Malden, Mass. u.a.: Wiley-Blackwell.

- Escobar, Anna María & Claudia Crespo del Río. 2021. La gramaticalización de la subjetividad en el español andino. El pretérito perfecto compuesto con valor evidencial. In Luis Andrade Ciudad & Sandro Sessarego (eds.), *Los Castellanos del Perú - historia, variación y contacto lingüístico*, 156–205. Abingdon/ New York: Routledge.
- Faller, Martina. 2002. *Semantics and pragmatics of evidentials in cuzco quechua*. Stanford University. (Doctoral dissertation).
- Feke, Marilyn Suzanne. 2004. *Quechua to Spanish cross-linguistic influence among Cuzco Quechua-Spanish bilinguals: the case of epistemology*. University of Pittsburgh. (Doctoral dissertation).
- García Tesoro, Ana Isabel. 2015. Valores emergentes del pretérito pluscuamperfecto en el español andino hablado en chinchoro (cuzco). *Boletín de filología* 50(2). 51–75.
- Grández Ávila, Magaly. 2010. *A functional approach to the subjectification of facultative meaning: The case of capaz in American Spanish*. University of Amsterdam. (MA thesis).
- Haboud, Marleen. 2008. Mujeres andinas migrantes: Creando fronteras, vulnerabilizando límites. In Isabel García Parejo, Azucena Palacios & Isabel Rodríguez (eds.), *Fronteras exteriores e interiores: Indigenismo, género e identidad*, 92–104. München: LINCOM.
- Haimovich, Gregory. 2016. From synthetic to analytic modality: Contact induced evolution of potential mood in Southern Quechua (Conference paper). In *The 46th Poznań linguistic meeting (PLM2016)*. Poznań: Adam Mickiewicz University.
- Hardman, Martha J. 1973. *Aymara grammatical sketch*. Washington, DC: Institute for International Studies.
- Hardman, Martha J. 1982. The mutual influence of Spanish and the Andean languages. *Word* 33(1-2). 143–157.
- Hardman, Martha J. 2001. *Aymara* (Lincom Series in Native American Linguistics 35). Munich: Lincom.
- Howard, Rosaleen. 2007. *Por los linderos de la lengua* (Lengua y Sociedad). Lima: IEP, IFEA, PUCP.
- Howard, Rosaleen. 2018. Evidence and evidentiality in Quechua narrative discourse. In Alexandra Y. Aikhenvald (ed.), *The Oxford handbook of evidentiality*, chap. 11, 222–242. Oxford: Oxford University Press.
- Jarvis, Scott & Aneta Pavlenko. 2008. *Crosslinguistic influence in language and cognition*. New York: Routledge.

- Johanson, Lars. 2008. Language contact and contact languages. In Peter Siemund & Noemi Kintana (eds.), chap. Remodeling grammar. Copying, conventionalization, grammaticalization, 61–79. Amsterdam: John Benjamins Publishing.
- Kalt, Susan E. 2021. Acquisition, loss and innovation in Chuquisaca Quechua – What happened to evidential marking? *Languages* 6(2).
- Klee, Carol A. & Alicia M. Ocampo. 1995. The expression of past reference in Spanish narratives of Spanish-Quechua bilingual speakers. In Carmen Silva-Corvalán (ed.), *Spanish in four continents: Studies in language contact and bilingualism*, 52–70. Washington D.C.: Georgetown University Press.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. Cambridge University Press.
- Lee, Tae Yoon. 1997. *Morfosintaxis amerindias en el español americano. Desde la perspectiva del quechua*. Madrid: Ediciones del Orto.
- Lehmann, Christian. 1982. Thoughts on grammaticalization. A programmatic sketch. *AKUP: Arbeiten des Kölner Universalien-Projekts Köln* 1(48). 1–186.
- Lipski, John. 2010. Spanish and Portuguese in contact. In Raymond Hickey (ed.), *The handbook of language contact*, 550–580. Malden, MA / Oxford: Wiley-Blackwell.
- Mamani Lopez, Marcelino. 2018. *Kunanpi Qhichwa: El quechua actual*. Cochabamba: Universidad Mayor de San Simón. (MA thesis).
- Manley, Marilyn S. 2015. Multidimensional markers of evidential, epistemic and mirative stance in Cuzco Quechua. In Marilyn S. Manley & Antje Muntendam (eds.), *Quechua expressions of stance and deixis*, 145–207. Leiden: Brill.
- Mannheim, Bruce. 1991. *The language of the Inka since the European invasion*. University of Texas Press.
- Martínez Vera, Gabriel A. 2020. *On the semantics of evidentials in Southern Aymara*. University of Connecticut. (Doctoral dissertation).
- Mendoza, José G. 1991. *El castellano hablado en La Paz: Sintaxis divergente*. Tall. Gráf. de la Facultad de Humanidades y Ciencias de la Educación.
- Merma-Molina, Gladys. 2007. *Contacto lingüístico entre el español y el quechua: Un enfoque cognitivo-pragmático de las transferencias morfosintácticas en el español andino peruano*. Universidad de Alicante. (Doctoral dissertation).
- Narrog, Heiko. 2012. *Modality, subjectivity, and semantic change: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Olbertz, Hella. 2005. Dizque en el español andino ecuatoriano: Conservador e innovador. In Hella Olbertz & Pieter Muysken (eds.), *Encuentros y conflictos: Bilingüismo y contacto de lenguas en el mundo andino*, 77–94. Frankfurt a.M. / Madrid: Vervuert / Iberoamericana.

- Palacios, Azucena & Stefan Pfänder. 2018. El pretérito pluscuamperfecto con valor (ad)mirativo. *Revista Internacional de Lingüística Iberoamericana (RILI). Sección temática dedicada a „Procesos de gramaticalización en situaciones de contacto“*, coordinada por Azucena Palacios y Stefan Pfänder Volumen XVI, N° 2(32). 119–134.
- Palmer, Frank R. 2001. *Mood and modality*. Cambridge: Cambridge University Press.
- Peralta Zurita, Elvira. 2006. *Descripción morfológica de la palabra quechua: Un estudio basado en el quechua de Yambata, norte de Potosí*. Universidad Mayor de San Andrés. (Doctoral dissertation).
- Pfänder, Stefan. 2009. *Gramática Mestiza - Presencia del quechua en el castellano boliviano*. Vol. II. (in collaboration with Juan A. Ennis and Mario Soto Rodríguez and España Villegas Pinto). La Paz (Bolivia): IBLEL.
- Pfänder, Stefan & Azucena Palacios. 2013. Evidencialidad y validación en los pretéritos del español andino ecuatoriano. *Círculo de Lingüística Aplicada a la Comunicación* (54). 65–99.
- Plaza Martínez, Pedro. 2009. Quechua. In Mily Crevels & Pieter Muysken (eds.), *Lenguas de Bolivia tomo I: Ámbito andino*, 215–284. La Paz: Plural Editores.
- Quartararo, Geraldine. 2017. *Evidencialidad indirecta en aimara y en el español de La Paz: Un estudio semántico-pragmático de textos orales*. Department of Romance Studies & Classics, Stockholm University. (Doctoral dissertation).
- Quelca Huanca, Heriberto. 2006. *Transferencias morfosintácticas del aymara al castellano en estudiantes de la ciudad de El Alto*. Universidad Mayor de San Andres. (Doctoral dissertation).
- Real Academia Española. 2009. *Nueva gramática de la lengua española*. Vol. 2. Madrid: Espasa Calpe.
- Sánchez, Liliana. 2004. Functional convergence in the tense, evidentiality and aspectual systems of quechua Spanish bilinguals. *Bilingualism: Language and Cognition* 7(2). 147–162.
- Schumacher de Peña, Gertrud. 1980. El pasado en el español andino de Puno / Perú. In Hans Dieter Bork, Artur Greive & Dieter Woll (eds.), *Romanica Europaea et Americana: Festschrift für Harri Meier*, 553–558. Bonn: Bouvier-Herbert Grundmann.
- Silva-Corvalán, Carmen. 1994. *Language contact and change*. Oxford: Clarendon Press.
- Slobin, Dan I. 2016. Thinking for speaking and the construction of evidentiality in language contact. In Mine Güven, Didar Akar, Balkız Öztürk & Meltem Kelepir (eds.), *Exploring the turkish linguistic landscape: essays in honor of eser*

- erguvanlı-taylan, 105–120. Amsterdam & Philadelphia: John Benjamins Publishing.
- Soto Rodríguez, Mario. 2002. *La interferencia del castellano en el quechua radial del norte de potosí a través del uso de los conectores*. Universidad Mayor de San Andrés. (MA thesis).
- Soto Rodríguez, Mario. 2006. *Diagnóstico en unidades educativas que atienden a niños menores de 6 años hablantes de una lengua originaria en la región quechua en nivel inicial*. Tech. rep. 2. Bolivia: Ministerio de Educación de Bolivia Viceministerio de escolarizada y alternativa, Dirección de desarrollo curricular.
- Soto Rodríguez, Mario. 2016. Nuevas ámbitos, nuevas construcciones en el quechua de Bolivia: Analitización verbal con gerundio en medios de prensa (Conference paper). In Ohio State University. Columbus, Ohio, USA: 1st Annual Symposium on Indigenous Languages, Cultures of Latin America (ILCLA) in conjunction with the 3rd Annual Symposium on Teaching & Learning Indigenous Languages of Latin America (STLILLA).
- Soto Rodríguez, Mario & Philipp Dankel. in print. Contacto, cambio multifactorial y dinámico en la zona andina - epistemicidad encubierta en construcciones con decir. In Álvaro Ezcurra & Carlos Garatea Grau (eds.), *Español de América: Historia y contactos*. Frankfurt am Main: Peter Lang.
- Stratford, Dale. 1991. Tense in altiplano Spanish. In Carol A. Klee & Ramos-García Luis A. (eds.), *Sociolinguistics of the Spanish-speaking world*, 163–181. Tempe, Arizona: Bilingual Press/Editorial Bilingüe.
- Taylor, Gerald. 2006. Ritos y tradiciones de huarochirí: Edición bilingüe / Quechua-Castellano. In Lima: Instituto de Estudios Peruanos.
- Urton, Gary & Primitivo Nina Llanos. 1997. *The social life of numbers: A Quechua ontology of numbers and philosophy of arithmetic*. Austin: University of Texas Press.
- Velasco Rojas, Pedro. 2015. *Aymaraña qhalt’itaki, mi aymara para el mañana: Estudio dialectal-sincrónico del aymara de los ayllus Laymi-Puraka y Jukumani del norte de Potosí*. Universidad Mayor de San Andrés. (MA thesis).
- Wright, Susan. 1995. Subjetivity and experiential syntax. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 151–172. Cambridge: Cambridge University Press.
- Yelin, Boris & Lori Czerwionka. 2017. Capaz and quizás in Argentine Spanish: Epistemic adverbs with mood variability. *Hispania* 100(1). 30–46.
- Zariquiey, Roberto & Gavina Córdova. 2008. *Qayna, kunan, paqarin una introducción práctica al quechua chanca*. Lima: Estudios Generales Letras - Pontificia Universidad Católica del Perú.



# Chapter 5

## What is the role of the addressee in speakers' production? Examples from the Griko- and Greko-speaking communities

Manuela Pellegrino<sup>a</sup> & Maria Olimpia Squillaci<sup>b</sup>

<sup>a</sup>CHS, Harvard University <sup>b</sup>University of Naples "L' Orientale"

In the literature on minority languages, language use and variation are commonly analyzed with reference to the speaker. In this contribution we instead focus on the addressees and how they can impact the speaker's language use and influence speech production. We will discuss these issues in relation to Griko and Greko, two endangered Italo-Greek varieties spoken in the south of Italy, Salento (Puglia) and Calabria respectively.

### 1 Introduction

Over the past decades, there has been an increase in Greko and Griko written production, involving both elderly mother-tongue speakers and especially so-called "semi-speakers", and local language experts (cf. Martino 2009 and Pellegrino 2016b). By contrast, the spoken use of both varieties has long been decreasing, progressively losing domains. At present, on an everyday basis, locals mostly communicate in the local Romance varieties – Salentine (Puglia) and Southern Calabrian (Calabria) – or in Italian, and they use Griko and Greko in limited contexts and with an increasingly smaller number of people. This applies in particular to Calabria (Area Grecanica), where at the community level Greko is used significantly less than Griko in Salento (Grecìa Salentina). The dynamics leading



to this decrease are multiple and complex, and concern the broader processes of language shift and abandonment.

In this paper, we provide a preliminary analysis of data in relation to speaker-addressee dynamics and how these affect language use and may potentially lead to what we refer to as “temporary variation”. In particular, we highlight the role of the addressee in such dynamics and demonstrate how the addressee’s linguistic competence, age, and shared linguistic repertoire with the speaker may lead to style-shift in speakers’ production; this, in turn, contributes to the emergence of puristic attitudes and even inhibits the use of the varieties themselves. This will also shed light on key differences between the two communities with regard to reception of language maintenance and/or revitalisation programs.

This chapter builds on the authors’ joint project, “Investigating the future of the Greek linguistic minorities of Southern Italy”. This project provides a comparative examination of the responses to language maintenance and revitalisation initiatives among the Griko- and Greko-speaking communities. It was part of Sustaining Minoritized Languages in Europe (SMiLE), an interdisciplinary research program developed by the Center for Folklife and Cultural Heritage (Smithsonian Institution). One goal was to produce ethnographic studies of six communities in Europe and analyze how language-related initiatives build on motivational responses to social, cultural, political, and economic factors.

Our research rested at the intersection between social and linguistic domains, using qualitative methods such as participant observation and ethnography of speaking. Between January 2018 and June 2019, we conducted semi-structured interviews with leaders of cultural associations, elderly speakers, new and non-speakers, young people, academics, and representatives of institutions, totaling over 70 individuals.<sup>1</sup> We also participated in and observed 40 local cultural activities, speaking with those involved whenever possible. These cultural activities included music festivals, seminars, poetry competitions, and school projects. This allowed us to compare the current maintenance and revitalisation activities being implemented in both areas, their reception by and their impact on the communities, the target audience, and the degree of involvement of the young people in such activities. Our analysis is also enriched by the previous anthropological and linguistic research that we independently carried out during our doctoral work, as well as by our personal connection to and long-standing engagement with the Griko- and Greko-speaking communities from which we authors hail, respectively.

The article is structured as follows. In Section 2 we provide some historical background on the Griko and Greko varieties. In Section 4 we move on to discuss

---

<sup>1</sup>All interviewees’ names in this contribution are pseudonyms.

language use; using some key examples we will argue that the age and minority-language competence of the addressee play a crucial role in inhibiting or favoring a speaker's use of Griko and Greko in conversational settings. This analysis will bring to light some of the main differences between the two communities, which are the result of past and current dynamics unique to each. In Section 5 we focus specifically on the language competence of the addressee in (standard and regional) Greek,<sup>2</sup> and show how this may bring about instances of "temporary variation" in the speech production of some Griko and Greko speakers. As we will demonstrate, such an influence, although limited to specific contexts, has significant repercussions for inter- and intra-community dynamics which transcend speech production itself. In this respect, we also draw attention to researchers' implicit or explicit attitudes towards interference and variation, which may promote prescriptivist and puristic values when dealing with minority languages.

## 2 Background information

There has been extensive debate surrounding the origin of Griko and Greko, as scholars have not reached agreement on whether the Greek of southern Italy originates in the Magna Graecia period, as claimed by Gerhard Rohlfs (1924, 1974) or the Byzantine period (Falcone 1973, among others). Indeed, southern Italy experienced two waves of Greek influence: one in the 8<sup>th</sup> century B.C. when the first Greek colonies were founded, forming what we know as Magna Graecia or Greater Greece, and the other in the 6<sup>th</sup> century A.D., when the Eastern Roman Empire, also known as the Byzantine Empire, reconquered the southern regions of the Italian Peninsula after the fall of the Western Roman Empire. Between these two periods, southern Italy was under Roman control. The question is hence whether Greek has been spoken continuously in southern Italy since the Magna Graecia period or if Griko and Greko descend from Byzantine Greek. The Italian linguist Fanciullo (2001: 69) highlighted the ideological nature of this controversy and defined it as a "false problem" since it was based on the assumption that Greek and Latin could not co-exist. Indeed, while Italian philologists have tended to support the Byzantine theory – as arguing the contrary would have jeopardised the "Italianness" of these people – Greek scholars have tended to favor the theory of continuity since ancient times. As discussed by Pellegrino (2015,

---

<sup>2</sup>In this work we distinguish between the word *Greek*, (which we use to refer to the language as it is spoken throughout Greece and Cyprus, including its regional and local varieties) and *Standard Modern Greek* (SMG) to refer to the official language as it is taught by state institutions.

2021) this represents a “language ideological debate” (Blommaert 1999) highlighting how contested language ideologies are appropriated differently in different historical periods, and by people with diverging aims.

What is undeniable is that the end of Byzantine rule marked the beginning of a language shift towards Romance, a slow process that was initially characterised by a long period of intense bilingualism between the local Greek and Romance varieties, which led to a progressive decrease in the number of speakers of the former. Nonetheless, the most abrupt decrease in the number of speakers occurred after the unification of Italy. In particular, from the beginning of the 20<sup>th</sup> century the newly formed Italian state fostered a wholesale Italianisation project of the peninsula, coupled with significant discrimination against people speaking local and in particular non-Romance varieties such as Greko and Griko. Furthermore, compulsory monolingual education in Italian, promoted particularly under the Fascist regime, the flow of emigration to the north of Italy and abroad, and later, the influence of mass media significantly contributed to the loss of the minority languages. In addition to this, in the case of Greko, the area’s geographical isolation, along with natural disasters in the 1950s and 1970s, played a major role in the depopulation of the Greko-speaking mountain villages. This displacement had major repercussions for the community, contributing to its disaggregation and favoring language abandonment (Stamuli 2008, Martino 2009, Squillaci 2018).

However, the profound socio-economic changes that took place, particularly following WWII, did not mechanically determine language shift. As noted by Pellegrino (2016a; 2021), there was instead “an existential shift” from a traditional to a modern worldview, causing communities to go through a difficult negotiation process of redefining their perceptions of themselves and of their group, along with their values and goals. These were then encoded through language. Although painful, abandoning Griko and Greko was considered a passe-partout for social enhancement (cf. Martino 1980, Stamuli 2008, and Squillaci in preparation).

Currently, the Griko-speaking community is composed primarily of middle-aged and elderly people with various degrees of language competence. Griko is mainly spoken in seven villages in the province of Lecce: Calimera, Castrignano dei Greci, Corigliano d’Otranto, Zollino, Sternatia, Martano, and Martignano (there were additionally Griko speakers in Melpignano and Soleto until the beginning of the 20<sup>th</sup> century). In Calabria, Greko is spoken today, after the aforementioned displacements, mainly in Conofuri, particularly in the hamlet of Gallicianò, in Roghudi Nuovo; in Bova; and in a few neighbourhoods in Reggio Calabria, Bova Marina as well as Melito P.S. The Greko-speaking community is smaller than its Griko counterpart, and the most concerning aspect is the age of the majority of speakers: most are over 80, with only a few in their 60s and 70s.

Thus, unlike Griko, Greko has very few speakers with limited competence aged between 40 and 50.

### 3 Griko and Greko language activism

Pellegrino (2016b, 2021) provides a diachronic account of Griko activism and refers to the first, middle, and current revivals of Griko. The first revival stretches from the end of the 19<sup>th</sup> century to the mid-1970s and is centred around the activity of the Philhellenic circle of Calimera. This was constituted by local intellectuals who were influenced by the contacts they had established with Greek folklorists. Because of the intellectualist nature of their efforts, however, language practice was not affected, so the revival did not prevent the shift from Griko to the local Romance variety and then to Italian. By the mid-1960s, the number of Griko mother-tongue speakers had dramatically decreased. The middle revival occurred in the late 1970s and 1980s. It was promoted by local cultural activists who funded cultural associations in the various Griko-speaking villages. This revival was not restricted to Griko and not limited to activities in support of the language, but included the local culture as a whole. It was a response to the break in cultural practices caused by the abrupt modernisation process in the years following WWII; it therefore presented itself as a form of redemption for a long-stigmatised South.

The current revival started in the 1990s and continues until the present. It is the outcome of interaction between the language policies and ideologies promoted by the EU, Italy, and Greece. Among the effects of the most recent revival of Griko, there is a sense of empowerment and pride in the rediscovered value given to the local cultural heritage (music and language included). However, while the revival has become a springboard for expressing a range of local claims, it does not include efforts specifically linked to the use of language as a tool of daily communication or to the training of new speakers. Crucially, over the years, the performative and artistic use of the language has increased while the use of Griko as a vehicle to convey information has progressively diminished. The more the language dies, the more it is resurrected performatively, as it were. The semiotic approach adopted by Pellegrino shows how Griko has now become a cultural and social resource, a form of performative post-linguistic capital, where the intentional, albeit limited, use of Griko becomes more important than “speaking” it as a means of exchanging information.<sup>3</sup>

---

<sup>3</sup>This also applies to the younger generation (starting from the 1970s), who typically are not speakers of Griko; however, they may use Griko by citing and re-appropriating words or entire expressions from memory, re-contextualised in the present, a practice which Pellegrino – building on Rampton (1995) – calls “generational crossing” (Pellegrino 2021).

With the exception of the first revival, which is only relevant to Griko, the chronological/analytical framework proposed by Pellegrino (2013) can be partially applied to the Greko community. In Calabria too, language activism developed in the late 1960s and 1970s generated a broader valorisation of the area as a whole which prompted the beginning of a decisive change in the use of and attitudes towards Greko at a community level. Martino (1980) defines this first phase as the *awakening*. During the second phase (from the late 1980s and 1990s) the activities dedicated to the Greko cultural and linguistic heritage have become more and more fragmented and often dependent on national and international program for minority languages, without long-term planning for revitalisation (Martino 2009). As noted by Pipirou (2016, among others), in the long run, the potential of minority language discourse and struggle turned into an appealing tool for many to achieve socio-economic benefits and prestige, often leading to tension and conflict within the community. At present, Greko has largely lost its communicative value in favor of a new symbolic function; however, rather than being used performatively and for performative purposes, as in the case of Griko reported above, what is mostly attested for Greko is its folklorisation: folklorised use of the language in specific contexts, such as official salutations and celebrations, to assert belonging to the Calabrian Greek heritage (Squillaci in preparation, drawing from Fishman 1991; see also Martino 2009 and Pipirou 2016).

Simultaneously however, we are witnessing a new awakening, which is mostly reflected in the recent coming together of a group of fifteen young people who, through the active involvement of Squillaci, have been carrying out language revitalisation activities. The group is part of the local association Jalò tu Vua, which has been working to promote the language since the early 1970s. Like the first movement of activism in the 1960s, this too stems from a sense of awareness of the cultural and linguistic heritage of the area, and is based in particular on a shared sense of responsibility towards language loss.

#### 4 The addressee's age and competence in Griko/Greko

To begin, we discuss the cases in which the addressee's age and competence in the minority language may influence the speaker's use/non-use of Griko and Greko. As argued by Pellegrino (2021), Griko is largely considered to belong to the older generation; this age-related factor seems to effectively exclude younger people from the world – the past – usually associated with Griko, since they did not live it. In a recursive way, perceptions of who can claim authority over Griko also define the authenticity of language and language practices.

## 5 What is the role of the addressee in speakers' production?

This also became clear during our joint fieldwork. As one of our middle-aged informants said, “The real speakers are the elderly, only what they speak is true Griko.” Consequently, anyone who is not old is not perceived as an “authentic” speaker, irrespective of competence or fluency. When Squillaci remarked that Pellegrino had indeed been speaking Griko throughout the evening, our informant commented, “The Griko that Manuela speaks is not the same language of the elderly”. Similarly, older speakers question the competence of middle-aged and younger speakers regardless of their actual production. We witnessed this attitude again in a conversational setting, when an older mother-tongue speaker, aged 92, commented on the production of a younger speaker in his mid-60s, pointing out that, “This is not Griko really” (*En’ ene probbiu griko*); this clearly links the age factor to language authority and competence. Likewise when we met another middle-aged informant in Grecia Salentina, we noticed that he would speak Griko with his elderly neighbours and also with us in the flow of the same conversation – but he spontaneously switched to Italian when interacting just with us. Pellegrino initially tried to keep the conversation in Griko, but as he clarified, he was not used to speaking Griko with her because of her age, and doing so would be odd (Pellegrino 2022).

Indeed, older speakers argue that it does not come “naturally” to speak Griko to younger people, and when approached by someone who makes an effort to speak the language, they tend to make fun of their mistakes. These instances reveal the power struggles embedded in the current revival of Griko, whereby the older generation claim authority over the language based on their embodied knowledge, and express skepticism and resistance towards younger speakers, and towards the more recent proliferation of language experts (Pellegrino 2016a, 2021).<sup>4</sup> This constant control over the language and resistance to change by the older generation or by language experts often leads younger speakers to switch and continue the conversation in the local Romance variety and/or Italian, as they feel sanctioned over the incorrect use of the language and thus disempowered.

Moving to southern Calabria we find a different picture. Here, middle-aged speakers are not *a priori* considered less competent due to their age, or because they are not necessarily native speakers, nor are they considered less authoritative. This is particularly true for long-standing activists – in our specific case the

---

<sup>4</sup>For a more in-depth analysis of locals’ ideologies with respect to the “purity” and “authenticity” of Griko, and the resulting power struggles within the community, see Pellegrino 2016a and 2021. Here it is argued that the multiple and competing criteria by which locals define the authenticity of language and language practices recursively determine who can claim authority over it, and vice versa.

interviewees have been engaging with Greko since their 20s and are considered part of the speakers' community on a par with older speakers. "Here in Bova Marina there is no one who speaks Greko anymore, just me [native speaker in his 80s], Demetrio [speaker in his 60s], Carmelo [speaker in his 60s], Salvatore [native speaker in his 80s], and Bruno [speaker in his 80s]". Pasquale thus includes as part of the pool of speakers those who – like Demetrio and Carmelo – are today in their 60s and 70s. Over fluency, in this case it is their long-standing engagement with the language which grants middle-aged speakers or semi-speakers legitimacy as full members of the speaker community.

The difference between Calabria and Salento becomes even more evident on consideration of the current intergenerational communication between older native Greko speakers and younger new speakers in their 20s and 30s. Here, fluency rather than age or long-standing language engagement seems to be the main criterion for establishing a conversation with elderly speakers. Fluency is intended by the majority of speakers as the capacity to *speak*, that is to answer back in Greko to the speakers' input, regardless of grammatical mistakes which are not taken into consideration by the majority, as we shall see. This is clear from the answer of an old Greko woman when a fellow villager commented on her use of the language with Squillaci, given that she usually refuses to speak Greko. The speaker replied, "*Egò to platego manachò me cinu ti to plategu!*" 'I speak it [Greko] only with those who do speak it', showing how fluency was crucial in establishing a conversation with her, over the age difference, as in Salento. Speakers' initial resistance is thus overcome, once they have the confirmation that the person wants to establish a conversation with them, not just to "hear them speak Greko", as speakers complain.<sup>5</sup> This often comes either after they are reassured by another speaker that their addressee does indeed have some competence in the language or after testing the addressee's competence themselves, usually asking the translation of some basic words. In the case of the recently formed group of new speakers, for instance, Squillaci's presence – as a young community member well known to the older speakers – and her organisation of multiple informal and formal intergenerational encounters facilitated this transition. Yet, it was the new speakers' ability to "answer back" that caught older speakers' attention and translated into respect for their efforts. At all the events we participated in dur-

<sup>5</sup>For many speakers, this is also because they feel treated as guinea pigs, as Petropoulou (1992) had already noticed back in the 1990s. The situation has been exacerbated over the years given the high numbers of visitors, journalists, and researchers who regularly visit the villages (when compared to the smaller numbers of inhabitants of such villages). In addition to this, many, especially women, refuse to be videoed or photographed and, given that their requests often are not honoured, they leave as soon as someone approaches them.

## 5 What is the role of the addressee in speakers' production?

ing our fieldwork, new and older/middle-aged speakers would naturally interact in Greko and the latter would rarely correct or comment on younger speakers' linguistic production in the way we observed in Grecia Salentina; when they did, it was clear that their aim was to explain the correct way unpretentiously. Some older speakers would at times clarify the fact that younger speakers "have not learned the language as we did" since they are the last generation "of this type" of Greko speakers (again including middle-aged speakers in this pool). However, different from Salento, it seems that they do so as a mere recognition of differences, which does not imply a delegitimisation of new speakers' language efforts. These are, instead, publicly recognised and valued by many as "the only possible future of the language", as phrased by one of the older speakers and an activist (Squillaci in preparation).

Like age, provenance also seems not to influence speakers' attitudes; many of these young people do not come from the last Greek-speaking area, but are nonetheless well accepted within the community. "Could you ever imagine we would have a Greko speaker in Campo Calabro [a non-Greko-speaking town]? This is a miracle!", an older speaker said after a conversation with one of the new Greko speakers. Particularly interesting in this respect was the fact that a very old speaker happily accepted as Greko language teacher a new speaker who does not originally come from the Greko area; he would instead dismiss his family members' language competence, as it is mainly passive. Indeed, Enzo considers self-evident that "*ecini en to plategu*" "they [his family members] don't speak it", as he put it, and therefore do not know it, inasmuch as they can hardly carry a full conversation in Greko. What they value over age and provenance is therefore that the addressee can *speak* Greko, regardless of potential mistakes. Old speakers adopted the same open attitude with Pellegrino regardless of the fact that she originally comes from another region and irrespective of her age; her knowledge of Griko facilitated conversations in Greko and thus her subsequent acceptance into the community of speakers.

We noticed, however, how the attitude of openness which characterises these older speakers with whom we have been in contact, does not hold true among all the members of the generation of language activists, most of whom are today in their 60s and 70s despite several attempts from the new speakers to also actively include them in the various events and activities.<sup>6</sup> We have indeed seen how, in the name of "authenticity", some long-standing activists are critical of

---

<sup>6</sup>During the first months of activities in particular, Squillaci regularly held one-to-one informal meetings with older and middle-aged Greko-speakers who have always been involved in language-related activities. These meetings were crucial for acquiring the support of many former activists for new speakers.

the activities that new speakers and activists carry out; this appears to be an attempt to undermine and perhaps delegitimise such language revitalisation efforts. As in the Salentine case, this situation reveals how such dynamics are in fact an expression of internal struggles over authority, attested in numerous minoritised contexts (O'Rourke & Ramallo 2013, Costa 2015, Sallabank 2017, Sallabank & Marquis 2018, among others). Unlike in Salento, however, in Calabria middle-aged activists' engagement with the language has actually decreased considerably over time, and their presence in language-related activities is limited nowadays, allowing younger activists the space to carry out their projects.

With respect to the factors behind the difference in attitudes towards using the language with new/younger speakers, we identified the different degree of language endangerment, community size, and different paths of language activism among the main contributing factors. In Grecìa Salentina, locals engage in multiple ways with Griko, and this translates into various forms of activism. These reveal how claims to language authority are diffused, and often translate into discussions around and about the language which generate tensions among activists, experts, and speakers. Such dynamics may lead not only to interpersonal but also intra-group estrangements based on locals' divergent positions on the form and the future envisioned for Griko (Pellegrino 2021: 157). In turn, this climate may discourage non-fluent and/or younger speakers from using the language, and from embarking on activities in its support. Indeed, as is often observed in minority language contexts, their efforts are to varying degrees delegitimised by the older generations, who consider themselves and are considered to be the gatekeepers of the language.

Interpersonal and intra-group estrangements were also commonly reported in the case of Greco up until the past decade, including extensive language monitoring and much negative judgment on the language activities that each association promoted; on the other hand, there also used to be substantially more activities, program, and events related to the Greco cultural and linguistic heritage (cf. Martino 2009 and Pipirou 2016). "If we had talked less about the language and more in the language we might not have been in this situation today," an activist told Squillaci. Today, instead, the even smaller size of the speaker community, its less active participation in language-related activities, the overall decrease in the number of activities and of active associations dedicated to the language, as well as the growing level of language "endangerment" lead to little intra-community discussion of language issues, and crucially to less conflict over language authority. As we have seen, this emerges mostly in relation to the former activists' position within the community, and it seems to be nonetheless limited. As is typical among dying language communities, "self-appointed monitors of grammatical

norms may become increasingly rare" (Dorian 1981: 154), favoring a "relaxation of internal grammatical monitoring"; this attitude creates a more favorable environment for language learners, who are not discouraged from continuing their efforts. "*Echome tunda pedia ti to sceru to greko, echome ta pilastria, tuto spiti den petti pleo*", "We have these young people who know Greko, we have the basement, our house won't fall anymore", an older activist proudly commented during a public event, publicly endorsing new speakers' language revitalisation efforts.

In conclusion, in addition to the generally limited use of the language discussed at the beginning of this section, language use in the Griko and Greko communities might be further inhibited by the addressee's younger age regardless of fluency, as seen in Salento, or favored by the addressee's language fluency, as shown in Calabria, regardless of his/her age. Such a difference between the two communities mirrors their different reception of language maintenance and/or revitalisation efforts. While in Salento multiple claims over language authority and widespread internal monitoring might discourage such attempts (cf. O'Rourke & Ramallo 2013, Costa 2015, Sallabank 2017, Sallabank & Marquis 2018, among others.), in Calabria the more open and supportive attitude towards new speakers/learners favors the language revitalisation activities they promote.

In the next section, we focus on the temporary variation produced by the speaker under the influence of the addressee's linguistic competence. In particular, we analyze this with reference to Greek, since this is an aspect which has not yet been investigated. We leave for further research the analysis of how Italian and the local Romance varieties (Salentine and Southern Calabrese) may trigger temporary variation.

## 5 The addressee's language competence in Greek

As is well known, multiple linguistic and extralinguistic factors influence speech production in minority language contexts, leading to temporary and/or permanent variation. Within the Griko and Greko communities at least three different types of variation can be identified. First, historical variation: both Griko and Greko display a great deal of internal variation, which has historically led to the emergence of Griko and Greko varieties specific to individual villages. This variation mostly amounts to phonological differences – the Greek consonant cluster  $\xi$  /ks/ has become /ʃ/ or /ts/ in Greko and /ʃ/, /ts/, /s:/, and /fs/ in Griko, for instance – as well as lexical differences among villages. Speakers are well aware of these distinctions, and may use them to claim the authenticity of one variant

over another. Second, “emergent variation”: both communities have long experienced increasing variation due to the endangered status of the varieties, which has affected speakers’ production and has led to the creation of idiolects. Third is “temporary variation”, which is linked to various factors, including speech production on demand (i.e., when speakers are asked to speak the language by researchers or tourists), speakers’ personal preferences, setting (e.g., interviews, public events), and temporary interference due to the spectrum of speakers’ linguistic resources (i.e., competence in Italian, local Romance varieties, and Standard Modern Greek).

In addition to this, we observed that speakers’ production may also be affected – albeit temporarily – by the addressee’s shared multiple linguistic competences. Interestingly, if this temporary interference involves Romance, it does not raise any questions at the community and academic levels about the competence of the speaker.<sup>7</sup> Different, instead, is the response to temporary variation, which involves Greek, and which regards, in particular, the speech of middle-aged and younger Griko and Greko speakers with a certain degree of knowledge of SMG. Instances of temporary variation are often perceived as a lack of competence in the minority language, and therefore may ultimately influence speakers’ use of the varieties.

The potential influence of SMG on Griko and Greko has indeed been a disputed topic both at the community and academic level over the years, and this dispute has intensified with the progressive increase in relations with Greece/Greek speakers, together with the introduction of SMG courses in both regions funded by the Greek Ministry of Education. Yet the language courses at schools have not proven successful in spreading SMG, so to speak, given the overall lack of interest in the subject and the fragmentation of the courses (these are also common issues regarding Griko/Greko language courses in schools). Similarly, those organised in collaboration with associations have attracted the curiosity of some locals; in particular in Salento, they tend to be attended by retired/elderly Griko speakers. Such courses, however, have not influenced the overall frequency of Griko and Greko usage.

Instead, what has played a bigger role in affecting the speech of some speakers – although to different degrees and in specific contexts – is the increase in contact with speakers of Greek (visitors, researchers, tourists, journalists) who regularly visit the Griko- and Greko-speaking communities: this applies, in particular, to the teachers sent by the Greek Ministry of education, who establish strong bonds

---

<sup>7</sup>Similarly, no questions at the community level are raised in the case of elderly speakers, who tend to code-shift if the addressee has limited or no knowledge of the minority language.

## 5 What is the role of the addressee in speakers' production?

with locals and take active part in the life of these communities. Moreover, the numerous exchange trips to Greece (and more rarely Cyprus) and the language scholarships it provided have also played a role. In Calabria, such contacts and exchanges have interested a larger part of the community than in Salento, and have left a mark in particular on the speech of younger (today, middle-age) generations who have been involved in language activism over the past 30 to 50 years, as well as of the older generations who have taken active part in these exchanges, albeit to a different degree. In Salento, instead, this dynamic has not played a significant role, particularly with regard to elderly speakers, as they have comparatively been less actively involved in these activities. This difference is in part linked to the actual size of the two communities: as the Greko-speaking community is comparatively smaller, more people had the opportunity to visit Greece or to come into direct contact with Greek visitors in Calabria, thereby acquiring varying degrees of knowledge of Greek. Interestingly, however, we also found that Greko speakers and activists have and had a more open attitude than their Griko counterparts with respect to the use of SMG loanwords; this is also reflected in less judgment from within the community over such use; we will come back to this topic in Section 6. In this section, we discuss the observed increase in their use of SMG loanwords when addressing people who do not originally come from the community.

In Salento, speakers who have taken courses in SMG and have achieved competence in the language also tend to be more directly involved in welcoming Greek visitors. During such encounters, therefore, they may use for instance *dromo* instead of the Salentine borrowing *stra/strata*, *oxi* instead of the Griko *de/nde/degghe*, or pronounce Griko words according to SMG phonological rules, as *ekino* over Griko *ecino* (with palatalisation of velar /k/) with the aim of facilitating comprehension – but sometimes also to demonstrate their competence in SMG. We noticed how they may equally reproduce such dynamics when interacting with anyone outside the community. For example, Michele used SMG words when speaking to Squillaci, since she was equally perceived as a visitor. In the first minutes of the conversation he used SMG words such as *pollà*, *kai*, *oikogenia*, before abandoning his role as a “tourist entertainer” and turning to Griko, also in response to Squillaci’s linguistic input.

Similar issues are attested in the case of Greko. For instance, we witnessed an increase in SMG words in the speech of a Greko native speaker and activist (in his 70s) when talking to a journalist of Greek origin. The speaker used *mikrò* over Greko *cceḍdi* “small”, *vunì* over Greko *oscia* “mountain”, *katalavennise* over Greko *kapegghise* – “you understand”, *dimarko* “mayor”, *taxidi* “trip”, and other borrowings which do not have a Greko equivalent. We must reckon that these

words are attested in the speech and in the writings of the speaker in question, regardless of the context, inasmuch as they belong to the class of SMG borrowings which are more widely employed in Greko (see Section 6). However, we noticed a significant increase in their use when the speaker addressed the journalist, and a considerable decrease and, in some recordings, a total absence of these borrowings, when the speaker addressed older speakers. This is in line with studies of audience design which demonstrate the active role of the addressee – or hearer, in Bell’s terminology (Bell 1984, 2001: 144) – in shaping the stylistic production of the speaker. This factor must be taken into consideration when performing a linguistic analysis<sup>8</sup>.

This happens all the more so with Griko and Greko speakers who have acquired a certain degree of knowledge of SMG, or who visit Greece regularly. In this case, in addition to lexical borrowings, we also observe temporary confusion between the two codes, which may partially and temporarily affect the morphosyntax of Greko and Griko. For instance, after returning from Greece, Giuseppe would consistently code-mix Greko with SMG for the first few days we spoke to him during our fieldwork. Similarly Giuseppe code-mixed and at times code-shifted to SMG when asked to speak Greko with Greek speakers, thus showing the effort to keep Greko and SMG apart in these specific interactions. In these cases, we mostly attested: (i) the use of the imperfective stem in finite complements introduced by the subordinator *na*, as shown in (1), rather than the perfective stem, which would be obligatory in Greko, shown in (2):

- (1) thelo na arotàò  
I.want NA I.ask.IMP
- (2) thelo na arotìo  
I.want NA I.ask.PERF

Notably, this phenomenon may be due not only to the influence of SMG, which permits both perfective and imperfective here, but also to language decay (Sasse 1992, among others). Indeed, speakers in general – not only those potentially influenced by SMG – often use the imperfective stem of the verb over the perfective one (cf. Squillaci in preparation).

---

<sup>8</sup>This is different from the case attested in Calabria, which concerns non-fluent speakers who have a passive knowledge of Greko and have been in contact with Greek people. Taking into account the generally limited use of Greko within the community, these speakers end up putting the language into practice mostly when interacting with Greeks. Consequently, they tend to employ Greek words or expressions in conversation, regardless of the addressee. This type of interaction reveals that opportunities to speak SMG, through visits by Greek tourists for instance, are becoming more frequent than opportunities to speak Greko.

(ii) The use of the *-ate* ending rather than the Greko *-ete* for the second person plural of the imperfect and aorist.

- (3) eplatèggate / eplatègghe

speak.IPFV.2PL

- (4) epiàsate / epiàete

take.AOR.2PL

Interestingly, however, Katsoyannou (1995: 288–290) reports the use of the *-ate* ending as part of the Greko morphological system. Given that this is not attested in other descriptions of the language, nor does it appear in our corpus unless under influence of Greek, it becomes difficult to establish whether Katsoyannou reported a previously undocumented case of an additional morphological ending in Greko or whether this is one of the first attestations of the use of this new ending under Greek influence.

(iii) We also observed, albeit extremely infrequently, the use of the SMG pluperfect form of the type *είχα πει* ‘I had said’ – with the auxiliary HAVE + the invariable form of the lexical verb, rather than the Greko *immon iponda*, composed of the imperfect of BE + the invariable participial form of the lexical verb.

Similarly, in Griko we find examples of SMG endings, such as the use of *-a* instead of Griko *-i* for the 2sg of the present indicative of *-ao* verbs, or the *-ate* desinence for the second person plural imperfect instead of Griko *-ato*:

- (5) milà [SMG pattern *milàs*]

- (6) milì [Griko]

you.speak

- (7) irtate [SMG pattern *irthate*]

- (8) irtato [Griko]

However, all these changes (as well as any others that might occur) are restricted to specific contexts. Most typically, contexts in which Griko and Greko speakers who also have some competence in SMG converse with Greek individuals, with people outside of the community, or when they have recently returned from Greece, sometimes also emulating SMG intonation. SMG influence seems to be otherwise attested in the lexis.

## 6 Speakers' responses to temporary language variation

In the case of language use in intergenerational settings, discussions about language variation also take on an ideological dimension, which in turn impinges on the perceived authenticity of the language. Griko speakers and activists remain concerned about the role that should be attributed to SMG. Some among them consider it “an agent of renewal” of Griko, helping to enrich and update its vocabulary through borrowings and adaptation. In this respect, contact with Greek visitors and/or the availability of SMG language courses has partially affected linguistic “taste” and choices. In some instances, such interplay and influence have generated or strengthened the “drive to ‘purity’” on behalf of some locals. This leads to practices of “verbal hygiene” (Cameron 1995) wherein speakers avoid using old or new borrowings from Salentine/Italian, or sanctioning such use. Salentine has indeed long been perceived as an agent of contamination of the perceived authenticity of Griko in a pre-contact past (hence the longstanding label of Griko as a “bastard language”; see Pellegrino 2016a). Such an unfortunate definition as an “agent of corruption”, of “contamination” of Griko is now increasingly extended to SMG, since the majority of local activists and speakers alike tend to condemn its use as an “artificial intervention” which may “kill Griko” by erasing its historical specificities. Indeed, conscious attempts to integrate Griko with SMG occurred in the past but were criticised for creating an “abstract” language. The linguistic boundaries between SMG and Griko are therefore under constant surveillance. Examples of even momentary confusion between the two, and of interference from Greek, are promptly noticed, commented on, and subject to negative judgement, thus casting doubt on the competence and hence the authority of the speaker, and in turn highlighting the moral dimension embedded in the perceptions of authenticity (Pellegrino 2016a, 2021). This phenomenon can be taken to extremes. Indeed, as we attested, a Griko speaker argued that he would not even use the greeting *kalimera*, since Griko speakers would instead greet each other in Italian.

In Calabria, as mentioned previously, there seems to be a greater tolerance towards loanwords in general, including those from SMG. In fact, some of the most widespread SMG borrowings are today also known to a wider number of locals who have not been in direct contact with Greek speakers. Most of these words first entered the language via contact and are attested in interviews, documents, and writings from the late 1970s. See for instance *charistò*, a Greko adaptation from SMG ευχαριστώ to say ‘thank you’ (instead of Greko *tosso obbligato* or *grazzi* from Romance), or SMG λουλούδι for ‘flower’ (instead of the Greko *attho* or the more productive Romance loanword *chiuri*), which has been used

## *5 What is the role of the addressee in speakers' production?*

in poems and songs since the late 1970s and 1980s (Squillaci 2021). Additionally, since the early years of the Greko language movement, several activists have proposed that SMG should be used as the source language for Greko neologisms and borrowings as Greko is a variety of Greek. On 21 November 2004, during a conference of local associations at the Regional Institute for the study of the Greko language, the decision was officially taken to use SMG as a source language for loanwords although, in written texts, people always had to include the Italian translation in parentheses so that older speakers would be able to understand (Condemi 2006: 10). Despite this decision, however, the question has remained open over the years, being revisited and discussed from time to time in local meetings. Interference and temporary variation from SMG is thus overall not sanctioned within the community in the way we see in Salento and in fact, many of those who officially do not embrace the use of SMG loanwords often show cases of interference, in specific contexts and according to their addressee, as shown in the previous section. Recently, however, speakers have increasingly begun to pay attention to such temporary changes as they feel under external pressure regarding authenticity.

Indeed, in addition to intra-community discussion, the academic community has also been actively involved in discussing the role of SMG. Whereas the impact/influence of local Romance varieties on Griko and Greko is regarded by scholars as the outcome of centuries-long contact, the potential insertion of SMG loanwords has been considered to dehistoricise both Griko and Greko (see Rohlf's in Petropoulou 1992, Karanastasis 1984), and would prove detrimental to local maintenance and revitalisation efforts. Particularly in Calabria, the attempt to use SMG loanwords has been defined as "fanciful and contaminant" (Martino 2009: 263), and as a threat to the authentic Greko language spoken by the older generations (Karanastasis 1984, Katsoyannou 2017, among others). As shown above, in most cases, the influence of SMG seems to be only a perceived threat rather than an actual abrupt and substantial linguistic change.<sup>9</sup>

The academic debate has nevertheless reached the communities and, in particular in Calabria, it has had significant repercussions on language use. In recent years, this has led to the emergence of linguistic practices of verbal hygiene (cf. Pellegrino 2021), with the aim of achieving what is assumed to be authenticity in

---

<sup>9</sup>What contributes to a degree of confusion is that some locals with limited competence in the minority language may instead be fluent in SMG. They may equally be involved in activities to valorise/promote Griko/Greko, access funding, and enjoy respect as minority language experts, locally as well as abroad. We wish to emphasise that these are specific and isolated cases which should be analyzed separately, as they highlight some of the controversial dynamics embedded in minority language contexts.

the language, i.e. avoiding any SMG loanwords. Greko speakers have thus begun to “clean up” their speech or their older poems, removing the SMG words they had inserted back in the 1990s and replacing them with their Romance counterparts to avoid criticism. For instance, they may replace *daskali* with *maistri* ‘teachers’, *elpizo* with *speregguo* ‘to hope’, *vivlio* with *libbro* ‘book’. Similarly, interviews and public events often arouse anxieties, as people feel the need to declare that what they speak is “real Greko” and often excuse themselves if some Greek word or expression “escapes from their mouth”, as they say. On the other hand, people who have been engaged with the language for years but who are not fluent have started to justify their limited use of the language in public by saying that they prefer to avoid speaking the language altogether rather than inserting SMG words. These people are then viewed externally as an expression of authenticity, thereby leading to significant power imbalances among community members. More crucially, these discussions have significant repercussions for language use, as they cast doubt on the authenticity of the language as well as of individual speakers, and favor a sort of monitoring and self-monitoring mechanism which in turn might increasingly discourage use of the language by those who actually speak it/can speak it.

## 7 Conclusions

In this chapter, we have highlighted a number of extra-linguistic factors that influence speech production, and that might inhibit or encourage the use of the minority languages. By analysing the role of the addressee we showed how these factors are not necessarily related to the linguistic competence of the speakers, and are instead the result of the wider dynamics at play within the communities. In the first half of the paper, we provided examples of the different responses of the older speakers in interacting with younger and/or non-fluent addressees. This has highlighted how in Grecìa Salentina, elderly speakers tend to be more reluctant to use the language with younger or new speakers. Moreover, the larger size of the community compared to the Calabrian case, along with the more widespread knowledge of the language – albeit with different levels of competence – leads to multiple claims of language authority, which results in various degrees of resistance to younger and new speakers, and to language change. This also affects speakers’ production, as internal monitoring and metalinguistic discussions tend to favor puristic attitudes and to delegitimise attempts to use the language.

In contrast, older speakers in Calabria favor the use of the language and related activities regardless of the age or provenance of the addressee, provided that

## 5 What is the role of the addressee in speakers' production?

they are fluent. This more inclusive attitude is also reflected in people's positive response to ongoing language revitalisation programs. Here, the smaller size of the community and the very old age of the speakers have led to less conflict over language authority, to a significant decrease in language-internal monitoring, and less resistance to language change.

In the second half, we discussed cases of temporary variation from SMG. Interestingly, in both communities we found an increase in the number of borrowings from SMG which are not necessarily related to speakers' competence in the minority language; instead, they are linked to the specific multilingual contexts in which the language is used, and crucially to specific characteristics of the addressee, among which language competence in Greek and provenance. This further proves the active role of the addressee in influencing speakers' production (Bell 2001) and it highlights, we argue, the need to include such dynamics in linguistic analyses, as these too might lead to predictable cases of temporary variation.

To conclude, this paper has also shown how the long tradition of prescriptivism has had direct repercussions on language use and activism in parallel, albeit opposite, ways in the two communities. In the case of Greko, speakers feel judged – by themselves and by others – to “fail” at speaking Greko if they insert SMG borrowings; Griko speakers instead may avoid inserting borrowings from Salentine/Italian, as they increasingly perceive doing so as failing to speak “Griko”. Such widespread resistance to and monitoring of language variation and change tend to undermine efforts to maintain or revitalise Griko and Greko. In this respect, both communities seem to have reached a stalemate.

## References

- Bell, Allan. 1984. Language style as audience design. *Language in society* 13(2). 145–204.
- Bell, Allan. 2001. Back in style: Reworking audience design. Style and sociolinguistic variation. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*, 139–169. Cambridge: Cambridge University Press.
- Blommaert, Jan. 1999. The debate is open. In Jan Blommaert (ed.), *Language ideological debates*, 1–38. Berlin: De Gruyter Mouton.
- Cameron, Deborah. 1995. *Verbal hygiene*. London: Routledge.
- Condemi, Filippo. 2006. *La lingua della valle dell'amendolea*. Vocabolario fraseologico. Reggio Calabria: Edizioni Ellenofoni di Calabria.
- Costa, James. 2015. *New speakers, new language: On being a legitimate speaker of a minority language in provence*. Vol. 2015. De Gruyter Mouton. 127–145.

- Dorian, Nancy. 1981. *Language death: The life cycle of a Scottish Gaelic dialect*. Philadelphia: University of Pennsylvania Press.
- Falcone, Giuseppe. 1973. *Il dialetto romanesco della bovesia*. Milan: Istituto Lombardo di Scienze e Lettere.
- Fanciullo, Franco. 2001. On the origins of modern Greek in Southern Italy. In Mark Janse, Brian Joseph & Angela Ralli (eds.), *Proceedings of the 2nd international conference on modern Greek dialects and linguistic theory*, 67–77. Patras: University of Patras.
- Fishman, Joshua A. 1991. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Vol. 76. Multilingual matters.
- Karanastasis, Anastasios. 1984. *Ιστορικόν λεξικόν των ελληνικών ιδιωμάτων της κάτω ιταλίας*. Athens: Academy of Athens.
- Katsoyannou, Marianna. 2017. *Οι ποτέμκιν της μεγάλης ελλάδας*. <https://sarantakos.wordpress.com/2017/05/22/magnapotemkin/>.
- Katsoyannou, Marianne. 1995. *Le parler gréco de Gallicianò (Italie): Description d'une langue en voie de disparition*. Paris 7. (Doctoral dissertation).
- Martino, Paolo. 1980. L'isola grecanica dell'aspromonte: Aspetti sociolinguistici. In Federico Albano Leoni (ed.), *I dialetti e le lingue delle minoranze di fronte all'italiano: Atti del XI congresso internazionale di studi*, 305–341. Rome: Bulzoni.
- Martino, Paolo. 2009. L'affaire bovesia: Un singolare irredentismo. In *Alloglossie e comunità alloglotte nell'Italia contemporanea: Teorie, applicazioni e descrizioni, prospettive, atti del XLI congresso internazionale di studi della società linguistica italiana (SLI)*. Consultato 21 aprile 2020. <https://doi.org/10.95>. Italy: Bulzoni Editore.
- O'Rourke, Bernadette & Fernando Ramallo. 2013. Competing ideologies of linguistic authority amongst new speakers in contemporary galicia. *Language in Society* 42(3). 287–305.
- Pellegrino, Manuela. 2013. “*Dying*” language or “*living monument*”? *Language ideologies, practices and policies in the case of Griko*. London: University College London. (Doctoral dissertation).
- Pellegrino, Manuela. 2015. I glossa griki: Ittè ce avri—la lingua greco-salentina tra passato e futuro. In Giovanni Azzaroni & Matteo Casari (eds.), *Raccontare la Grecia: Una ricerca antropologica nelle memorie del Salento Griko*, 515–547. Martano: Kurumuny.
- Pellegrino, Manuela. 2016a. *Performing Griko beyond ‘death.’* *Palaver* 5(1). 137–162.
- Pellegrino, Manuela. 2016b. I toponimi griki e le ideologie linguistiche riguardanti il griko. In Cirino Carluccio & Manuela Pellegrino (eds.), *Topos=logos, La vita dei luoghi*, 13–28. Martano: Kurumuny.

## 5 What is the role of the addressee in speakers' production?

- Pellegrino, Manuela. 2021. *Greek language, Italian landscape: Griko and the re-storying of a linguistic minority*. Washington DC: Center for Hellenic Studies. <https://chs.harvard.edu/book/greek-language-italian-landscape-griko-and-the-re-storying-of-a-linguistic-minority/>.
- Pellegrino, Manuela. 2022. "I was there": Agency, authority and morality among the Griko linguistic minority of Southern Italy (Apulia). In M.S. Linn & A. Dayán-Fernández (eds.), *The periphery as centre: Examining social actors and agency in European language reclamation, multilingual matters*. Manuscript submitted for publication.
- Petropoulou, Christina. 1992. *Lingua e dialetto nella Grecia Calabrese*. Archivio storico per la Calabria e la Lucania LIX. 153–170. Aspetti linguistici e culturali.
- Pipyrou, Stavroula. 2016. *The Grecanici of Southern Italy: Governance, violence, and minority politics*. Philadelphia: University of Pennsylvania Press.
- Rampton, Ben. 1995. *Crossing: Language and ethnicity among adolescents*. London: Longman.
- Rohlfs, Gerhard. 1924. *Griechen und romanen in unteritalien: ein Beitrag zur Geschichte der unteritalienischen Gräzität*. LS Olschki. 1–177.
- Rohlfs, Gerhard. 1974. Scavi linguistici nella magna grecia. *Galatina (Puglia)*.
- Sallabank, Julia. 2017. Purism, variation, change and 'authenticity': Ideological challenges to language revitalisation. *European Review* 26(1). 164–178.
- Sallabank, Julia & Yan Marquis. 2018. "We don't say it like that": Language ownership and (de)legitimising the new speaker of guernesiais. In M. Hornsby & C. Smith-Christmas (eds.), *New speakers of minority languages: Linguistic ideologies and practices*, 67–90. Basingstoke: Palgrave Macmillan.
- Sasse, Hans-Jürgen. 1992. Language decay and contact-induced change: Similarities and differences. In Matthias Brenzinger (ed.), *Language death: Factual and theoretical explanations with special reference to East Africa*, 59–80. Berlin: De Gruyter Mouton.
- Squillaci, Maria Olimpia. 2018. *Weaving identities through language*. Paper presented at the conference 'Political Logic of Mediterranean Landscapes', University of Notre Dame, Rome.
- Squillaci, Maria Olimpia. 2021. *From contact to 'contamination' and beyond. What's left of Calabrian Greek and its speakers in 2021*. (Paper presented at the MGDLT 9, Leonidio, 5–6 June 2021).
- Squillaci, Maria Olimpia. in preparation. *SMiLE Report*. Tech. rep. Center for Folklife & Cultural Heritage, Smithsonian Institution.
- Stamuli, Francesca. 2008. *Morte di lingua e variazione lessicale nel greco di calabria. Tre profili dalla bovesia*. Naples. (Doctoral dissertation).



# Chapter 6

## Innovations in the Contemporary Hasidic Yiddish pronominal system

Zoë Belk<sup>a</sup>, Lily Kahn<sup>b</sup>, Kriszta Eszter Szendrői<sup>c</sup> & Sonya Yampolskaya<sup>b</sup>

<sup>a</sup>Department of Linguistics, University College London <sup>b</sup>Department of Hebrew and Jewish Studies, University College London <sup>c</sup>Department of Linguistics, University of Vienna

Although under existential threat in the secular world, Yiddish continues to be a native and daily language for Haredi (Hasidic and other strictly Orthodox) communities, with Hasidic speakers comprising the vast majority of these. Historical and demographic shifts, specifically in the post-War period, in the population of speakers have led to rapid changes in the language itself. These developments are so far-reaching and pervasive that we consider the variety spoken by today's Haredi speakers to be distinct, referring to it as Contemporary Hasidic Yiddish. This chapter presents a study involving 29 native Contemporary Hasidic Yiddish speakers, and demonstrates that significant changes have occurred in the personal pronoun, possessive, and demonstrative systems. Specifically, the personal pronoun system has undergone significant levelling in terms of case and gender marking, but a distinct paradigm of weak pronominal forms exists, independent possessives have lost case and grammatical gender distinctions completely, and a new demonstrative pronoun has emerged which exhibits a novel case distinction.

### 1 Introduction

In this chapter, we describe the innovative features of the pronominal system of Contemporary Hasidic Yiddish. We begin with a brief historical introduction to the language and to its pronominal system.



Yiddish is the traditional language of the Ashkenazic (Central and Eastern European) Jews. It can be divided into two major varieties, Western and Eastern Yiddish. Western Yiddish was used by Jews living in the Western European regions corresponding to present-day Germany, France, the Netherlands, and Switzerland throughout the medieval and early modern periods, but was largely abandoned over the course of the 18<sup>th</sup> century in favour of the dominant co-territorial languages and had become largely moribund by the 19<sup>th</sup> century, though it retained a small spoken presence into the 20<sup>th</sup> century (see Fleischer 2018). Eastern Yiddish was used by Jews in Eastern Europe, with the largest populations in regions corresponding to present-day Poland, Hungary, Romania, Ukraine, Lithuania, and Latvia. Henceforth all references to Yiddish in this chapter will denote Eastern Yiddish. Eastern Yiddish is characterised by a core Germanic morphosyntactic structure and lexis with a substantial Semitic (Hebrew and Aramaic) lexical component and a smaller but relatively high-frequency Slavic (chiefly Polish and Ukrainian) lexical component, with some Slavic-influenced grammatical features. It can itself be subdivided into three chief dialect areas, termed Northeastern or Lithuanian Yiddish (traditionally spoken in areas corresponding primarily to present-day Lithuania, Latvia, and Belarus), Mideastern or Central Yiddish (traditionally spoken in areas corresponding primarily to present-day Poland and Hungary), and Southeastern or Ukrainian Yiddish (traditionally spoken in areas corresponding primarily to present-day Ukraine). Yiddish in Eastern Europe served not only as a vernacular but also as a written language; it existed in a diglossic relationship with Hebrew, with the former typically used for more low-prestige types of writing and the latter functioning as the high-prestige written vehicle though not as a vernacular. (See Harshav 1990 for discussion of the traditional Hebrew-Yiddish diglossia in Eastern Europe.) In the 1920s and 1930s, a standardised variety of Yiddish based largely on the Northeastern dialect was developed by the Vilna-based YIVO Institute, an organisation devoted to Yiddish pedagogy and linguistic research as well as other scholarly activities focused on Eastern European Jewry (see Kuznitz 2010).

Hasidism is a Jewish spiritual movement which originated in late 18<sup>th</sup>-century Ukraine and grew to become a prominent force in Eastern European Jewish society over the course of the 19<sup>th</sup> century (see Biale et al. 2018 for a historical overview of Hasidism). Like other Eastern European Jews, most followers of the Hasidic movement spoke Yiddish as their native and main language. During the 19<sup>th</sup> century the various Hasidic rebbes, or spiritual leaders, established dynasties that were typically named after the locations where they were founded. Most of the Hasidic dynasties (e.g. Belz, Satmar, Skver, Tosh, Vizhnitz) were based in the Mideastern (Polish and Hungarian) and Southeastern (Ukrainian) Yiddish

dialect areas, with a smaller number (e.g. Chabad and Karlin) based in the North-eastern (Lithuanian) dialect area. The Yiddish spoken by followers of the Hasidic movement did not differ from that of their non-Hasidic counterparts but rather varied by regional dialect. This traditional situation changed dramatically with World War II and the cataclysmic destruction of the majority of Yiddish speakers during the Holocaust. Surviving members of the various Hasidic dynasties were dispersed from their traditional homelands in Eastern Europe and resettled in new locations around the world, chiefly the New York area in the US, a variety of communities in Israel, the Montreal area in Canada, London's Stamford Hill neighbourhood in the UK, and Antwerp in Belgium. This rapid geographical shift led to a complete realignment whereby speakers of different varieties of Eastern European Yiddish were now living side by side, which contributed to a high degree of dialect contact and mixing. Moreover, these post-War Hasidic communities had little contact with secular Yiddish speakers in the new locations (in contrast to Eastern Europe, where Hasidic and non-Hasidic Yiddish-speaking Jews had typically lived in the same areas), and as such Hasidic Yiddish began to develop separately. This rapid shift was compounded by the fact that these post-War Hasidic communities absorbed a substantial number of L2 speakers who adopted the language as adults. These post-War Hasidic centers also came to accommodate smaller groups of non-Hasidic Haredi (strictly Orthodox) Jews, primarily from Northeastern Yiddish dialect areas, which became increasingly integrated with their Hasidic counterparts as both communities had much in common due to their shared Haredi cultural tradition. As a result of these inter-related factors, the Yiddish of these new Haredi (primarily Hasidic) communities developed very quickly, to the extent that in the 21<sup>st</sup> century, it can be regarded as a distinct variety of the Yiddish language. Indeed, our previous research has suggested that a greater amount of dialect mixing and L2 speakers at a community level in the years since World War II is associated with increased use of innovative features; see Belk et al. (In press) for further discussion. We term this new variety Contemporary Hasidic Yiddish.

One of the most prominent of these is a complete absence of morphological noun case and gender, which contrasts dramatically from the pre-War tripartite case and gender system (see Belk et al. 2020, In press), although see Assouline (2014) for the claim that animate nouns in Jerusalemitic Yiddish reliably retain gender but not case morphology.<sup>1</sup> This rapid loss of morphological case and gender, as well as the above-mentioned dialect mixing which contributed to the

---

<sup>1</sup>Note that Northeastern Yiddish only had two morphological genders, although it had the same three cases as other pre-War varieties of Yiddish (Jacobs 1990).

emergence of Contemporary Hasidic Yiddish, are linked to a number of innovations in the pronominal system. There has been very little research into Hasidic Yiddish pronouns in general, namely Assouline's (2010) study of the first person plural pronoun in Haredi Jerusalemitic Yiddish, Nove's (2018) study of case syncretism in the first and second person objective singular pronouns in New York Hasidic Yiddish, and Sadock & Masor's (2018: 95, 103) observation on the use of the demonstratives *dey* and *deye* in the language of Bobover Hasidic speakers in New York; Assouline (2014) has also briefly discussed this form, while Krogh (2012) mentions the existence of *deys*. We seek to complement these previous studies by examining the major innovative features of the Contemporary Hasidic Yiddish pronominal system based on elicited spoken data provided by native speakers in the main Hasidic centers globally.

Our investigation focuses on the personal pronoun paradigm, strong vs. weak pronouns, possessives, and demonstratives. Our research found that Contemporary Hasidic Yiddish has undergone a realignment of the personal pronoun paradigm, with an increase in case syncretism in both the singular and plural whereby the singular forms exhibit a two-way nominative / objective distinction and the plural forms tend to exhibit no case distinctions, as well as the introduction of a distinct 2PL.HON pronoun. The pronominal system includes a distinction between strong and weak forms, which have different morphological and syntactic properties. With respect to the possessives, our main findings are that there is a clear distinction between singular and plural dependent forms, and between dependent and independent forms, with the independent forms exhibiting a complete lack of case and gender morphology (in contrast to pre-War and Standard Yiddish). Our main findings with respect to the demonstratives are that there is a novel distinction between nominative and objective independent demonstratives suggesting the innovation of a distinct pronominal form, as well as a complete lack of case, gender, and number distinctions. Furthermore, the distribution of the 'proximal' and 'distal' forms differs from that of pre-War dialects, English and Modern Hebrew.

### 1.1 The pronominal system of pre-War and Standard Yiddish

Almost all traditional geographical dialects of Eastern Yiddish, as well as Standard Yiddish, exhibit the same tripartite case system for personal pronouns as for full nominals (Kahn 2016). Pronouns decline in the nominative, accusative, and dative, as shown below (based on Kahn 2016: 678); note that there is a degree of syncretism present in the paradigm, provided in Table 1.<sup>2</sup>

<sup>2</sup>Published descriptions of the different dialects of pre-War Yiddish are often incomplete or only cover larger dialect areas. It is therefore possible that some of the forms described in this

Table 1: Case and gender marking on personal pronouns in pre-War/Standard Yiddish

		Singular			Plural		
		NOM	ACC	DAT	NOM	ACC	DAT
1 <sup>st</sup>		<i>ikh</i>	<i>mikh</i>	<i>mir</i>	<i>mir</i>	<i>undz</i>	
2 <sup>nd</sup>	FAM	<i>du</i>	<i>dikh</i>	<i>dir</i>	<i>ir</i>	<i>aykh</i>	
	HON	<i>ir</i>	<i>aykh</i>				
3 <sup>rd</sup>	MASC	<i>er</i>	<i>im</i>				
	FEM	<i>zi</i>		<i>ir</i>			
	NEUT	<i>es</i>		<i>im</i>			

Within this system there is a degree of regional variation; for example, the second person plural form in certain Mideastern Yiddish varieties is *ets* (NOM) or *enk* (Acc/DAT) rather than *ir* (Jacobs 2005: 70). Similarly, there is syntactic variation whereby particular verbs that take an accusative in most dialects of Eastern Yiddish instead take a dative in certain local varieties; this can be seen in examples involving the first person singular and third person feminine singular pronouns. For example, the verb *kenen* ‘to know [a person]’ typically takes the accusative, i.e. *er ken mikh* ‘he knows me’, *er ken zi* ‘he knows her’, but the dative in certain (mostly Northeastern but also some Mideastern) varieties, i.e. *er ken mir*, *er ken ir* (Wolf 1969: 142–147). Jacobs (2005: 184) goes further than Wolf in claiming that pre-War Northeastern Yiddish lacked the historical three-way distinction in the 1SG, 2SG, and 3FS entirely, employing the historically dative forms in both accusative and dative settings. Conversely, use of the accusative in contexts where the dative would typically be used do not seem to be attested (Wolf 1969: 142–146). It is important to note that, as with the noun gender and case variation discussed above, these phenomena seem to be restricted to specific individual verbs, rather than pointing to a breakdown in the pronominal case system as a whole. However, the first person phenomenon may be at least partly based on a lack of phonological distinction between the sounds /r/ (often realised as a uvular trill or fricative) and /χ/ in the dialects in question, and may have constituted the first step in a more widespread hypothetical future merger (Wolf 1969: 149).

chapter existed in pre-War dialects, but this claim is difficult to substantiate. In any case, it is clear that our findings represent a much more widespread and generalised phenomenon across the Hasidic Yiddish speech community than existed in pre-War Yiddish.

### 1.1.1 Verbal agreement

Verbal agreement for pre-War and Standard Yiddish is shown in Table 2. Note that the Mideastern first person plural pronoun *undz* and the second person plural pronoun *ets* are dialectal, and take the same verbal agreement as the equivalent pronouns *mir* and *ir*, namely -(e)n and -t respectively.<sup>3</sup>

Table 2: Subject-verb agreement morphology in pre-War and Standard Yiddish

			Singular		Plural	
1 <sup>st</sup>			<i>ikh</i>	<i>V-Ø</i>	<i>mir</i> (/ <i>undz</i> )	<i>V-(e)n</i>
2 <sup>nd</sup>	FAM		<i>du</i>	<i>V-st</i>	<i>ir</i> (/ <i>ets</i> )	<i>V-t</i>
	HON		<i>ir</i>	<i>V-t</i>		
3 <sup>rd</sup>	MASC		<i>er</i>			
	FEM		<i>zi</i>	<i>V-t</i>	<i>zey</i>	<i>V-(e)n</i>
	NEUT		<i>es</i>			

### 1.1.2 Reflexives

The reflexive pronoun in Northeastern and Southeastern Yiddish, as well as in the standardised variety, is the invariant form *zikh* ‘oneself’, which is used for all persons and numbers without case or gender distinctions. In Mideastern Yiddish, by contrast, the 1<sup>PL</sup> and 2<sup>PL</sup> accusative forms of the personal pronouns (*mikh* and *dikh* respectively) are used as reflexive forms, while the 1<sup>PL</sup> and 2<sup>PL</sup> accusative/dative forms of the personal pronouns (*undz* and *enk* respectively) are used as reflexive forms, though less consistently than in the singular (Jacobs 2005: 184–185).

### 1.1.3 Possessive pronouns

The pre-War and Standard Yiddish possessive pronouns show a distinction between dependent and independent forms. Dependent forms lack any case and gender distinctions, but distinguish between forms used to modify a singular

<sup>3</sup>The 1<sup>PL</sup> dialectal variant *undz* also appeared with the verbal suffix *-mer/mir*, e.g. *undz zog(n)mer/undz zog(n)mir*, see Jacobs (2005: 70, 189). For discussion of the verbal suffix *-ts* as agreeing with the pronoun *ets*, see Section 3.1.1.

noun and forms used to modify a plural noun. The pre-War/Standard Yiddish dependent possessive pronouns are shown in Table 3.

Table 3: Pre-War/Standard Yiddish dependent possessive pronouns

		Sg. possessor		Pl. possessor	
		Sg. possessum	Pl. possessum	Sg. possessum	Pl. possessum
1 <sup>st</sup>		<i>mayn</i>	<i>mayne</i>	<i>undzer</i>	<i>undzere</i>
2 <sup>nd</sup>	FAM	<i>dayn</i>	<i>dayne</i>	<i>ayer</i>	<i>ayere</i>
	HON	<i>ayer</i>	<i>ayere</i>		
3 <sup>rd</sup>	MASC	<i>zayn</i>	<i>zayne</i>	<i>zeyer</i>	<i>zeyere</i>
	FEM	<i>ir</i>	<i>ire</i>		
	NEUT	<i>zayn</i>	<i>zayne</i>		

The independent possessive pronouns, by contrast, take case and gender markings in accordance with the case and gender of the associated noun. This system is exemplified in Table 4, which shows the various forms of the first person singular independent possessive pronoun. There are also postpositive forms of the dependent possessive pronouns, and these take the same case and gender endings shown in Table 4 (e.g. *mitn khaver maynem* ‘with-DEF.M.SG.DAT my-M.SG.DAT friend.M’). See Katz (1987: 108–112) and Jacobs (2005: 183–184) for further discussion of the possessive pronouns in pre-War and Standard Yiddish.

Table 4: Pre-War/Standard Yiddish independent 1SG possessive pronoun forms

	MASC	FEM	NEUT	PL
NOM	<i>mayner</i>	<i>mayne</i>	<i>mayns</i>	
ACC	<i>maynem</i>	<i>mayne</i>	<i>mayns</i>	<i>mayne</i>
DAT	<i>maynem</i>	<i>mayner</i>	<i>maynem</i>	

#### 1.1.4 Demonstratives

The pre-War and Standard Yiddish demonstrative pronouns exhibit morphological case and gender, with masculine, feminine, and neuter forms (except for

Northeast Yiddish, which has only masculine and feminine forms). There was a set of proximal demonstratives and a set of distal demonstratives, but no distinction between dependent and independent forms. The stressed definite article was used for the proximal demonstratives, as shown in Table 5. The definite article could additionally be accompanied by prepositive *ot* or by postpositive *dozik* (plus case and gender suffixes) to reinforce the demonstrative sense. See Katz (1987: 112–114) for further discussion of the proximal demonstratives in pre-War and Standard Yiddish.

Table 5: Pre-War/Standard Yiddish proximal demonstratives

	MASC	FEM	NEUT	PL
NOM	<i>der</i>	<i>di</i>	<i>dos</i>	
ACC	<i>dem</i>	<i>di</i>	<i>dos</i>	<i>di</i>
DAT	<i>dem</i>	<i>der</i>	<i>dem</i>	

The distal demonstratives are based on the stem *yen-*, which takes case and gender endings, as in Table 6. Note that Northeastern Yiddish typically lacked the neuter form *yens*, instead employing the masculine or feminine forms. See Katz (1987: 115–116) and Jacobs (2005: 186) for further discussion of the distal demonstratives in pre-War and Standard Yiddish.

Table 6: Pre-War/Standard Yiddish distal demonstratives

	MASC	FEM	NEUT	PL
NOM	<i>yener</i>	<i>yene</i>	<i>yens</i>	
ACC	<i>yenem</i>	<i>yene</i>	<i>yens</i>	<i>yene</i>
DAT	<i>yenem</i>	<i>yene</i>	<i>yenem</i>	

### 1.1.5 Road map

In the remainder of this chapter, we provide an overview of our participants and study design (section 2) and introduce the personal pronoun system of Contemporary Hasidic Yiddish including strong and weak pronoun forms (Section 3). We then discuss innovations in possessives (Section 4) and demonstratives (section 5). Section 6 provides some concluding remarks.

## 2 Methodology

### 2.1 Participants

Our analysis of the pronoun system in Contemporary Hasidic Yiddish is based on interviews with 29 native speakers between the ages of 18 and 72 who were born and raised in Haredi communities worldwide. We worked with 15 participants from the New York area (six female) from a range of Haredi neighbourhoods in Brooklyn, as well as several Haredi communities in upstate New York. Five participants (two female) were from communities in Israel such as Bnei Brak, Ashdod, and in and around Jerusalem. In addition, five participants (four female) were from London's Stamford Hill community, two (both female) were from the Montreal area, and two (both male) were from Antwerp.<sup>4</sup>

Contemporary Hasidic Yiddish constitutes a distinct variety of the language, although it is most closely related to the historical Mid- and Southeastern dialects. Our participants largely identify as speaking *khsidiske yidish* 'Hasidic Yiddish', which is associated with a vowel profile most closely matching that of traditional Mid- and Southeastern dialects. Participants from Chabad and those non-Hasidic Haredi speakers who refer to themselves as 'Litvish' or 'yeshivish' typically speak a variety more closely related to the historical Northeastern dialect, with the associated vowel profile.<sup>5</sup> These groups are often distinguished by their pronunciation of the word *ונן* 'what', with the former group, who pronounce it as [vʊs], described as speaking 'vus' and the latter, who pronounce it [vɔs], described as speaking 'vos'. We will use these terms throughout the rest of the chapter. In our sample, 'vos' speakers include participants from Chabad and those non-Hasidic Haredi speakers who refer to themselves as 'Litvish' Yiddish speakers. Those who speak 'vus' in our dataset include a wide range of other Hasidic affiliations, including Belz, Dushinsky, Karlin, Pupa, Satmar, Skver, Toldos Avrohom Yitzchok, Tosh, Tsanz, Vizhnitz, Vizhnitz-Monsey, and so-called 'klal Hasidish', i.e. non-specific/general Hasidic. It is important to note also that while some Hasidic sects are associated with one pronunciation or the other (e.g. Satmar and Belz are associated with 'vus' while Chabad and Karlin are associated with 'vos'), individual speakers inside those communities might differ. Our sample includes 24 speakers of 'vus' (14 female), four speakers of 'vos' (none female),

<sup>4</sup>Given these sample sizes, the Montreal and Antwerp data should be viewed as some indication that the changes we describe are generalised, but more data would be necessary to get a more reliable picture.

<sup>5</sup>Certain other groups use a similar vowel profile, notably the Yerushalayimer or Jerusalem Haredi community. See Belk et al. In press for further discussion.

and one (female) who speaks a ‘mixed’ phonological variety skewing mostly towards ‘vos’. Three of the ‘vos’ speakers are from the New York area and one is from Israel.<sup>6</sup>

All of our participants were raised in Yiddish-speaking homes and were largely educated in Yiddish, particularly in their early years.<sup>7</sup> Many of our participants use Yiddish on a daily basis with family, friends and business contacts. Others employ it only on occasion (e.g. when talking to particular family members or friends). Some speakers also employ English, Modern Hebrew, and/or French (depending on the location) on a regular basis, while others speak Yiddish almost exclusively. All participants are comfortable reading and writing in Yiddish, though not all of them regularly employ the language in these ways.

The participant codes we use in this paper take the following form. The first character indicates the geographical community in which the participant grew up (N=New York, I=Israel, S=Stamford Hill, M=Montreal, A=Antwerp), the second character represents the participant speaks ‘vus’ or ‘vos’ (U=‘vus’, O=‘vos’, M=mixed), and the third character is serves to provide a unique identifier.

## 2.2 Description of tasks

Our study of developments in the Contemporary Hasidic Yiddish pronominal system is primarily based on elicited oral data. Participants were presented with a series of short sentences in either English or Modern Hebrew (according to the participant’s preference) and asked to translate each into Yiddish. The sentences target personal pronouns in each of the person, gender and case combinations, as well as second person familiar and honorific contexts; independent and dependent possessives for each of the pronominal persons in both singular and plural contexts; and independent and dependent demonstratives (both proximal and distal) in singular and plural nominative, accusative, and dative contexts.

In addition, some participants completed a written task which targeted the familiar/honorific distinction in more detail. This task presented participants with a variety of scenarios in Yiddish involving direct address to a range of interlocutors, and participants were asked to provide the form of address they would most naturally use.

<sup>6</sup>Again, given the relative sizes of the samples of ‘vus’ and ‘vos’ speakers, data on the ‘vos’ variety should be taken as indicative. However, it should be noted that relatively little variation was observed in the language of the ‘vos’ speakers.

<sup>7</sup>The medium of education varies by gender, especially in the later years, with boys largely being educated in Yiddish and *loshn koydesh* (the traditional Yiddish term for pre-modern Hebrew, and girls receiving more education, especially in secular subjects, in a co-territorial language).

Participants also engaged in Yiddish-, English-, and Modern Hebrew-medium metalinguistic discussion with the experimenter which both provided additional unelicited data and insight into their linguistic choices. In certain cases, we also made use of Yiddish-language Hasidic internet resources to verify tendencies observed in our elicited data.

The interview recordings were transcribed in ELAN by fluent or native Yiddish speakers using a modified version of the YIVO standard transliteration system and a modified IPA.

### **2.3 A note on the transliteration system used in this paper**

In most instances, the data presented in this paper are represented in writing according to the standard transliteration system developed by the YIVO Institute, which is in widespread use for Romanisation of Yiddish. This transliteration system is based on the Hebrew-script orthography designed by the same Institute, which is the standard orthography throughout the non-Haredi Yiddish-speaking world. This system is based primarily on the North- and Southeastern vowel profiles of Eastern Yiddish, and as such the representation employed here obscures some of the pronunciation features characteristic of the Mideastern dialect region, which is more typical of the majority of our participants. In most instances, the differences in vowel patterns between these dialect areas are predictable. For example, the vowel /u/ is regularly pronounced as [i] in the Mideastern dialect area, and consequently also in the speech of our ‘vos’ participants; conversely, the diphthong /ɔɪ/ is regularly pronounced as [eɪ] in the Northeastern dialect area, and consequently also in the speech of our ‘vos’ participants. As the YIVO transliteration system serves to represent a variety of actual phonetic realisations, we have supplemented it with IPA representations where required for clarity.

## **3 Personal pronouns**

The paradigm of personal pronouns in Contemporary Hasidic Yiddish exhibits a number of innovative uses of the forms known from the pre-War and Standard varieties of Yiddish. In this section we discuss the basic personal pronoun paradigm, and describe the distinction between the strong and weak forms of these pronouns.

### 3.1 The personal pronoun system of Contemporary Hasidic Yiddish

There is a considerable amount of variation in the Contemporary Hasidic Yiddish pronoun paradigm, primarily between different speakers, but also sometimes within the same speaker. This variation seems to reflect a pronominal paradigm in flux, a situation which is likely ascribable to the increased degree of dialect mixing in the speech of previous generations of speakers who came together in the new Hasidic centers in the immediate post-War period; see also Nove (2018) for a discussion of the role that sociolinguistic factors play in the tendency towards accusative/dative syncretism in first and second singular pronouns in New York Hasidic Yiddish. Five main pronominal paradigms can be distinguished among the speakers. The first two, shown in Tables 7 and 8, are the most common and, combined, account for most speakers.<sup>8,9</sup>

Together these two paradigms reflect the speech of many speakers from the ‘vus’ vowel profile. The only difference between them is that in the paradigm shown in Table 7, the 1SG and 2SG forms appear as *mikh* and *dikh* respectively in both the accusative and dative positions, whereas in Table 8, these same forms

---

<sup>8</sup>We transliterate the 3MS objective form as *em* despite the fact that in the standard YIVO transliteration system it is represented as *im* because the vast majority of our participants pronounce it as *em* and it does not represent a phonologically predictable vowel change (in contrast to other forms represented here by the YIVO transliteration system, such as *undz*, which speakers with a ‘vus’ vowel profile typically pronounce as [indz] or [ints] in accordance with a predictable regionally based *u/i* vowel alternation plus a final devoicing process). One Chabad speaker with a ‘vos’ vowel profile produces some tokens of [im], which corresponds to the spelling used in most Hebrew-script orthographic conventions and in the standard YIVO system. The form *eym* was also found in some, especially Mideastern, pre-War dialects (Weinreich 1964, Jacobs 2005).

<sup>9</sup>In the third person singular, the masculine and feminine forms are almost exclusively used for animate entities by speakers, replacing a former grammatical gender distinction with a semantic one. The former neuter form, *es*, is now used as an inanimate third person singular pronoun. However, there is quite a lot of variation in our data with respect to the (morpho)phonological form of this morpheme and our data questionnaire was not designed to probe this issue. Therefore, in this paper, we do not provide inanimate third singular pronominal forms and leave this topic for future research.

We also note that while the masculine and feminine singular pronouns are reserved for (usually human) animates for most speakers, Israeli speakers sometimes deviate from this general rule by employing the masculine *er* or feminine *zi* 3SG pronouns in conjunction with inanimate nouns in accordance with the gender of those nouns in Modern (Israeli) Hebrew, as exemplified in (i).

- (i) neyn, de seyfer iz nisht mayne, er              iz dayne (IU1)  
no      the book is not mine    it (lit: he) is yours

Table 7: ‘Vus’ paradigm with *mikh* and *dikh* in both accusative and dative positions

		Singular		Plural	
		NOM	OBJ	NOM	OBJ
1 <sup>st</sup>		<i>ikh</i> /iχ/	<i>mikh</i> /miχ/	<i>undz</i> /inz/	
2 <sup>nd</sup>	FAM	<i>du</i> /di/	<i>dikh</i> /diχ/	<i>enk</i> /εŋk/	
	HON	<i>ir</i> /ir/	<i>aykh</i> /aχ/	<i>ir</i> /ir/	<i>aykh</i> /aχ/
	MASC	<i>er</i> /εr/	<i>em</i> /εm/, /εm/		
	FEM	<i>zi</i> /zi/	<i>zi/ir</i> /zi/, /ir/	<i>zey</i> /zai/	

Table 8: ‘Vus’ paradigm with *mir* and *dir* in both accusative and dative positions

		Singular		Plural	
		NOM	OBJ	NOM	OBJ
1 <sup>st</sup>		<i>ikh</i> /mir/	<i>mir</i> /mir/	<i>undz</i>	
2 <sup>nd</sup>	FAM	<i>du</i> /dir/	<i>dir</i> /dir/	<i>enk</i>	
	HON	<i>ir</i> /aχ/, /arχ/	<i>aykh</i> /aχ/, /arχ/	<i>ir</i> /aχ/, /arχ/	<i>aykh</i> /aχ/, /arχ/
	MASC	<i>er</i> /zi/	<i>em</i> /zi/ir		
	FEM			<i>zey</i>	

appear as *mir* and *dir* respectively in both case positions. These two paradigms reflect different patterns of case syncretism: in the first, the traditionally accusative 1SG and 2SG forms have been extended to the dative as well, while in the second, the traditionally dative 1SG and 2SG forms have been extended to the accusative. This syncretism may have been triggered to some extent by the variation in use of accusative vs. dative forms attested in certain regional pre-War varieties of Yiddish as discussed above (and analyzed in more detail in Wolf (1969: 142–147); see also Jacobs (2005) on the lack of *mikh/dikh* forms in Northeastern Yiddish): given the increased mixing of speakers from different Yiddish dialect areas in the post-War period, it is possible that speakers who acquired the language in the late 1940s and later were exposed to a large amount of variation in these forms and this contributed to a paradigm levelling whereby only one form was selected for both the accusative and dative.<sup>10</sup> However, there do not seem to be any clear patterns governing the use of one pattern (*mikh/dikh* vs. *mir/dir*) for any particular speaker (e.g. age, Hasidic affiliation, gender), except that our Israeli participants all have the *mir/dir* pattern.

Apart from these two distinct patterns for the 1SG and 2SG accusative/dative forms, both paradigms are identical and contain a number of other innovative features. First, the 1PL is *undz* in the nominative, accusative, and dative positions, and as such has no case distinctions. This phenomenon has precedent in pre-War Polish and Hungarian dialects of Yiddish (Weinreich 1964, Jacobs 2005: 70). However, the use of nominative *undz* has spread to younger ‘vos’ speakers, resulting in the innovative nominative form *undz*, which never existed in Lithuanian Yiddish (where the nominative form of the 1PL was exclusively *mir*, with *undz* reserved for the accusative/dative). For example, IO1, a younger ‘Litvish’ speaker, uses nominative *undz*, in contrast to older ‘Litvish’ speaker NO1, who uses the traditional nominal form *mir*. This points to a trend observed elsewhere in the grammar of contemporary Haredi Yiddish whereby the speech of the larger ‘vus’ speaking population has exerted a noticeable influence over that of the smaller ‘vos’-speaking counterparts (see Belk et al. In press).

Second, the 2PL form in both of these paradigms is uniformly *enk*, with no case distinctions (like the 1PL). However, in contrast to the 1PL, the use of *enk* in the nominative position seems to be without precedent in pre-War varieties of Yiddish. In the pre-War Mid- and Southeastern dialects, the 2PL nominative form was *ets* (Jacobs 2005: 70), with *enk* reserved for the accusative and dative. Speakers of Contemporary Hasidic Yiddish, particularly the younger generations

<sup>10</sup>We believe the direction of influence to be from ‘vus’ to ‘vos’ speakers; see Belk et al. (In press) for details.

of ‘vus’ speakers, appear to have lost the traditional nominative *ets* and adopted the accusative/dative form in all positions. This may have evolved on analogy with the earlier use of *undz* in all positions, as well as with the more universal traditional use of the 3PL *zey* in all positions in all forms of Eastern European Yiddish. A small minority of our participants still maintain nominative *ets*, but use it in free variation with nominative *enk*, suggesting that they had acquired the traditional form *ets* from their parents and other older-generation speakers, but had already lost the clear case distinction between *ets* and *enk*, which they can use in all positions.

The 3FS pronoun is also employed differently than in the pre-War and Standard varieties of Yiddish. As shown in Section 1, the traditional 3FS forms were *zi* in the nominative and accusative, and *ir* in the dative. By contrast, in Contemporary Hasidic Yiddish, some speakers employ *zi* only in the nominative, and *ir* in both the accusative and dative, while a much smaller number of speakers employ *zi* in all positions, and do not have *ir* in their feminine pronominal paradigm. As discussed with respect to the 1SG and 2SG accusative and dative forms, both of these innovations may be ascribable in part to the fact that the distribution of *zi* and *ir* in the accusative varied to some extent by region in pre-War Yiddish dialects. This variation may have led to a) a shift from the use of *ir* in the dative only to a single objective form, *ir*, among most speakers, and b) the loss of *ir* and adoption of *zi* in all positions among a small subset of speakers. Whichever pattern our participants employ, none of them follows the pre-War model of *zi* in the nominative and accusative, and *ir* in the dative. This difference is consistent with the rest of the paradigms shown in Tables 7 and 8, in which there are two possibilities with regard to case: a) there is only one form that is used in all case environments, such as *undz*, *enk*, and *zey*, or b) there is one form used for the nominative and a second form used for the objective (accusative/dative), e.g. *ikh* vs. *mikh* and *er* vs. *em*.<sup>11</sup>

Another innovation concerns the emergence of a distinct T/V (familiar vs. honorific) distinction in the plural. To the best of our knowledge, in the pre-War and Standard varieties of Yiddish, there was a T/V distinction in the singular, with the familiar variant *du* in contrast to the honorific variant *ir* (accusative/dative *aykh*). The 2SG honorific variant *ir* was also employed as the generic 2PL form, with no T/V distinction.<sup>12</sup> Our oral questionnaire indicated a reluctance to produce

<sup>11</sup>Note however that ‘vus’ speakers appear to distinguish between nominative and objective weak 3PL pronouns, as discussed further in Section 3.2.

<sup>12</sup>We find no discussion in the literature about a distinct honorific form in varieties that used *ets* in the 2PL. We therefore conclude that, like varieties using *ir* in such contexts, *ets* varieties did not distinguish between the 2PL and the 2HON.

honorific forms in some of the discourse contexts we targeted, either because participants provided a familiar form or because they made use of an alternative honorific strategy such as avoiding direct address. Nevertheless, our data still indicate a shift away from the pre-War Standard system, both in terms of the morphemes currently used as honorific forms, which we give in Tables 7 to 10, and in the realms of usage for the different honorific forms. We discuss the newly emerging T/V system and its use elsewhere (see Belk et al. In prep).<sup>13</sup>

Table 9: Mixed ‘vus’ paradigm

		Singular		Plural		
		NOM	OBJ	NOM	OBJ	
1 <sup>st</sup>		<i>ikh</i>	<i>mikh/mir</i>	<i>undz</i>		
2 <sup>nd</sup>	FAM	<i>du</i>	<i>dikh/dir</i>	<i>enk</i>	<i>aykh</i>	
	HON	<i>ir</i>	<i>aykh</i>	<i>ir</i>		
3 <sup>rd</sup>	MASC	<i>er</i>	<i>em</i>			
	FEM	<i>zi</i>	<i>ir</i>	<i>zey</i>		

A third paradigm is provided in Table 9. This paradigm resembles those shown in Tables 7 and 8 except that there is intraspeaker variation in the 1SG and 2SG accusative/dative form. Instead of employing only a) *mir/dir* or b) *mikh/dikh* in both the accusative and dative positions, speakers who pattern according to Table 9 may employ both *mikh/mir* and *dikh/dir* in free variation in either position. Alternatively, some speakers consistently employ the 1SG form *mikh* in the accusative and dative, but use the 2SG form *dir* in the same positions. The opposite pattern is also attested, whereby a speaker employs the 1SG form *mir* in the accusative and dative, but *dikh* in the same positions. This phenomenon again points to a pronominal system in flux, whereby there is a high degree of flexibility in the selection of particular forms.

The next paradigm, given in Table 10, resembles that shown in Table 8 except that speakers employ a different 2PL pronoun, *aykh* instead of *enk*. This pattern seems to be attested only in the paradigms of speakers from Israel and from London. As in the use of nominative *enk*, the use of nominative *aykh* is innovative because in pre-War and Standard Yiddish this form was solely objective. As

<sup>13</sup>A small number of speakers who use *enk* in the 2PL do not appear to have a nominative/objective distinction in the honorific, and use *aykh* as the honorific pronoun in both the nominative and objective. We are unable to determine a pattern governing this distribution with respect to the rest of the paradigm (e.g. the *mikh/mir* distinction).

## 6 Innovations in the Contemporary Hasidic Yiddish pronominal system

Table 10: ‘Vus’ paradigm with 2PL *aykh*

		Singular	Plural		
		NOM	OBJ	NOM	OBJ
1 <sup>st</sup>		<i>ikh</i>	<i>mir</i>	<i>undz</i>	
2 <sup>nd</sup>	FAM	<i>du</i>	<i>dir</i>	<i>aykh</i>	
	HON	<i>aykh</i>			
3 <sup>rd</sup>	MASC	<i>er</i>	<i>em</i>	<i>ir</i>	<i>zey</i>
	FEM	<i>zi</i>			

with *enk*, it is possible that the Contemporary Hasidic Yiddish usage is based on analogy with other plural forms, such as nominative *undz*, which was attested in the pre-War period, and with *zey*, which has long been employed in all case environments in Yiddish. Some speakers may employ *enk* alongside *aykh*, which suggests that the two originally dialectal forms have been more widely adopted among ‘vus’ speakers in general.<sup>14</sup>

Table 11: ‘Vos’ paradigm with 1PL *mir* and 2PL *ir*

		Singular	Plural		
		NOM	OBJ	NOM	OBJ
1 <sup>st</sup>		<i>ikh</i>	<i>mir</i>	<i>mir</i>	<i>undz</i> /unz/
2 <sup>nd</sup>	FAM	<i>du</i>	<i>dir</i>	<i>ir</i>	<i>aykh</i> /ayχ/
	HON	<i>ir</i>	<i>aykh</i> /ayχ/		
3 <sup>rd</sup>	MASC	<i>er</i>	<i>em, im</i> /ɛm/, /im/	<i>ir</i>	<i>zey</i>
	FEM	<i>zi</i>			

<sup>14</sup>These speakers also used *aykh* in the honorific. This is a relatively rare pattern among our participants so we hesitate to draw strong conclusions. We believe that these speakers do not use *ir* in the nominative at all, but they may distinguish the familiar and honorific paradigms through the use of nominative *ir* in the latter.

The final paradigm is shown in Table 11. This paradigm is attested primarily in the speech of older ‘vos’ speakers (i.e. those over 40). This older ‘vos’ paradigm resembles that of the pre-War Northeastern dialect of Yiddish. It retains the distinction between nominative *mir* and objective *undz* in the 1PL, and the distinction between nominative *ir* and objective *aykh* in the 2PL. In addition, (Jacobs 2005: 184) suggests that pre-War Northeastern Yiddish already lacked the historical three-way distinction in the 1SG, 2SG, and 3FS, employing the historically dative forms in both accusative and dative settings. If true, the 1SG, 2SG, and 3FS two-way distinction shown in Table 11 corresponds to this traditional Northeastern pattern (i.e. using *mir*, *dir*, *ir* in objective cases) rather than being innovative, in contrast to the ‘vus’ paradigms where the same synchronic pattern is innovative.

Interestingly, one ‘vos’ speaker [NO1], who was born in the immediate post-War period, partially follows the pre-War Mideastern Yiddish pattern, employing *mikh* for the 1SG accusative but *mir* for the 1SG dative (though he exhibits the traditional Northeastern syncretism in the 2SG, employing *dir* in both accusative and dative contexts). This *mikh/mir* distinction may actually point to influence from another dialect, such as Contemporary Hasidic Yiddish or a traditional Mid- or Southeastern dialect, but more research would be needed in order to confirm this possibility. The younger ‘vos’ speaker mentioned above (IO1) has a different paradigm, employing *undz* in the nominative instead of the more conservative *mir* and using the innovative nominative *aykh* alongside the older *ir*. Conversely, another younger speaker whose vowel profile corresponds largely to the ‘vos’ model (IM1) provided the 1PL nominative form *mir* alongside the objective form *undz*, but employed the traditionally objective 2PL form *aykh* in nominative contexts.<sup>15</sup>

### 3.1.1 Verbal agreement with novel forms

Verbal agreement with the 1SG, 2SG, 3SG, and 3PL personal pronouns is the same in Contemporary Hasidic Yiddish as in other varieties of the language. However, there are some innovative or otherwise noteworthy features relating to verbal agreement in the 1PL and 2PL. First, speakers who employ the nominative 1PL

<sup>15</sup>There are also a few ‘vus’ speakers, including younger ones, who provide the form *mir* for the 1PL; however, these speakers typically also provide *undz*, rather than employing *mir* exclusively. In such cases, metalinguistic discussion with speakers indicates that speakers consider the *mir* variant to be more literary and higher register (as it is frequently seen in writing but is less common in everyday speech), and as such are more likely to provide it in formal contexts while they may tend to use it less in spontaneous conversation.

pronoun *undz*, especially those who also use the 1SG objective form *mikh*, may use a verbal form ending in *-mir*, e.g. *undz geyenmir* [indz gammir]. This form is historically attested in certain varieties of Mideastern and Southeastern Yiddish (Jacobs 2005: 70, 189), such as those spoken in the Polish and Hungarian regions to which many ‘vus’ speakers trace their ancestry. This form can be used interchangeably with the variant *-n/-en*, e.g. *undz geyen*, which is also the verbal suffix employed with the 1PL pronoun *mir*. In Contemporary Hasidic Yiddish the verbal form ending in *-mir* may be a marker of more informal speech, though this needs further research.

Second, the emergence of the innovative 2PL *enk* and *aykh* in nominative contexts has resulted in the development of a new type of verbal agreement modelled on the 1PL and 3PL agreement, whereby the verbal suffix is *-(e)n*. Thus, we see forms such as *enk geyen* ‘you.PL go’, *enk kumen* ‘you.PL come’, *aykh hobn* ‘you.PL have’. These findings are in keeping with those of Assouline (2007: 86), who discusses the nominative form *aykh* with verbal agreement in *-(e)n*. These forms are to the best of our knowledge unattested in pre-War Polish and Hungarian varieties of Yiddish, in which the 2PL nominative pronoun was *ets* and this was used in conjunction with verbs ending in the suffix *-t*, e.g. *ets hot* ‘you.PL have’ (Jacobs 2005: 190).<sup>16</sup> In our study, participants who employ *ets* in the nominative do so in conjunction with the verbal suffix *-ts*, e.g. *ets geyts* ‘you.PL go’, which is actually based on the plural imperative suffix *-ts* that was historically employed with *ets* in Mideastern Yiddish, e.g. *gayts* ‘go-2PL!’ (Jacobs 2005: 70). In fact the imperative suffix *-ts* continues to be used by speakers even if they use *enk* as the second person plural personal pronoun and V-*(e)n* as the corresponding verb form. In contrast, speakers who use 2PL nominative *ir*, whether as an honorific or as a general pronoun use verbal agreement in *-t*, which is unchanged from pre-War varieties.

### 3.1.2 Reflexive forms

Most of our Contemporary Hasidic Yiddish participants have an invariable reflexive pronoun, *zikh*, which is used for all persons. A minority of participants exhibit some variation in the 1SG and 2SG, using the invariable *zikh* seemingly interchangeably with the objective 1SG and 2SG personal pronouns *mikh* and *dikh*. Even fewer participants invariably employ *mikh* and *dikh* in 1SG and 2SG reflexive

---

<sup>16</sup>An anonymous reviewer points out that the form *ets hots* is attested in pre-War Yiddish. We have found only one such example (Okrutny 1953: 277) in the Yiddish Book Center archive, although it is possible that the form was more common in speech than in writing. A detailed examination of this issue is beyond the scope of this paper.

contexts without ever providing *zikh* in the same contexts. The latter patterns, whereby objective personal pronouns are used as reflexive pronouns, has precedent in a more extensive older Mideastern Yiddish pattern whereby the 1SG, 2SG, 1PL, and 2PL personal accusative or dative pronouns are used as reflexive pronouns, in contrast to other Eastern Yiddish dialects which use *zikh* in all persons and cases, see Jacobs (2005: 184–185).<sup>17</sup> The Contemporary Hasidic Yiddish use is more restricted than its Mideastern Yiddish predecessor as it is limited to the singular, and is rarer than the generalised *zikh* pattern even among ‘vus’ speakers. This tendency towards paradigm levelling of the reflexive pronoun is consistent with the general trend towards syncretism seen elsewhere in the personal pronoun paradigm.

### 3.1.3 Overall trends in the personal pronoun paradigm

The above patterns indicate a trend of increased syncretism in the Contemporary Hasidic Yiddish personal pronoun paradigm vis-à-vis the pre-War and Standard varieties of the language. The most common patterns can be summarised in the abstracted paradigm shown in Table 12.

Table 12: Abstract Contemporary Hasidic Yiddish personal pronoun paradigm

		Singular		Plural	
		NOM	OBJ	NOM	OBJ
1 <sup>st</sup>		NOM	OBJ	SINGLE FORM	
2 <sup>nd</sup>	FAM	NOM	OBJ	SINGLE FORM	
	HON	NOM	OBJ	NOM	OBJ
3 <sup>rd</sup>	MASC	NOM	OBJ	SINGLE FORM	
	FEM	NOM	OBJ		

This typical Contemporary Hasidic Yiddish paradigm reflects syncretism in both the singular and the plural. In the singular, the paradigm is the result of a merger of the previously distinct accusative and dative cases in the 1SG and 2SG, so that instead of a three-way distinction there is now a two-way distinction between nominative and objective (though the objective form varies from speaker

<sup>17</sup>Though see Katz (1987: 125–126) for the observation that the 1PL and 2PL forms were used less consistently than their 1SG and 2SG counterparts.

to speaker, with some employing *mikh/dikh*, others employing *mir/dir*, and still others employing a mix). This may be based on analogy with other historical pronoun forms, e.g. the 3MS (nominative *er* vs. objective *em*), and the 2PL (nominative *ir* or *ets* vs. objective *aykh* or *enk*). This pattern seems to have spread to the 1SG and 2SG, possibly due in part to historically varying distribution of the 1SG and 2SG accusative and dative forms with different verbs. With respect to the 3FS, the development seems to have followed a slightly different route: while, like the 3MS, it had only a two-way distinction, this was between the nominative/accusative *zi* and the dative *ir*. This pattern has not been retained by any of our participants, indicating that the emergence of the two-way distinction between the 1SG and 2SG forms may have contributed to a realignment of the 3FS forms to a nominative/objective distinction, resulting in a regularised singular paradigm with this pattern. In the plural, where there was historically a two-way distinction between the nominative and objective cases in the first and second persons, the objective case was extended to use in nominative contexts as well. As in the singular, this shift may have happened by analogy with the third person plural, which historically had only one form in all case environments (though, as in the 1SG and 2SG, the 2PL form employed may differ from speaker to speaker between *enk* and, more rarely, *aykh*). Thus, in both the singular and the plural, the Contemporary Hasidic Yiddish paradigm reflects one degree of syncretism greater than that found in the pre-War and Standard varieties, having shifted from a three-way distinction in the first and second person singular and a two-way distinction in the 1PL and 2PL to a two-way distinction across the board in the singular and a single form across the board in the plural.<sup>18</sup>

These trends suggest to us that in future, the remaining variations in usage (e.g. the continued optional employment of the pre-War and Standard nominative *mir* and *ets* in the 1PL and 2PL respectively) will decrease (a tendency supported by the fact that our participants under the age of 40 are much less likely to supply them), resulting in a more stable levelled paradigm with a clear two-way nominative/objective distinction in the singular and a single form in the plural. The only likely exception to this is the possible retention of the 1PL form *mir* in nominative contexts, which may continue to persist for longer due to the fact that it is frequently found in writing and as such may be preserved as a higher-register form in speech as well. This is reinforced by the fact that *mir* is the only exclusively nominative option for the 1PL, in contrast to the 2PL, for which two different traditional nominative variants exist (*ir* and *ets*), a situation which may have contributed to the increased instability of the nominative 2PL forms.

---

<sup>18</sup>Although such forms can be difficult to elicit, at least some speakers retain a nominative/objective distinction in the 2PL.HON.

### 3.2 Weak vs. strong pronouns

Yiddish personal pronouns typically occupy different syntactic positions than corresponding full noun phrases. The phenomenon, which resembles Scandinavian object shift, is known to have existed already in pre-War and Standard Yiddish. As (1) shows, Yiddish personal pronouns typically appear right-adjacent to the auxiliary, which is in the second position.<sup>19,20</sup>

- (1) 'kh vil <enk> RUFN \*<enk> ven supper iz GREYT.<sup>21</sup>  
I will <you.PL> call <you.PL> when dinner is ready  
'I will call you when dinner is ready.'

More than one pronominal can occur in this position; in this case, the pronouns form a phonological unit, or cluster, with the preceding auxiliary which can also include the subject pronoun in the initial position. In fast speech, such clusters can be phonetically substantially non-transparent:

- (2) a. er hot es ir nisht gegebn  
[ɛtəsi]  
he has her not given  
'He did not give it to her.'  
b. er hot dikh nisht gezen  
[ɛdəχ]  
he has you not seen  
'He did not see you.'

<sup>19</sup>Weak pronouns can occur in the neutral clause final position if they are inside a prepositional phrase, as shown in (i). In this case main stress falls on the participle, as indicated. As Ruys (2008) observes, the fact that PPs are differently affected by pronoun shift suggests a motivation for pronoun shifting that is related to case assignment, as suggested by Neeleman & Reinhart (1998). We will not be concerned with motivating pronoun shift in Yiddish, but we note that in this respect the Yiddish data seems to pattern with Dutch.

(i) du host es geshikt [tsə i].  
You have-2SG it sent to her  
'You sent it to her.'

<sup>20</sup>Grammaticality judgments in this section were provided by a young 'vus' speaker from Stamford Hill. He has a typical mixed paradigm using both *mikh/dikh* and *mir/dir*. A nominative-objective contrast is also present in the 3fs (*zi* vs *ir*, pronounced [i:] as this speaker has deletion of /r/ in coda position, which is typical of Stamford Hill Yiddish). He uses *enk* in 2PL.

<sup>21</sup>In transliterated examples and English glosses and translations, we indicate primary stress/focus with small capitals. In IPA transcriptions we indicate it with the primary stress symbol, '.

Although this phenomenon is a distinctive characteristic of spoken Yiddish, it has not yet, to the best of our knowledge, been subject to linguistic analysis. We would like to propose that in Contemporary Hasidic Yiddish, and most likely also in pre-War spoken dialects, personal pronouns can be divided into strong and weak categories in the sense of Cardinaletti & Starke (1999). Strong pronouns can be used deictically or accompanied by a pointing gesture. As they identify their own referent, they can also be easily contrasted. Weak pronouns are referentially deficient. They are always dependent on the immediate discourse context for reference resolution. They rely on the accessibility of a given referent for reference assignment. A summary table of all the forms used by our participants is given in Table 13. Recall that some of the speakers use *mikh/dikh* forms while others use *mir/dir* forms, and some use both interchangeably. Correspondingly, the speakers would use the weak variant of the forms they use as strong forms, which in Table 13 appear in the same line. Strong pronouns are stressed, while weak are unstressed.<sup>22</sup>

Table 13: Strong and weak pronominal forms used by ‘vus’ speakers

		NOM		OBJ	
		Strong	Weak	Strong	Weak
SG	1 <sup>st</sup>	[iχ]	[ix, εx, χ]	[mix, 'mir]	[mix, məx, mir, mi]
	2 <sup>nd</sup> FAM	[di]	[dɪ, -i, -ε]	[dix, 'dir]	[dix, dəx, dir, di]
	HON	[ir, i]	[ir, i, i]	[aix, 'a:χ]	[aix, a:x, aχ]
	3 <sup>rd</sup> MASC	[ər]	[ər, ε]	[əim, 'em]	[əim, em, əm, m]
	FEM	[zi]	[zi, z]	[zi, 'ir, 'i]	[zi, ir, i, i]
	1 <sup>st</sup>	[ints]	[nts]	[ints]	[nts]
PL	2 <sup>nd</sup> FAM	[ɛŋk]	[ɛŋk]	[ɛŋk]	[ɛŋk]
	HON	[ir, i]	[ir, i, i]	[aix, 'a:χ]	[aix, a:x, aχ]
	3 <sup>rd</sup>	[zai]	[zai, za]	[zai]	[zai, ze]

<sup>22</sup>We abstract away from specific length and quality differences in the high front vowels, using /i/ for long and/or tense vowels and /ɪ/ for short and/or lax vowels. The specific realisation of these vowels depends to a certain extent on the vowel inventory of the co-territorial language(s) of which the speaker has command. Where such languages do not have a tense/lax distinction, the speaker’s Yiddish vowel inventory will also lack this distinction.

Table 14: Strong and weak pronominal forms used by ‘vos’ speakers

		NOM		OBJ	
		Strong	Weak	Strong	Weak
SG	1 <sup>st</sup>	[iχ]	[ix, χ]	['mir]	[mir, mi]
	2 <sup>nd</sup> FAM	['du]	[du]	['dir]	[dir, di]
	HON	['ir, i]	[ir, i, i]	['aix]	[aix]
	3 <sup>rd</sup> MASC	['er]	[er, ε]	['εim, 'em, 'im]	[εim, em, əm, m]
	FEM	['zi]	[zi, z]	['ir]	[ir]
	1 <sup>st</sup>	['mir]	[mir, mir]	['unts, 'undz]	[unts, undz]
PL	2 <sup>nd</sup>	['ir, i]	[ir, i, i]	['aix]	[aix]
	3 <sup>rd</sup>	['zεɪ]	[zεɪ, ze]	['zεɪ]	[zεɪ, ze]

Two points of interest emerge from these paradigms. The first is that there is a distinction between the weak paradigms of ‘vos’ and ‘vus’ speakers. For the former group of speakers, the weak 3PL is /zε/ in both nominative and objective contexts, while the latter group have a case distinction that is not present in the strong form, where the weak 3PL pronoun is /za/ in the nominative and /zε/ in the objective.<sup>23</sup>

Some examples from our elicited data are given in (3), which were provided by a typical young ‘vus’ speaker, MU1. In all of these examples the context provides an appropriate environment for a weak pronoun, as the referent is established in the previous sentence. MU1 produces clear examples of /i/ as the weak form of the 3FS objective pronoun. In (3b) the weak pronominal form [zε] is used, while in (3c) the participant uses the weak second person objective pronoun [dəχ] which clearly contrasts with the form [diχ].

- (3) a. rokhl iz a gute meydl. ikh hob ir lib.

[iχɔbi 'lib]

Rokhl is a good girl I have her love

‘Rokhl is a nice girl. I like her.’

<sup>23</sup>The observations as they relate to ‘vos’ speakers are based on a much smaller subset of speakers and so must be considered somewhat provisional.

- b. rokhl un sheyni zenen gute meydlekh. ikh ze zey a SAKH in the  
 Rokhl and Sheyni are good girls I see them a lot in the  
 STREET.

street

'Rokhl and Sheyni are nice girls. I see them often in the street.'

- c. de lerer hot dikh LIB. er hot dikh zayer shtark lib.  
 the teacher has you love he has you very strong love  
 'The teacher likes you. He likes you very much.'

MU1 uses strong pronouns in other environments. This typically occurs when the pronoun appears at the end of the sentence, as in (4a) and (4b) and in sentences that she utters with slow tempo and with comma intonation like (4c).

- (4) a. ikh geb es far DIR  
 [ɛχ,geb̥es fa'dir]  
 I give it to you
- b. enk gebn es far ZEY  
 [ɛŋk 'gəbn̥es fa'zai]  
 You.PL give it to them
- c. zi get em de PAPER  
 ['zi 'get 'eim də'pʰeipər]  
 She, gives, him, the paper

MU1's data clearly illustrates that the same speaker may use different forms in different discourse situations and in different syntactic environments. This supports our claim that Contemporary Hasidic Yiddish distinguishes between separate weak and strong pronominal forms.

Cardinaletti & Starke (1999) identify a host of syntactic tests to distinguish strong from weak pronouns. One such characteristic difference between the two is that weak pronouns typically have a different syntactic distribution to corresponding full noun phrases, while strong pronouns behave syntactically like full nominals. As the data in (5)–(7) show, the Yiddish personal pronouns we identified as weak pattern with Cardinaletti and Starke's weak pronouns in all relevant respects. In addition to not being allowed in the clause final position, they also cannot occupy the clause initial position, which is restricted to contrastively stressed strong pronouns.

- (5) [ɛɪm]/\*[(ə)m] hob ikh nisht gezen, ober [i:]/\*[i], yo.  
him have I not seen but her yes  
'HIM I didn't see, but HER, yes.'

In (6a), we illustrate that only strong pronouns can be coordinated. (6b) shows that only strong pronouns can be fragment answers.

- (6) a. <MIKH un DIKH> hobn zey shoyn gezen <\*MIKH un DIKH>  
<me and you> have they already seen <me and you>  
'They have already seen me and you.'
- b. Q: vemen kenstu hern?  
who can.2SG hear  
'Who do you hear?'  
A: DIKH / IR / EM / \*dəkh / \*dkh / \*i / \*əm / \*m  
you / her / him / you / you / you / him / him  
'YOU/HER/HIM/\*you/\*her/\*him.'

(7) illustrates that only strong pronouns can be modified.

- (7) a. ikh hob <\*nor 'm> gezen <nor EM>.  
I have <\*only him> seen <only him>  
'I only saw him.'
- b. \*Ikh hob 'm aleyn gezen.  
I have him alone seen
- c. EM aleyn hob ikh gezen.  
him alone have I seen  
'I saw only him.'

The same strong-weak contrast that is prevalent among 'vus' speakers can be also be observed in the speech of older 'vos' speakers.<sup>24</sup> In (8a), provided by NO1, illustrates this point the 3MS objective pronoun, with (8b) exemplifying weak forms, and (8c) providing the strong form. Similarly, (9a) illustrates the weak 3PL objective pronoun and (9b) its strong variant.

<sup>24</sup>Note that three among our four 'vos' speakers displayed a strong-weak distinction. Whether the fourth person does not have a strong-weak distinction or whether perhaps their generally slow speech tempo and formal attitude towards the test situation prevented them from using a faster speech tempo or a more colloquial register is not clear.

- (8) a. dovid iz a fayner mentsh. ikh glaykh im a SAKH. ikh hob im

[χ,glaiχəm a'sax χ,hɔbəm  
Dovid is a nice guy I like him a lot I have him  
LIB a sakh.

libasax]

love a lot

‘Dovid is a nice guy. I like him. I like him a lot.’

- b. zi git EM... de PAPIR.

[zi.git 'em... də pa'pir]  
she gives him the paper

‘She gives him... the paper.’

- c. du, gib es tsu EM

[du,gibətsə 'em]  
you give.IMP2SG it to him

‘(Hey) you, give it to him.’

- (9) a. rokhl un sheyni zenen fayne meydlekh. ikh ze zey zeyer oft

[iχ,zεzε zejer ,oft  
Rokhl and Sheyni are nice girls I see them very often

oyf de GAS

[ɔfdə'gas]

on the street

‘Rokhl and Sheyni are nice girls. I see them very often on the street.’

- b. gib es tsu ZEY

[gibes tsə'zeɪ]  
give.IMP2SG it to them

‘Give it to them.’

In sum, Contemporary Hasidic Yiddish has a set of corresponding strong and weak pronominal forms with distinct syntactic distributions. Strong pronominal forms can be coordinated or modified and can be used as fragment answers. They can be fronted, and in fact are ideally fronted to a clause-initial position. Weak pronominal forms, in contrast, cannot be coordinated, modified, or used as fragment answers. In subject-initial clauses, weak pronouns follow the finite auxiliary (or verb) in the V2 position. They form a phonological word with the auxiliary, which can also include the subject, if that is also a weak pronoun. One question for future research is whether weak pronominals are actually clitics

fitting into a predetermined order or morphological template.<sup>25</sup> It is not clear whether the distinction between strong and weak pronouns constitutes an innovation in the pronominal system due to the lack of literature on the issue, but the distinction clearly holds in Contemporary Hasidic Yiddish.

## 4 Possessives

Like the personal pronouns, possessive pronouns in Contemporary Hasidic Yiddish also exhibit various innovations. Parts of the pre-War possessive system are maintained in Contemporary Hasidic Yiddish, such as the distinction between singular and plural forms for dependent possessives. At the same time, certain pre-War and Standard case and gender morphemes have been reanalyzed in this dialect, as is the case in independent possessive pronouns. Compared to the relative variability found in the personal pronoun system, the possessive pronouns are noticeably more stable and exhibit a number of innovations as well as retaining a number of characteristics of the pre-War system.

### 4.1 Dependent possessive pronouns

Dependent possessive pronouns in Hasidic Yiddish are very similar to those of pre-War and Standard varieties of Yiddish as shown in Table 15.

This paradigm exhibits several noteworthy characteristics. The possessive stem *enker-*, which is attested in the 2PL in some pre-War varieties, is nearly universal among our ‘vus’ speaking participants. This is in line with our findings in the personal pronoun system, which demonstrated that a novel distinction has emerged between 2PL and 2HON: *enker-* is used for the former and *ayer-* for the latter. ‘Vos’ speakers typically retain the more conservative pattern of using a single form for both 2PL and 2HON, *ayer-*.

In pre-War and Standard Yiddish, dependent adnominals agree with the noun in number only. This pattern largely survives in Contemporary Hasidic Yiddish. Forms with a Ø-ending are used with singular possessa and the -e ending is used with plural possessa, as demonstrated in (10).

<sup>25</sup> Another issue that could be considered further concerns empty pronominals, or *pro-drop*. As in many Germanic dialects, the 2SG nominative pronoun is often dropped in Contemporary Hasidic Yiddish, especially in questions.

- (i) (du) host epes?  
have.2SG-2SG.*pro* something  
'Do you have anything?'

Table 15: Possessive pronoun stems in Contemporary Hasidic Yiddish

		Singular	Plural
1 <sup>st</sup>		<i>mayn-</i> /ma:n/, /mam/	<i>undzer-</i> /inzer/, /unzer/
2 <sup>nd</sup>	FAM	<i>dayn-</i> /da:n/, /dam/	<i>enker- (ayer-)</i> /ɛŋker/ (/ajer/)
	HON	<i>ayer-</i> /ajer/	<i>ayer-</i> /ajer/
3 <sup>rd</sup>	MASC	<i>zayn-</i> /za:n/, /zam/	<i>zeyer-</i> /zajer/, /zejer/
	FEM	<i>ir-</i> /ir/	

- (10) a. undzer tish vs. undzere tishn  
       our     table vs. our-E    tables  
     b. dayn feder vs. dayne feders  
       your pen vs. your-E pens  
     c. ir bukh vs. ire bikher  
       her book vs. her-E books

This pattern is especially strong with singular possessors; the plural possessors, *undzer*, *enker*, *ayer*, *zeyer*, show somewhat more variation with more of them appearing with a Ø-suffix with plural possessa than expected. There may be a phonological explanation for this phenomenon: some participants, especially those from Stamford Hill and certain communities in the New York area, often delete /r/ in syllable codas, and reduce /r/ between two unstressed syllables especially in fast speech. These factors mean that the difference between *undzer* and *undzere* is often difficult to perceive.<sup>26</sup> Nonetheless, there appears to be a strong overall tendency to use dependent possessive pronouns with a Ø-ending

<sup>26</sup>The results of our questionnaire revealed an unexpected difference in the behaviour of singular vs. plural possessors, and we therefore made use of the Haredi Yiddish-medium online forum Shtiebel (n.d.). Using Google, we searched the forum for relevant written examples. A small number of examples of the form *ענדזֶה* *undze* are attested on this forum with both singular and plural posessa, including *undze kehile* ‘our community’, *undze kop* ‘our head’, *undze menahel* ‘our director’ along with *undze nemen* ‘our names’, *undze gvirim* ‘our rich and influential people’. However, three linguistic consultants all rejected these forms.

for singular possessa and dependent possessive pronouns with a *e*-ending with plural possessa.

The fact that the singular/plural distinction survives in Contemporary Hasidic Yiddish is surprising given the fact that the corresponding agreement in attributive adjectives is no longer productive. In pre-War and Standard Yiddish, attributive adjectives agreed with the noun for case, gender, and number, while in Contemporary Hasidic Yiddish the pre-War agreement morpheme *-e* has been reanalyzed as a marker of attribution and is applied to all attributive adjectives regardless of number, case, or gender (Belk et al. 2020, In press; see also Krogh 2012: 489–496 and Assouline 2014: 42 for a slightly different view). Thus, unlike in attributive adjectives, dependent possessives retain a number distinction.<sup>27</sup>

#### 4.2 Independent possessive pronouns

In independent possessive pronouns, we find two competing sets of forms: one with an *-e* suffix, and the other with a *-s* suffix, as demonstrated in Table 16.<sup>28</sup> This pattern is in contrast to the situation in pre-War and Standard Yiddish, where independent possessive pronouns inflected according to gender and number: *-er* for MS, *-e* for FS, *-s* for NS; and *-e* for plural (Jacobs 2005: 183–184). In contemporary Hasidic Yiddish *-e* forms (historically FEM SG or PL forms) and *-s* forms (historically Neut SG forms) used regardless of noun gender and number and represent two competing realisations of independent possessive pronouns. To the best of our knowledge, this constitutes an innovative pattern.

Both forms are used regardless of case, gender, and number. Some speakers make use of both forms: (12) and (15) were both produced by the same participant, IU1, while the other examples were produced by IU2 (11), NO1 (13), NU1 (14), and MU1 (16).

- (11) de pene iz mayne  
the pen is mine

- (13) di pens zenen zayne  
DEM pens are his

- (12) er iz dayne  
it (lit: he) is yours

- (14) deye tish iz mayns  
DEM table is mine

<sup>27</sup>We did not expect that our questionnaire would provide these results, and we therefore felt that we did not have enough data for certain informative forms. We therefore searched Shtiebel (n.d.) for possessive pronouns with all possible endings. We checked approximately 250 of the results, which conformed to our findings.

<sup>28</sup>The *-e* suffix is pronounced [ɛ] or [ə], while the *-s* suffix is /s/. After a liquid, /s/ often surfaces as [ts] due to a process of /t/ insertion.

Table 16: Independent possessive pronouns in Contemporary Hasidic Yiddish

		E-forms		S-forms	
		Singular	Plural	Singular	Plural
1 <sup>st</sup>		<i>mayne</i>	<i>undzere</i>	<i>mayns</i>	<i>undzers</i>
2 <sup>nd</sup>	FAM	<i>dayne</i>	<i>enkere (ayere)</i>	<i>dayns</i>	<i>enkers (ayers)</i>
	HON	<i>ayere</i>	<i>ayere</i>	<i>ayers</i>	<i>ayers</i>
3 <sup>rd</sup>	MASC	<i>zayne</i>	<i>zeyere</i>	<i>zayns</i>	
	FEM	<i>ire</i>		<i>irs</i>	<i>zeyers</i>

- (15) yene feders zenen nisht dayns      (16) deye bukh iz nisht zayns  
DEM pens are not yours                            DEM book is not his

'Vos' speakers and Israeli speakers from various groups typically prefer -*e* forms, while speakers in the New York and Montreal areas, especially Satmar Hasidim, tend to prefer -*s* forms. However, two individual exceptions prove instructive. IO1 is a 'vos' speaker who now speaks Yiddish mostly with 'vus' speakers. She notes that she used to use -*e* forms of independent possessives, but now that she often speaks to 'vus' speakers, she is now more likely to use -*s* forms than previously. Contrastingly, NO2, an older 'vos' speaker, follows a slightly different pattern, using -*er* forms for singular nouns and -*e* forms for plural nouns. Nevertheless, while tendencies towards one form or another can be observed at a group level, on an individual level most participants produce both forms in seemingly free variation.

Independent possessive pronouns with  $\emptyset$ -endings are rare in our dataset, which is consistent with the findings of (Mark 1978: 242–243) for pre-War Yiddish. However, the phonological factors discussed in Section 4.1 play a role here as well: in some cases it is difficult to determine whether a participant produces e.g. *undzer* (which might be pronounced /unze/) or *undzere* (which might be pronounced /unze:/) and thus we cannot rule out the existence of  $\emptyset$ -forms in stems ending in /r/.

We have also recorded a specific variant for the 3FS possessive pronoun, which has not, to the best of our knowledge, been documented before. The form *zire* (her/hers, parallel to *ire*), can be used as a dependent possessive pronoun, e.g. *zire kinder* ‘her children’. As an independent possessive pronoun it is only attested in our dataset after a 3SG auxiliary verb, e.g. *dos iz zire hoyz* ‘it is her house’, but not with a plural verb form. When asked about this form, all three of our consultants

recognised it, with one reporting that “*zire* for ‘her’ or for ‘hers’ is very frequent I’d say”. Another consultant notes that the form is used in speech, but not in writing: “We would sometimes use *zire* and not *ire*. I would definitely write *ire*, but sometimes you’d say *zire*”. While this form is much less widespread than *ire*, we have observed it interactions with native Contemporary Hasidic Yiddish speakers in the New York area, Israel, Stamford Hill, and the Montreal area.

#### 4.3 Overall trends in possessives

Possessive pronouns in Contemporary Hasidic Yiddish exhibit distinct dependent and independent morphological patterns. Dependent possessive pronouns have two variants: one (with an Ø-ending) appears with singular possessa while the other (with an *e*-ending) appears with plural possessa. This pattern, while not innovative as it existed in pre-War and Standard varieties of Yiddish, is nonetheless unexpected as it does not follow trends seen in attributive adjective and definite determiner morphology towards a single uninflected form. Independent possessive pronouns appear with one of two suffixes, *-s* or *-e*. A subset of speakers, particularly those from Israel and speakers of the ‘*vos*’ variety, uses only one suffix, *-e*, for all independent possessive pronoun stems, regardless of case, gender, or number features. The remaining speakers, who constitute a majority, use stems with the *-s* suffix in free variation with stems with the *-e* suffix. Both patterns represent a departure from pre-War and Standard varieties of Yiddish, in which independent possessives were inflected for case, gender, and number. Thus, the traditional case and gender morphology has been reanalyzed as markers of a distinct syntactic role.

Certain questions remain to be answered. The first concerns the use of a specific possessive-indefinite construction. This construction, where an inflected form of the possessive pronoun is followed by the indefinite article and associated noun, existed in pre-War Yiddish (Katz 1987: 109). We have found written evidence (on Shtiebel n.d.) that this construction is also used in Contemporary Hasidic Yiddish. These preliminary data suggest that, for singular possessives, the *-er* ending is used to mark this construction: *mayner a bakanter* ‘an acquaintance of mine’, *mayner a fraynd* ‘a friend of mine’ regardless of case, gender, and number: *men zingt zayner a nigen* ‘someone sings a *nigun* (traditional wordless melody) of his’, *mit mayner a noenter yedid* ‘with a close friend of mine’. However, as in pre-War Yiddish (Mazin 1927: 27–28) Ø-ending forms are used for plural possessors: *fun undzer a tayerer khaver* ‘from a dear friend of ours’.

The second question concerns dependent possessive pronouns in postposition, which preliminary written data suggest exhibit a strong tendency to use the *-er* ending regardless of the historical gender of a particular noun: *a khaver mayner*

‘my friend’, *mit a khaverte irer* ‘with her female friend’, *dos harts undzerer* ‘our heart’, *a gantse toyre enkerer* ‘all your theory’, *der mayse enkerer* ‘your story’, *di gefilen mayner* ‘my feelings’. Morphological case also appears to have no bearing on the morphology of the possessive in this construction, which is consistent with the rest of Contemporary Hasidic grammar. However, on rare occasions some speakers use a distinct -e ending for plural nouns: *di zikhroynes mayne* ‘my memories’, *di kinder mayne* ‘my children’. The tendency suggests the emergence of a new grammatical meaning for the -er ending in possessive pronouns, namely a general marker of an attributive possessive pronoun in postposition for any noun regardless case, gender, and number.

## 5 Demonstratives

Standard and pre-War varieties of Yiddish did not have a dedicated proximal determiner: either the stressed definite article or a deictic form such as *ot* or *dozik-* (both roughly meaning ‘this here’), or a combination of the two, could be used where English uses *this*. The status of distal demonstratives is somewhat less clear. Jacobs (2005: 186) claims that the inflected root *yen-* is used. However, Katz (1987) claims that in contexts where English would use *that*, the same approaches could be used as with proximal determiners, but that there also existed a distal demonstrative *yener/-e/-em*, which he says can be used neutrally or associated with aggression, ‘otherness’ or derogatory connotations.

Discussing contemporary varieties of Haredi Yiddish, both Assouline (2014: 58) and Sadock & Masor (2018: 95) describe a stressed form of the definite determiner, pronounced *dey/dei* or *deye*, which is used as a demonstrative. They link this form to a Hungarian Yiddish pronunciation of the feminine and plural definite determiner *di* as this variety often diphthongises stressed vowels, and they claim that it is absent in Polish Yiddish. Similarly, Krogh (2012) reports *des* and *deys* as variants of the neuter pronouns *es* and *dus/dos*.

In the oral translation task, participants were asked to translate sentences containing a variety of pronominal forms from either English or Modern Hebrew, according to their preference. These sentences included independent and dependent proximal and distal determiners in a range of case and gender contexts.<sup>29</sup>

<sup>29</sup>The questionnaire was originally composed in English and translated by a native speaker into Modern Hebrew. As the proximal/distal distinction in Modern Hebrew does not map on to that of English, we expect some differences in the choice of demonstrative stem between participants translating the English version of the questionnaire and those using the Modern Hebrew version. This issue is discussed further in Section 5.2.

Additionally, speakers produced spontaneous metalinguistic discussion, which was recorded and analyzed, complementing the elicited translation data. Most speakers produced between 40 and 50 demonstrative tokens. Our results indicate a different pattern than that of either Standard and pre-War varieties of Yiddish or previous descriptions of contemporary varieties of Haredi Yiddish. Specifically, we find innovations in demonstrative stems, in the distribution of what Jacobs (2005) refers to as proximant and distant demonstratives, and on the inflection associated with them. We discuss these issues in turn.

### 5.1 Demonstrative stems

All participants distinguish between two forms of the demonstrative, one beginning with /j/ and one with one beginning with /d/. We will refer to these as *y*-stems and *d*-stems, respectively. The *y*-stem for all participants, regardless of age, geographical origin or Hasidic affiliation, was *yen-*, corresponding to the dedicated distal demonstrative described by Katz (1987) and Jacobs (2005). However, much more variation is found in the *d*-stem.

All ‘vos’ speakers follow the traditional pattern of using the stressed definite determiner as a demonstrative, although younger ‘vos’ speakers may use another *d*-stem in addition to this option. The stressed definite determiner often co-occurs with elements such as *ot* or *dozik-*, as is the case in Standard and historical varieties. While these speakers use a stressed definite determiner in both independent and dependent contexts, they distinguish between these two contexts in the form of the determiner. They overwhelmingly use *dos* as an independent determiner (for some speakers, even in non-nominative contexts) in both the plural and singular, and prefer *di* as a dependent determiner, although *der* and *dem* are also found. This pattern suggests that, for ‘vos’ speakers, *dos* is emerging as a distinct, independent demonstrative, while the older strategy of using a stressed definite determiner persists for dependent demonstratives.

In dependent contexts, some ‘vos’ speakers use a single invariant definite determiner, while others use a variety of forms (i.e. *der*, *di*, *dos*, *dem*). The form of the determiner is not determined by the case or gender of the DP, as evidenced by the appearance of non-Standard-like usages such as *dem* in the nominative, or *der* in the plural. These findings are consistent with those of Belk et al. (In press).

‘Vus’ speakers show somewhat more variety in their choice of *d*-forms. In dependent contexts, all speakers use the stem *dey-*, which was described by Krogh (2012), Assouline (2014), Sadock & Masor (2018), although there are a small number of tokens of the more conservative stressed *di* (none use a stressed deter-

miner form other than *di*). In independent contexts, *dey-* is found alongside *dos* (pronounced /dɔs/) and even *diye*. There are no instances of ‘vus’ speakers using a stressed definite determiner other than *dos* as an independent demonstrative. Some participants exclusively use the stem *dey-* in independent contexts while others use a mix of the two (or three) stems, but no participant uses *dos* or *diye* exclusively. For those that mix stems, *dos* is never found outside of nominative contexts, but it is found alongside other *d*-stems in the nominative. The same pattern is found in singular as in plural nouns: *dos* is regularly used and, for some speakers, preferred in nominative contexts and *dey-* is used elsewhere.

Beyond the nominative/objective distinction and the ‘vos’/‘vus’ distinction, there appears to be no pattern to when participants use particular *d*-stems, or to which participants are likely to use particular forms. *Dey-* appears to be the preferred stem for most ‘vus’ speakers regardless of age, gender, Hasidic affiliation, or geographical origin, but these factors do not appear to influence whether a speaker makes use of other *d*-stems and, if so, which they use. ‘Vos’ speakers largely use the traditional strategy of using a stressed definite determiner as a *d*-stem, although younger speakers may be beginning to adopt the *dey-* stem.

## 5.2 Distribution of *d*- and *y*-stems

Regardless of which particular *d*-stems are used, all participants make use of both *d*- and *y*-stems. However, the distribution of these stems does not appear to be entirely determined by the proximal/distal distinction. Before discussing the results of this task, a note on the proximal/distal distinction in English and in Modern Hebrew is in order.

English demonstratives are proximal (*this, these*), i.e. close to the speaker, or distal (*that, those*), i.e. further from the speaker, although the border between the two is obviously subjective (see e.g. Stirling & Huddleston (2002) for further discussion). In Modern Hebrew, demonstratives are similarly categorised as either proximal ((*ha*)ze, (*ha*)ele) or distal ((*ha*)hu, (*ha*)hi, (*ha*)hem), but their distribution differs from that of English. The proximal demonstratives are employed much more frequently than their distal counterparts (Halevy 2013), with the latter often reserved for contrastive contexts. In such cases, the proximal demonstrative is used to denote a referent which the speaker views as emotionally close or relatable, with the demonstrative serving to denote a referent which the speaker regards as ‘remote or adversative’ (Halevy 2013). The Modern Hebrew demonstratives thus differ in function from those of English, as the Hebrew ‘proximal’ demonstratives are used as a default form, often irrespective of spatial deixis, while the ‘distal’ demonstratives are frequently restricted to specifically

contrastive settings. Jacobs' (2005) categorisation of demonstratives in Yiddish is thus much closer to that of English, whereas Katz's (1987) is perhaps closer to Hebrew, although he does not specifically mention the notion of contrast.

Participants translated from either an English or a Modern Hebrew version of the same questionnaire, with the Modern Hebrew utilising the distal (and less frequent) demonstratives (*ha*)*hu*, (*ha*)*hi*, (*ha*)*hem* where the English version used *that*, *those*. Given the differing distribution of demonstratives in these two languages, we might expect an effect of the language of the questionnaire on participants' responses.

Regardless of the language they were translating from, participants almost never use *y*-stems to translate English *this* or *these*, or Modern Hebrew (*ha*)*ze* or (*ha*)*ele*. This suggests that such stems cannot act as proximal demonstratives, which is consistent with both Jacobs (2005) and Katz (1987). However, like their proximal counterparts, English *that* and *those* are also usually translated with *d*-stems, contra Jacobs (2005). Participants using the Modern Hebrew version of the questionnaire were most likely to translate (*ha*)*hu*, (*ha*)*hi*, (*ha*)*hem* using *y*-stems, although even they did not do so consistently. This suggests that the Contemporary Hasidic Yiddish proximal/distal distinction does not map directly onto either the English system or the Modern Hebrew system.

The only factor that is consistently associated with the use of *y*-stems is contrast. All speakers translated at least some distal/proximal contrastive pairs using a *d*-stem and a *y*-stem, and only a very few produced other types of pairs.<sup>30</sup> We therefore suggest that *y*-stems are primarily a marker of contrast, rather than distance from the speaker. This conclusion is supported by metalinguistic discussion with one speaker in particular who, unprompted, proposed after completing the questionnaire that they "would only use *yene* when I mean *yene* and not *deye*".

### 5.3 Demonstrative morphology

The final innovation in the demonstrative system we will discuss is the inflection found on independent and dependent demonstratives. In Standard and pre-War varieties, all determiners (i.e. the stressed definite article and the stem *yen-*) were inflected for case, gender and number. However, in Contemporary Hasidic Yiddish, the role of morphology is less straightforward.

<sup>30</sup>Other contrastive pairs produced by our participants include two *d*-stems (fewer than five tokens overall), two *y*-stems (one token overall) and an unstressed determiner plus a *y*-stem, as well as forms such as *de andere* 'the other' and *di other* 'the other'.

For almost all speakers, dependent *yen-* appears exclusively with the ending *-e*, leading to the existence of a single invariant *y*-form, *yene*. In line with developments in the case and gender system on full nominals discussed in Belk et al. (2020, In press), this form is not inflected for case, gender or number when used as a dependent demonstrative. A small number of the ‘vos’ speakers produce inflected forms of *yen-*, such as *yener* and *yem* in dependent contexts, although such forms do not appear to be determined by nominal case or gender. Again, this is in line with the findings of Belk et al. (In press), who suggest that the loss of case and gender marking on full nominals in Contemporary Hasidic Yiddish happened somewhat later in ‘vos’ speakers than it did in ‘vus’ speakers, and that more vestiges of this system can therefore be found among the former group.

As discussed in Section 5.1, ‘vos’ speakers use a stressed definite determiner as the *d*-stem in dependent contexts. However, a significant minority of ‘vos’ speakers’ translations of English and Hebrew demonstratives made use of an unstressed definite determiner. In at least some of these cases, a demonstrative reading was clearly intended (i.e. where the participant pronounced the target sentence aloud in English with a demonstrative before translating it to Yiddish), but in other cases this is less clear. We leave this issue aside here.

For ‘vus’ speakers, *dey-* appears in dependent contexts with either a *-e* suffix or with a  $\emptyset$  ending. Due to diphthong smoothing, the distinction between the two can sometimes be difficult to perceive, but clear examples of both *dey* and *deye* in dependent contexts can be identified. Some speakers only use the form *deye*, but no speakers use only *dey*. For those speakers who do use *dey*, this form appears to be in free variation with *deye* as both are produced in all case and gender contexts for both singulars and plurals and for English and Modern Hebrew proximal and distal determiners.

Despite the existence of the novel form *dey*, it is striking that speakers do not use the unaffixed root \**yen*. One could easily imagine that the dependent demonstrative system could develop by analogy with the possessive system, where a distinction between singular and plural dependent possessives existed in Standard and pre-War varieties and continues to exist in large part in Contemporary Hasidic Yiddish. In the possessive system, singular nominals usually appear with an unaffixed form of the possessive, while plural nouns appear with either an unaffixed possessive or a possessive with the *-e* suffix. Such a system is not evident in the dependent demonstratives, where no bare form of the *y*-stem exists, and the bare and suffixed forms of the *d*-stem appear in free variation. While it is not impossible that these two systems will develop in such a way as to become more similar, we do not see any evidence of such a development in this study.

In the independent demonstratives a somewhat different pattern emerges. Some speakers use the -e-suffixed forms of both *dey* and *yēn* as invariant independent demonstratives *deye* and *yene*. These speakers typically use the same invariant forms in dependent contexts.

A much larger group of speakers distinguish between nominative and objective forms of the independent demonstratives. For these speakers, the objective form appears with a -e suffix, while the nominative ends with /s/. Thus, the y-stem appears as *yens* in the nominative and *yene* in the objective. However, the d-stem is somewhat different: the nominative can use either the stem *dey-*, producing the *deys* observed by Krogh (2012), or what appears to be a distinct (although likely etymologically related) stem, *dos*, pronounced /dɔs/. This pattern obtains for both singular and plural nouns: no number distinction is observed in either dependent or independent demonstratives, except among ‘vos’ speakers who do not use forms other than *di* in the plural.

A small number of other forms are also found, including *dis* and *diye*, but these fall in to the patterns described above. *Dis* acts as a *d*-stem with an /s/ ending (i.e. it appears in the nominative as an independent demonstrative), while *diye* acts as a *d*-stem with an -e ending (i.e. it appears in the objective as an independent demonstrative). None of these innovative forms appear as dependent demonstratives.

#### 5.4 Overall trends in demonstratives

Four patterns or tendencies emerge from our findings. Patterns 1–3 are primarily found in ‘vus’ speakers while Pattern 4 is typical of ‘vos’ speakers.

However, very few speakers fit into one of these patterns without exceptions. Some speakers produce both *dos* and *deys* in free variation, some speakers from Patterns 2 and 3 occasionally use *deye* or *yene* in the nominative, some speakers from Patterns 1–3 occasionally produce stressed definite determiners as dependent demonstratives, and even some ‘vos’ speakers, who largely follow Pattern 4, produce *deye* on occasion. Pattern 1 appears to be more common among Israelis, while Pattern 2 is more prevalent in the New York area and Montreal. Stamford Hill speakers are split between Patterns 2 and 3, and speakers of Patterns 2 and 3 can be found in all geographical communities.

The restriction of *dos* and *deys* to nominative contexts, as well as the fact that these forms are not found dependently, suggests that independent demonstratives are pronominal forms. As discussed in Section 3, the singular personal pronouns also distinguish between nominal and objective forms. However, perhaps

Table 17: Patterns of demonstrative use in Contemporary Hasidic Yiddish

	Dependent	Independent	
		NOM	OBJ
Pattern 1	<i>deye</i> <i>yene</i>	<i>deye</i> <i>yene</i>	
Pattern 2	<i>deye/dey</i> <i>yene</i>	<i>deys</i> <i>yens</i>	<i>deye</i> <i>yene</i>
Pattern 3	<i>deye/dey</i> <i>yene</i>	<i>dos</i> <i>yens</i>	<i>deye</i> <i>yene</i>
Pattern 4	<i>di/der/dem</i> <i>yene/yener/yenem</i>	<i>dos</i> <i>yene, yens</i>	<i>dos/dem</i> <i>yene/yenem</i>

unexpectedly, for demonstratives but not for personal pronouns, the nominative/objective distinction persists in the plural. This appears to be a striking innovation in the demonstrative system compared to Standard and pre-War varieties of Yiddish: while the stems *dey-* and *-yen* existed historically, they do not appear to have been used as pronouns. Rather, as they carried the same case and gender morphology as definite determiners and dependent demonstratives, they appear to have acted as determiners even in dependent contexts. Like other functional categories, pronouns are usually considered a closed class and therefore resistant to new members. It is therefore surprising that the demonstrative pronouns observed in this study, *yens/yene* and *deys (dos)/deye* appear to have been so quickly and pervasively adopted, at least among ‘vus’ speakers.<sup>31</sup>

It is similarly striking that speakers, especially ‘vus’ speakers, do not appear to distinguish between singular and plural demonstratives, in either dependent or independent contexts. This innovation can clearly not be a result of contact with English or Modern Hebrew, as demonstratives in both of these languages agree with their nouns for number. It remains to be seen whether such a distinction will emerge.

<sup>31</sup>As novel pronouns, it is expected that they should carry the same case distinctions as other pronominal forms, such as 1SG *ikh* vs. *mikh/mir*, 2SG *du* vs. *dikh/dir*, 3SG *er* vs. *em*, etc. Indeed, innovative gender neutral pronouns in English, such as *ze/zir* and *sie/hir*, typically have the same case distinctions as other English pronouns (*I/me*, *she/her*, etc.).

Demonstratives constitute a distinct pattern to both possessive pronouns and attributive adjectives. The former, like demonstratives, distinguish between dependent and independent forms, although dependent possessives are unlike demonstratives in that they consistently distinguish between singular and plural possessa. Attributive adjectives have an invariant suffix *-e* that distinguishes them from predicative adjectives, which appear in the bare form. In demonstratives, there is no bare form of the *y*-stem and the dependent forms *dey* and *deye* appear in free variation. Given that none of these systems has a precedent in pre-War or Standard varieties of Yiddish it is surprising that Contemporary Hasidic Yiddish has innovated three distinct systems in the realms of attributive adjectives, possessive pronouns, and demonstratives.

The development of a single, uninflected form of the dependent demonstrative, as well as ‘vos’ speakers’ use of a variety of inflected forms, is in line with developments observed in definite determiners (Belk et al. 2020, In press). Additionally, the variety of individual strategies we find in the demonstratives and the overall heterogeneity of the system is consistent with our findings on the personal pronoun and possessive systems in suggesting a system in flux. How this situation will resolve itself is yet to be seen, and a number of questions remain open.

The first regards the apparently free variation between the dependent demonstratives *dey* and *deye*. Further study is required to determine whether their distribution is predictable due to some factor to which our questionnaire was not sensitive. If they are indeed in free variation, this situation may persist; one form might die out in favour of the other (leaving a system more similar to the definite determiners); or the two forms might differentiate into, for example, a singular and a plural form, as appears to be the case with dependent possessives. Similarly, it remains to be seen whether Patterns 2 and 3 will merge, or whether a distinction between *deys* and *dos* will emerge.

There is also the question of how robust Pattern 4 is or, put more generally, the extent to which the ‘vos’ variety of Contemporary Hasidic Yiddish will remain distinct from the ‘vus’ variety. Belk et al. (In press) suggests that the loss of morphological case and gender on full nominals happened later among ‘vos’ speakers as these communities historically mixed less with other communities, but that this loss is nonetheless complete. They argue that the change in the ‘vos’ variety may have been driven by contact with ‘vus’ speakers, among whom the innovations in the morphological case and gender system were more established. The same may be true of developments in the demonstrative system, an idea which is supported by the fact that the youngest ‘vos’ speaker in our study produced some tokens of *deye*, a characteristically ‘vus’ form.

Finally, given the relative rarity of Pattern 1, we might wonder how long it will persist. As the personal pronoun system seems to crystallise into one that distinguishes nominative from objective forms, at least in the singular, it may be that speakers of Pattern 1 adopt one of the other patterns and thereby develop a nominative/objective distinction in demonstrative pronouns as well.

## 6 Conclusions

In this chapter we have shown that the Contemporary Hasidic Yiddish pronominal system has undergone a number of innovations vis-à-vis the pre-War and Standard Yiddish varieties. Our results are based on a survey questionnaire which systematically elicited data for the personal pronouns, reflexive pronouns, dependent and independent possessive pronouns, and dependent and independent demonstratives from 29 native speakers of Contemporary Hasidic Yiddish in the main Hasidic centers worldwide (Israel, the New York area, London's Stamford Hill, the Montreal area, and Antwerp). Our findings indicate a system in flux, with a high degree of variation present both between and within speakers regardless of geographical location. This variation applies to all of the pronominal categories that we examined.

With respect to the independent personal pronominal paradigm, we found a widespread trend towards paradigm levelling, with the traditional three-way nominative/accusative/dative distinction in the singular shifting to a two-way nominative/objective distinction, and the traditional two-way nominative/objective distinction in the 1PL and 2PL shifting to a single unchanging form. These changes appear to have been driven by the effects of substantial dialect mixing, with the historical Mideastern Yiddish pronominal paradigm exerting the greatest influence, both in terms of the structure of the paradigm and in terms of actual forms used (e.g. the 2PL form *enk*). The reflexive paradigm has also undergone a degree of levelling in comparison with the pre-War Mideastern variety of Yiddish. Our survey also allowed us to map a system of strong and weak personal pronouns, which are likely to have existed in some form or another in pre-War Yiddish varieties, especially the Mideastern ones, but, perhaps because they are an effect of colloquial speech, have not been clearly documented previously. It is noteworthy that 'vus' speakers appear to have a nominative/objective distinction in the weak 3pl, which does not to the best of our knowledge have historical precedent and stands in contrast to the general trend towards simplification in the personal proun paradigm.

The possessive pronouns have dependent and independent variants. The dependent variants exhibit one form (with a Ø-ending) when modifying a singu-

lar noun and another one (with an *-e* ending) when modifying a plural noun. This is in keeping with the pre-War model and goes against our prediction that the Contemporary Hasidic Yiddish dependent possessive pronouns would have undergone or be undergoing the same streamlining process as attributive adjectives, which we have elsewhere (Belk et al. 2020, In press) demonstrated to have lost their pre-War case and gender distinctions in favour of an invariant attributive marker *-e*. The independent variants are morphologically distinct from their dependent counterparts, but have lost all of the pre-War case and gender distinctions in favour of a more streamlined innovative model with two paradigms, *-s* and *-e*, which have the same function and can be used interchangeably even within the speech of a single participant.

The demonstrative pronouns likewise have dependent and independent variants. The proximal/distal distinction exhibited in both the dependent and independent variants differs from those of both English and Modern Hebrew, with the “distals” serving primarily to mark contrast. The stem used for the proximal demonstratives varies somewhat and thus seems to be in flux, though ‘vos’ speakers tend to use the traditional definite determiner forms (*der*, *di*, *dos*, and *dem*, though without the pre-War case and gender distinctions), while ‘vus’ speakers tend to use the novel stem *dey-*. The dependent variants behave morphologically like (Contemporary Hasidic Yiddish) definite determiners, i.e. they show no case or gender distinctions. However, the independent variants show a novel nominative/objective case distinction (in contrast to pre-War Yiddish, which exhibited the same case and gender morphology as the definite article and adjectives). This suggests that the demonstratives have been reanalyzed on analogy with the personal pronouns and follow the same pattern of a two-way nominal/objective distinction that the singular personal pronouns exhibit. As such, we posit that Contemporary Hasidic Yiddish has innovated a novel demonstrative pronoun.

These innovative features of the pronominal system support our claim that the Yiddish spoken in 21<sup>st</sup>-century Hasidic communities constitutes a distinct variety of the language, which, though descended from the pre-War Eastern European dialects, has evolved away from them to such an extent that it can no longer be analyzed within this older dialectal framework. While the pronominal system has not lost case and gender in the same way that the nominal system has, the pronominal innovations are of the same magnitude as those affecting the nominal case and gender system. Our analysis shows that Contemporary Hasidic Yiddish has not simply lost forms and functions in comparison with the pre-War varieties, but rather has innovated them. These innovative features are not determined directly by contact with the dominant co-territorial languages, but rather

are internal developments which bear witness to the linguistic vibrancy of Contemporary Hasidic Yiddish.

## Abbreviations

1	first person	HON	honorific
2	second person	IMP	imperative
3	third person	INAN	inanimate
ACC	accusative	MASC/M	masculine
DAT	dative	NEUT	neuter
DEF	definite	NOM	nominative
DEM	demonstrative	PL/P	plural
FAM	familiar	SG/S	singular
FEM/F	feminine		

## Acknowledgements

We gratefully acknowledge Eli Benedict for his help in administering the questionnaire, as well as for his insightful discussions. We are also thankful for the contributions of Shifra Hiley, and for the cooperation of our participants. This research is generously funded by the UK Arts and Humanities Research Council and the Leverhulme Trust.

## References

- Assouline, Dalit. 2007. *Shimur u-teruma ba-yidish ha-haredit be-yisra'el*. Hebrew University of Jerusalem. (PhD diss.).
- Assouline, Dalit. 2010. The emergence of two first-person plural pronouns in Haredi Jerusalemitic Yiddish. *Journal of Germanic Linguistics* 22(1). 1–22. DOI: 10.1017/S1470542709990183.
- Assouline, Dalit. 2014. Language change in a bilingual community: The preposition *far* in Israeli Haredi Yiddish. In Marion Aptroot & Björn Hansen (eds.), *Yiddish language structures*, 39–62. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110339529.39. <https://www.degruyter.com/view/book/9783110339529/10.1515/9783110339529.39.xml>.
- Belk, Zoë, Lily Kahn & Kriszta Eszter Szendrői. 2020. Complete loss of case and gender within two generations: Evidence from Stamford Hill Hasidic Yiddish. *Journal of Comparative Germanic Linguistics* 23. 271–326.

- Belk, Zoë, Lily Kahn & Kriszta Eszter Szendrői. In press. Absence of morphological case and gender marking in Contemporary Hasidic Yiddish worldwide. *Journal of Germanic Linguistics*.
- Belk, Zoë, Lily Kahn, Kriszta Eszter Szendrői & Sonya Yampolskaya. In prep. T/V distinctions in Contemporary Hasidic Yiddish. m.s.
- Biale, David, David Assaf, Benjamin Brown, Uriel Gellman, Samuel Heilman, Moshe Rosman, Gade Sagev & Marcin Wodzinski. 2018. *Hasidism: A new history*. Princeton, NJ: Princeton University Press.
- Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In Henk van Riemsdijk (ed.), *Eurotyp. volume 5/part 1: Clitics in the languages of Europe*, 145–234. Berlin, New York: De Gruyter Mouton.
- Fleischer, Jürg. 2018. Western Yiddish and Judeo-German. In Benjamin Hary & Sarah Bunin Benor (eds.), *Languages of Jewish communities past and present*, 239–275. Boston & Berlin: De Gruyter Mouton.
- Halevy, Rivka. 2013. Deixis: Modern Hebrew. In Geoffrey Khan, Shmuel Bolozky, Steven Fassberg, Gary A. Rendsburg, Aaron D. Rubin, Ora R. Schwarzwald & Tamar Zewi (eds.), *Encyclopedia of Hebrew language and linguistics*. Leiden, The Netherlands: Brill Online.
- Harshav, Benjamin. 1990. *The meaning of Yiddish*. Berkeley: University of California Press.
- Jacobs, Neil G. 1990. Northeastern Yiddish gender-switch: Abstracting dialect features regionally. *Diachronica* 7(1). 69–100.
- Jacobs, Neil G. 2005. *Yiddish: A linguistic introduction*. Cambridge, UK: Cambridge University Press.
- Kahn, Lily. 2016. Yiddish. In Lily Kahn & Aaron D. Rubin (eds.), *Handbook of Jewish languages*, 639–747. Leiden, The Netherlands: Brill.
- Katz, Dovid. 1987. *Grammar of the Yiddish language*. London: Duckworth.
- Krogh, Steffen. 2012. How Satmarish is Haredi Satmar Yiddish? In Marion Ap-troot, Efrat Gal-Ed, Roland Gruschka & Simon Neuberg (eds.), *Leket: Yiddish studies today*, 483–508. Düsseldorf: Düsseldorf University Press.
- Kuznitz, Cecile Esther. 2010. YIVO. In *YIVO encyclopedia of Jews in Eastern Europe*. <https://yivoencyclopedia.org/article.aspx/YIVO>.
- Mark, Yudel. 1978. *Gramatik fun der yidisher klal-shprakh*. New York: Congress for Jewish Culture.
- Mazin, R. 1927. *Zelbst lehrer, idish*. London: n.p.
- Neleman, Ad & Tanya Reinhart. 1998. Scrambling and the PF interface. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments: Lexical and compositional factors*, 309–353. Stanford: CSLI.

## *6 Innovations in the Contemporary Hasidic Yiddish pronominal system*

- Nove, Chaya. 2018. Social predictors of case syncretism in New York Hasidic Yiddish. *University of Pennsylvania Working Papers in Linguistics* 24(2). 11: 1–10.
- Okrutny, Iosef. 1953. *Baluter vebers*. Buenos Aires: Farband fun Lodzher un umgegnt in Argentine.
- Ruys, Eddy G. 2008. Stranding, weak pronouns, and the fine structure of the Dutch Mittelfeld. *Natural Language & Linguistics Theory* 26(3). 547–587.
- Sadock, Benjamin & Alyssa Masor. 2018. Bobover Yiddish: “Polish” or “Hungarian?” *Journal of Jewish Languages* 6(1). 89–110. DOI: 10.1163/22134638-06011133.
- Shtiebel, Kave. N.d. <http://www.kaveshtiebel.com>.
- Stirling, Lesley & Rodney Huddleston. 2002. Deixis and anaphora. In Rodney Huddleston & Geoffrey K. Pullum (eds.), *The cambridge grammar of the english language*, 1449–1564. Cambridge, UK: Cambridge University Press. DOI: 10.1017/9781316423530.018.
- Weinreich, Uriel. 1964. Western traits in Transcarpathian Yiddish. In Lucy S. Dawidowicz (ed.), *For Max Weinreich on his seventieth birthday: Studies in Jewish languages, literature, and society*, 245–264. The Hague: Mouton & Co.
- Wolf, Meyer. 1969. The geography of Yiddish case and gender variation. In Marvin I. Herzog, Wita Ravid & Uriel Weinreich (eds.), *The field of Yiddish: Studies in language, folklore and literature*, 102–215. The Hague: Mouton. DOI: 10.1515/978311699509-008.



# Chapter 7

## Validity of crowd-sourced minority language data: Observing variation patterns in the Stimmen recordings

Nanna Hilton<sup>a</sup>

<sup>a</sup>University of Groningen

Minority languages are underrepresented in linguistic research, and a possible reason for this is the lack of accessible speech recordings from lesser-used languages. This paper considers the usability of crowd sourced minority language data for research, focussing on the speech recordings and reported dialect knowledge collected with the smartphone application Stimmen, ('Voices' in Frisian). In this paper variation patterns in Frisian speech data from the Stimmen project (2017-2019) are compared with findings from previous sociolinguistic research in Fryslân. The comparison focusses on three phonological variables in Frisian speech: the coda cluster /sk/, the vowel in 'eye', and the realisation of post-vocalic coda /r/. The analysis conducted on the crowd sourced recordings show the same variation patterns as that of previous research, giving validity to the data for use in sociolinguistic studies of variation and change in minority language settings. This gives hope for future research that relies on the collection of speech samples in a remote capacity.

### 1 Background

#### 1.1 The European bias in (socio)linguistics

The 'Stimmen fan Fryslân' project, Stimmen for short, is a citizen science project with the aim of collecting speech data from lesser used languages for research and technological development. While we estimate that there are 5,000-7,000 languages used across the globe, publications within the field of linguistics largely



represent (speakers of) large languages, and especially English. Nagy & Meyerhoff (2008) conclude, for example, that in 449 publications and presentations in four key outlets for research in (variationist) sociolinguistics 50-70% concerns the English language.

For a field such as sociolinguistics, that deals predominantly with diversity in language, how and why language changes, and what the linguistic and social consequences of such changes are (cf. Weinreich et al. 1968), it is clear that biases in the empirical foundations are problematic. This is a likely reason why in recent decades more studies have appeared that incorporate the social reality outside that of majority populations into sociolinguistic theory. Stanford & Preston (2009) and Stanford (2016) point out how minority language communities offer a welcome chance for variationist sociolinguistics to revisit principles of linguistic variation and change, most of which take a European, or Anglo-American, linguistic and societal constellations for granted. Nagy (2009: 400) notes in her study of Faetar that minority languages without a codified and accepted standard give sociolinguists an opportunity to do research without influence from a standard language ideology. Studies outside the Anglo-American sphere also reconsider the arrangements of the traditional social variables class, gender and age: Noglo (2009), for example, shows how factors such as monetary income, property ownership and fortune are not part of what constitutes social class in many societies, providing a case study where ethnicity and community membership is used to define layers of social hierarchies. Brunelle (2009), similarly, argues how the Labovian principles of gender and language cannot be assumed before carefully considering the type of access to political and financial privilege and prestige that women have in specific communities outside the US and UK. O'Shannessy (2009) indicates in her study of language change in northern Australian indigenous languages that the variationist sociolinguistic principles of age and apparent time cannot be assumed in communities where contact with a majority language is widespread and the younger generations have a higher degree of bilingualism.

Perhaps more than for other linguistic disciplines the study of social variation in language is reliant on previous descriptive work having been done for the language in question. A researcher wanting to enter a community to understand the social role of language is reliant on already being a speaker of the language, or on available teaching material for acquiring the language. Arguably, linguistic documentation is also key for a number of applications of our linguistic research: the development of language technology, monitoring of learning abilities, and forensic analyzes concerning language, cannot exist without proper documentation. It is problematic then, that the majority of the world's languages are still

under-documented (Hammarström and Nordhoff, 2011). More than a thousand varieties are only described with a wordlist, a text collection, or not at all. One of the challenges the linguistic community is faced with as of now is that available data, and particularly speech data, from lesser used languages is hard to come by. While collections of speech can often be found on the internet, it may not be in a format that lends itself well for research, without annotation and translations, for example, or of poor quality due to recordings made in noisy environments.

## 1.2 Citizen science and crowd sourcing language data

A way to increase representation of under-resourced language communities in science is to engage in “citizen science”: the participation of the general public alongside scientific researchers, in some or all steps of the research project (cf. Bonney et al. 2016). While the label “citizen science” is relatively recent, the practice itself is not. In linguistics, the contribution of the general public as providers of empirical data has a history that goes back centuries. Early descriptive linguistic work, such as dialectological studies where speakers of localised varieties provided data for the analysis of regional variation in language (e.g. Wenker 1881), can be seen as examples of “citizen science”. However, recent technological developments have made public involvement in linguistic research substantially easier to facilitate, see, for instance, examples of dialectological surveys being conducted using the internet such as Vaux (2004) and Möller & Elspaß (2015). Smartphones offer further functionalities that can be employed to collect, store, share and analyze large amounts of language. Applications for iOS or Android have been used to, e.g. collect speech for development of language technology (De Vries et al. 2014), to make high-quality recordings for acoustic phonetic research (De Decker & Nyčz 2011), or to collect reported language use for the creation of new dialect maps (Kolly & Leemann 2015). Yet, an issue with remote crowd sourcing of speech data remains its anonymous nature and the lack of control on the side of the researcher.

## 1.3 What constitutes “good” crowd sourced minority language data?

A commonly held conception about crowd sourced data is that it is problematic for research purposes due to the lack of control that the researcher can execute over the data collection process and context. Researchers could receive contributions that are done merely in jest, there could be participants who contribute several times, there could be contributors who have not read instructions, or contributions that do not meet the expectations of the researcher in another way.

Studies of the validity and reliability of crowd sourced data indicate that these worries are not entirely justified. Lind et al. (2017) find in their study of “crowd-coding” (coding of content analysis of news articles in German) that paid volunteers (through the platform Mturk) performed comparably to five research assistants, but that there was variability across different types of tasks, and within the group of coders. Horn (2019) concludes in a study of whether non-expert coders can perform quite complex content analysis of political messages, and conclude that crowd sourced coding is equally reliable to that of experts. However, there seems to be a limit to task complexity before the validity and reliability suffers: Shing et al. (2018) compare experts’ and crowds’ (CrowdFlower) annotations of online postings for authors’ suicide risk, and find, unsurprisingly, that the experts outperform the crowd, even when the crowd has been given detailed task instructions. One previous study has considered the validity of crowd sourced linguistic judgements for the purpose of studying language variation and change. Leemann et al. (2016) find that the dialect judgements crowd sourced with the smartphone based Dialekt Äpp correlate to a high degree with speech samples collected with traditional dialectological methods.

The findings in Leemann et al. (2016) bode well for a study of the validity of crowd sourced speech data using smartphones. However, a number of validity-related concerns exist specifically for the crowd sourcing of minority language data. Shameem (1998) points out, for example, that self-reported minority language proficiencies do not always compare to actual test results. In her study of Indo-Fijians living in New Zealand she finds some over-reporting of oral minority language proficiency. Contrarily, Nicholas (1988) finds that minority children typically under-report their knowledge of the heritage language in surveys of British language diversity. There is also reason to question whether surveys designed for the whole of a minority speech community does in fact reach all those that could interest the researcher. “New speakers”, for instance, could experience a lack of ownership and legitimacy in the use of the minority language (cf. O’Rourke & Ramallo 2013) and not see themselves as “real” minority language speakers.

Minority language communities cannot be approached with the traditional dialectological methods employed for instance in Leemann et al. (2016) either. While many minority languages may have a written standard, the acceptance levels of the standard language can be highly variable within the communities. Dorian (1978: 592) concludes that a majority language, in her case English, can act as the formal register in bilingual speakers who are less proficient in a minority variety, Gaelic in the case at hand. In communities where the minority language is not used throughout the educational system the majority language

generally enjoys higher instrumental value. This means that the reliance of written language in crowd sourcing of minority language can be problematic, with respondents wanting to perform well and give the most correct form as a response. This could lead to questionable outcomes, complicating the design of the study.

This study is the first to consider the validity of crowd-sourced speech data from a minority language. It does so by comparing the speech that users have themselves, as the citizen scientists, categorised as “Frisian”, collected in a picture naming task in the Stimmen application, with previous data collected in Fryslân to investigate language variation and change. Additionally, data from a gamified dialect task for Frisian is used for the comparison. Before moving on to the analysis of the speech data the design and the context of the study is laid out below.

## 2 The current study

### 2.1 Frisian and the basis for data comparison

Fryslân is the only officially multilingual province in the Netherlands, and is a region in which several (regional and migrant) minority languages are used. Frisian-Dutch bilingualism is widespread in the province: 75% of the 9915 inhabitants that the Province of Fryslân surveyed in 2015 report being able to speak Frisian (TaalAtlas 2015: 3). Three quarters of the population would correspond to some 485,000 speakers of Frisian in Fryslân (Centraal Bureau voor de Statistiek 2018). These speakers are all presumed bilingual as secondary schooling is taught partially, or only, in Dutch. The province Fryslân is also home to mixed languages (results of long-term contact between Dutch and Frisian) ‘Bildts’ and ‘Town Frisian’ (van Bree 1994, Hoekstra & van Koppen 2000), as well as to varieties of another West-Germanic language family: Low Saxon, spoken along the southern border of the province of Fryslân.

An assumption for sociolinguistic work on Frisian is that the language is converging towards Dutch on all linguistic levels, cf. Breuker (1992) and De Haan (1997). The study of contact phenomena between Frisian and Dutch has been given considerable attention in Frisian linguistics, e.g. Sjölin (1976), yet generally in an introspective manner, or with anecdotal evidence. Very few empirical studies of speech variation in Frisian exist, and those that have been conducted have all been studies of phonetics and phonology. They indicate no evidence of convergence between Frisian and Dutch: Feitsma et al. (1987, data collected

1982–1984) is the oldest study available, looking at sandhi phenomena in an apparent time study. The study finds no signs of convergence between Frisian and Dutch, but rather indications of divergence for some phonetic variables. Van Bezooijen (2009) investigates the variation and change in pronunciation of /r/ in Frisian, which traditionally has the alveolar trill, and concludes that the approximant variant of /r/ that is gaining ground in the Dutch speech community does not occur at all in the Frisian data (Van Bezooijen 2009: 312). Finally, Nota et al. (2016) find gender and age differences in the realisation of intonation contours in 40 bilingual Frisian-Dutch speakers, but no direct indications that Dutch and Frisian intonation contours are converging.

In this paper, the research question does not directly concern the convergence between Frisian and Dutch, but rather asks what the usability, that is, the validity, of Frisian data collected by crowd sourcing is for conducting sociolinguistic analyzes. However, the data will be compared to other work concerned with the question of whether Frisian is changing towards Dutch and can in that sense possibly corroborate some previous findings. One of the data sets used for comparison is that of Van Bezooijen (2009), used to compare variation patterns in the pronunciation of (r). Furthermore, variation in the consonant coda cluster (sk) is looked at alongside data from Hilton & Weening (2014). Variation in the vowel in the word *each* “eye” is also considered, by comparing crowd-sourced data to that from an ongoing study by Stefan, Klinkenberg & Versloot (cf. Stefan et al. 2014).

## 2.2 The citizen science project “Stimmen”

The crowd sourced data discussed in this paper was collected with the smartphone application Stimmen. The app was part of the larger “Stimmen fan Fryslân” project funded by the program “Lân fan taal” in the European Capital of Culture project for Leeuwarden 2018. The citizen science components of the Stimmen project consisted of collaboration with twelve secondary schools in Fryslân as well as numerous online and offline activities held throughout the year of 2018. An estimated 2,000 respondents were reached through the offline activities, that gave the researchers an opportunity to share findings with the general public, and, crucially, for the public to approach the researchers to post questions about language. Since the start of the project the smartphone application has been downloaded 6,039 times (iOS: 2,989; Android: 3,050. In 2017, 28% of smartphones in the Netherlands was an iPhone (Steemers et al. 2017), but the application could also be downloaded on an iPad which may eschew the iOS user numbers), yet this is not equivalent to the amount of people who have submitted data through

the application. The dialect quiz has been used a total of 15,131 times. The discrepancy between the two figures can be explained by the fact that the dialect quiz, see details below, is also available as an easily shared web application on [stommen.nl](http://stommen.nl).

### 3 The Stimmen smartphone application

The Stimmen application is available for Android or iOS and consists of the following components: a start screen with choice of interface language (English, Dutch, or Frisian), a tutorial, and a root menu with choices of four options: a picture naming task, a dialect quiz, a speech map in which free speech can be recorded, and an “about” section. In this paper, the data collected with the picture naming task and the dialect quiz are considered only. For an overview of all the functionalities of the application, the researchers’ ethical considerations, and the collection of other speech and language attitude data, see Hilton (2021).

#### 3.1 The Stimmen picture naming task

In the Stimmen application, a picture naming task can be found with 87 hand-drawn images of everyday objects in the Netherlands. The scientific purpose of this module is to collect data that can be used to document phonological and phonetic patterns for any language recorded, but the nameability and selection with regards to avoiding heteronyms was done on the basis of the varieties spoken in Fryslân, only. The 87 constructs represent all the phonemes and their allophonic realisations in the most spoken varieties found in the region where the activities surrounding Stimmen were primarily conducted: in the north of the Netherlands. The current list of languages in which one can record includes 38 different languages, but more can be added by contacting the Stimmen team, or the author of this paper.

When opening the picture task from the root menu a prompt informs the participants that the aim of the task is to name as many pictures as possible, and in as many languages as they can. Next, participants are asked to fill in a short questionnaire about where they are from (to indicate this on a map), their gender, their age bracket, which languages they are most fluent in, which languages they actively use in their life, and to answer the open question whether there is anything they would like to share with the researchers about their own variety. Upon considerations with the ethics committee and legal experts associated

with the project it was decided that further biographical data could not be collected due to privacy concerns, as the speech recordings are publicly available alongside the social information collected.

The screenshot shows a mobile application interface titled 'Stimmen'. At the top, there is a green header bar with a back arrow, the title 'Stimmen', and a close button. Below the header, the text 'About you' is displayed. There are three dropdown menus: 'Where are you from?' with a location pin icon, 'What is your gender?' with a 'Choose...' option, and 'What is your age bracket?' with a 'Choose...' option. Under the 'Languages' section, there are two dropdown menus: 'Which language are you the most proficient in?' and 'Which languages do you actively use in your life?'. Both have 'Choose...' options. At the bottom, there is a text input field with the placeholder 'Tell us anything that you think is important to know about your dialect'.

Figure 1: The meta-data questionnaire used for all tasks in 'Stimmen'.

Participants are then asked which language they would like to record in. After choosing a language, a randomised selection of a picture is made and shown to the user. They can then press the screen to record, and thereafter send the recording to the researchers. A prompt asks whether the user is sure they want to share the recording publicly. After naming 10 pictures the user can choose to go on or go to the gallery of named pictures to check their progress.

### 3.1.1 The recordings made in the picture naming task

Some 2,000 distinct individuals have used the picture naming task up until 2020, creating 41,553 individual recordings of words. 24,214 of these were created by female speakers, 17,028 by male speakers and 311 by speakers identifying as 'other'. The age distribution of the recordings is found in Figure 5. Almost half of the recordings are made by respondents younger than 30 years, which means that the recordings in Stimmen represent younger language. For the sake of investigating language change in progress this may not be a disadvantage, but for any

## 7 Validity of crowd-sourced minority language data

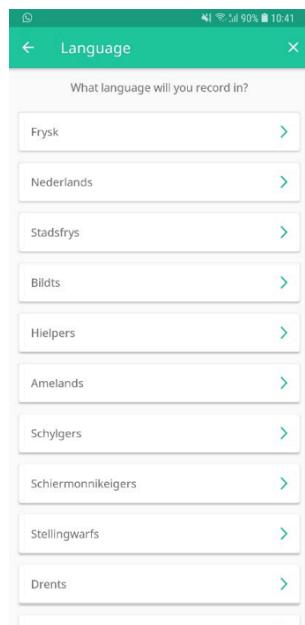


Figure 2: Some of the possible recording languages in 'Stimmen'.

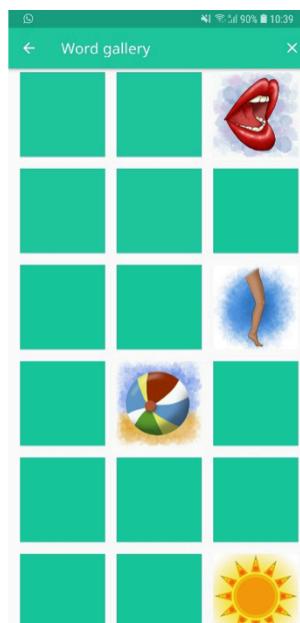


Figure 3: Word gallery after naming four pictures in 'Stimmen'.

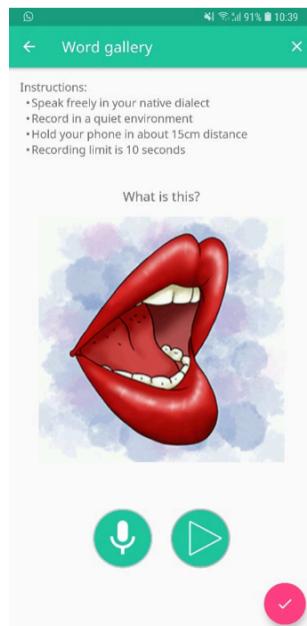


Figure 4: Screenshot of the picture task in 'Stimmen', showing the picture 'mouth'.

study reliant on balanced age data, one would need to take this imbalance into account in the analysis.

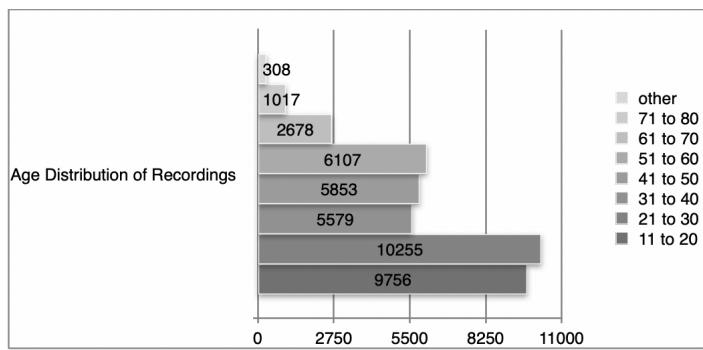


Figure 5: Distribution of recordings from the picture task in 'Stimmen' across age groups.

The top eight languages recorded in the picture naming task are shown in Table 1. Nearly three quarters of the recordings are made in Frisian, or varieties spoken within Fryslân.

Table 1: Number of Recordings Made by Languages in Stimmen Picture Task

Language	No. of recordings	Language	No. of recordings
Frisian	31,580	English	1,246
Dutch	5,858	German	207
Town Frisian	914	French	168
Bildts	540	Others	1,040

### 3.2 The Stimmen dialect quiz

A second component of the Stimmen application is a gamified task that guesses where people hail from within the Province of Fryslân. This task, “dialect quiz”, was based on Leemann & Kolly (2013) and their subsequent applications. Data gathered with this module provides researchers with the public’s knowledge of traditional linguistic variants, and regional and social (age, gender) variation in reported dialect use can be analyzed. A straight-forward way to develop a dialect quiz is to use an existing corpus of speech samples and create a set of questions that will result in a unique combination of answers for each of the locations represented in the corpus. Naturally, this is not easily done for minority languages. However, for the case of Frisian, mixed languages and Low Saxon as spoken in Fryslân previous dialectological work facilitated this. The prediction in Stimmen was created based on the GTRP database (2017) with data from 58 informants from the 1980s in Fryslân. The GTRP allowed for creation of a multilingual prediction for 58 different locations in Fryslân in which varieties of Frisian, Low Saxon and mixed languages are spoken.

While the picture naming task described in 3.1 was developed for use in *any* language, this part of the application was directed only at inhabitants of Fryslân. The quiz game asked users to provide their local variant for 19 Standard Dutch words (see Table 1), as all inhabitants of Fryslân are believed to be fluent in Dutch. 2-10 possible variants were available to the user to listen to and to choose between. After giving their personal variant to all 19 words the app makes three guesses of where a user could be from. The user can then indicate whether this prediction is correct or not and fill in the same meta-data questionnaire as used for the picture naming task (see Figure 1).

Table 2: Words in Stimmen's Dialect Quiz

Standard Dutch prompt	Translation	No. of variants
armen	'arms'	9
avond	'evening'	9
bij	'bee'	10
blad	'leaf'	3
borst	'breast'	8
dag	'day'	4
deurtje	'door' (+ diminutive)	12
geel	'yellow'	5
(ik ben) gegaan	'I have gone'	11
(ik heb) gezet	'I have placed'	6
heel	'whole'	5
kaas	'cheese'	7
koken	'cook' (infinitive)	11
oog	'eye'	7
(toe)sprak	'spoke'	8
tand	'tooth'	8
trein	'train'	5
vis	'fish'	2
zaterdag	'Saturday'	7

### 3.2.1 The data collected in the dialect quiz component

Of the 15,131 times the dialect quiz has been used, the survey has been filled in 3,340 times: 1,688 times by a female user; 1,633 by a male user and 19 times by a user identifying as 'other'. This rather low proportion (some 22%) of respondents filling in meta data could have to do with the fact that the survey was easily skipped in the popular web version of the Quiz (it was the very last component and the respondents had already received their guessed location). The age distribution of those using the Dialect Quiz and filling in the survey is given in Figure 6. Note that, as was the case in the picture task recordings, the youngest generations are over-represented in the collected data.

The project website houses interactive maps of the distributions of the answers for all variables and all variants, on <http://stimmen.nl/uitspraakkaarten/>.

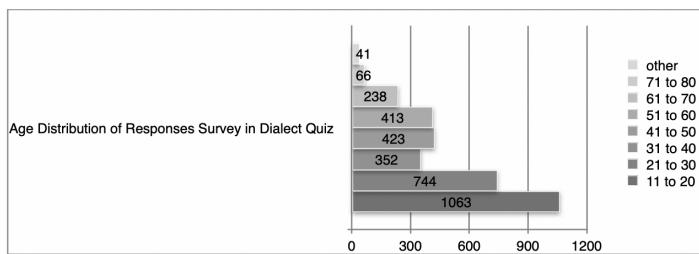


Figure 6: Distribution of survey responses in the Dialect Quiz in Stimmen across ages.

## 4 Analysis of variation in the Stimmen corpus

The manner chosen to validate the data collected with the picture naming task and the dialect quiz in the Stimmen project is to compare patterns of variation in the data with that found in studies collected with different methodologies, in approximately the same time period, in the Province of Fryslân. There are three comparisons that can be made with data in the Stimmen corpus: i) a comparison between the variation patterns in production of the coda cluster (sk) in the picture naming task, alongside reported usage of the variants from the dialect quiz data, with usage patterns found by Hilton & Weening (2014), ii) a comparison between the reported usage of the pronunciation variants for the Frisian word for ‘eye’ *each* in Stimmen with reported usage of the variants in Stefan, Klinkenberg & Versloot (2014), and iii) a comparison between the variation patterns in (r) in the data from the picture naming task with that in Van Bezooijen (2009).

### 4.1 Variation in (sk)

The coda cluster (sk) in Frisian has two variants: [sk] is the canonical Frisian variant but variant [s] is also commonplace. [s] is the variant used in equivalent cognates in the majority language Dutch. The variable occurs in two morphological contexts: (1) as the coda to a number of nouns and verbs that had <sk> as (part of the) coda in Old Frisian and other older Germanic varieties (OED Online, 2015); i.e., *fisk* ‘fish’; *wask* ‘wash’; *bosk* ‘forest’ or, more frequently throughout the lexicon, as (2) the coda in the adjectival or adverbial suffix equivalent to English ‘ic’: a) *histoarysk* ‘historic’ *fantastysk* ‘fantastic’, in which it occurs in the loan-morpheme –ysk from High German and Dutch. The Dutch equivalent phoneme in cognate words in both (1) and (2) above is /s/ cf. Dutch *vis*; *was*; *bos*; *historisch*; *fantastisch*. Hilton & Weening (2014) find lexical constraints exist on

the variation with certain adjectives and nouns less likely pronounced with the full cluster, and some words, such as *gånsk* ‘whole’ never undergoing variation. Furthermore, a higher writing proficiency in the minority language correlates with a higher proportion of the minority variant [sk] in speech. Hilton & Weening (2014) further argue that the variant [sk] is associated with being authentically Frisian. The data from Hilton & Weening (2014) has been made available for comparison with the data collected in the Stimmen app. That data was collected using a translation task and a ‘map task’ (where two participants must lead each other through incongruent maps) with 31 participants aged 15-62 ( $M=39.9$ ,  $SD=14.2$ ); where 15 participants were male, 16 female.

In the Stimmen application, the lexical item ‘fish’ exists both in the dialect quiz, where 3144 of the informants who filled in questionnaires reported their variant usage, as well as in the picture naming task, where 419 recordings were made. 14 recordings were discarded from the corpus, due to the recording being inaudible, or the naming of the picture as a particular breed of fish (most often *bears* ‘perch’). In the data from Hilton & Weening (2014), 76 tokens of ‘fish’ exist. The distribution of the pronunciation variants of ‘fish’ in the three different data sets are rendered in Table 3–5. The Dialect Quiz and Picture Task data render the same distributions as the data collected by Hilton & Weening (2014). It is also clear that female speakers use a higher proportion of [s] than males in all three data sets.

Table 3: The distribution of variants [s] and [sk] in the three data sets compared

	Stimmen dialect quiz data	
	female	male
[sk] (%)	1118 (70.7%)	1123 (72.8%)
[s] (%)	464 (29.3%)	420 (27.2%)
Total	1582 (100%)	1543 (100%)

## 4.2 Variation in (each) - ‘eye’

A second linguistic variable in the crowd sourced Stimmen data that can be compared to other research is the reported pronunciation of ‘eye’ *each*. This linguistic variable has a standardised variant in written Frisian, which is representative

Table 4: The distribution of variants [s] and [sk] in the three data sets compared

	Stimmen picture task data	
	female	male
[sk] (%)	184 (76.3%)	135 (82.3%)
[s] (%)	57 (23.7%)	29 (17.7%)
Total	241 (100%)	164 (100%)

Table 5: The distribution of variants [s] and [sk] in the three data sets compared

	Hilton & Weening (2014)	
	female	male
[sk] (%)	28 (71.8%)	29 (78.4%)
[s] (%)	11 (28.2%)	8 (21.6%)
Total	39 (100%)	37 (100%)

of one of three main variants that exist in spoken language [ɪ.əx], while variants [e:x] and [ɛ:x] are associated with particular regional varieties of Frisian (cf. Breuker 1992: 19), making the variable particularly suitable to consider when assessing whether regional dialect levelling, or convergence, is ongoing in the Frisian speech community.

No published studies of this variable exist, but the authors of an on-going linguistic survey project (Stefan, Klinkenberg & Versloot p.c.) in Fryslân have kindly provided a map with distributions of the three Frisian variants for *eye* in 250 survey responses. The study uses a traditional dialectological methodology with a survey including direct questions about language use and attitudes (see Stefan et al., 2014 for more details of the methodology). The resulting dialectological map, shown in Figure 7, can be compared to the dialect maps made with the answers from the dialect quiz in the Stimmen project.

In figure 7, the regional distributions of the variants for *eye* in the comparison dataset is shown. The variant is indicated by the colour of the municipality as well as in the pie charts. The western and southern municipalities as shown in Figure 8 have a majority of *eech* [e:x], while the central north and north-west has

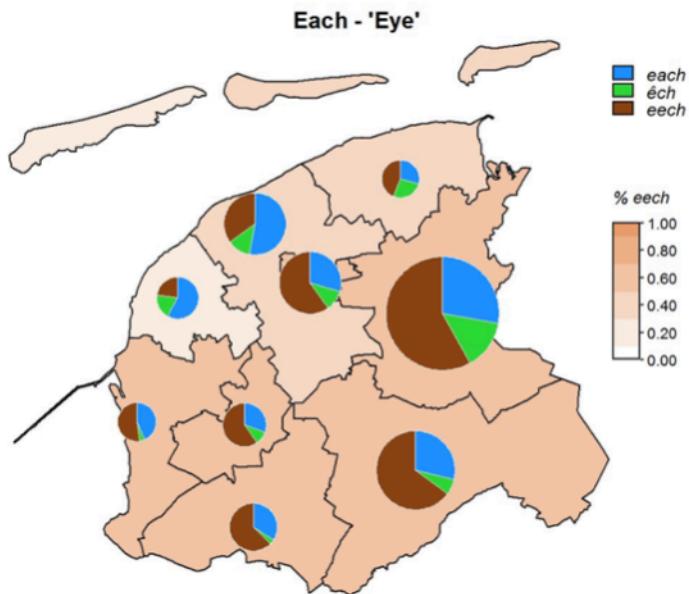


Figure 7: Distribution of the variants for 'eye' in an ongoing study by Stefan, Klinkenberg & Versloot (p.c.).

a majority of standard variant *each* [i.əx]. The final variant *éch* [ɛ:x], indicated with green colour, is particularly frequent in the north-eastern municipality. If looking at distribution maps<sup>1</sup> for variant usage in the Stimmen corpus in Figures 8–10, we see a very comparable picture in the distribution of the variants (despite using lower level neighbourhood borders, as opposed to municipal borders). Each map contains the number of realisations of the different variants in the top right corner. Each realisation is a unique user. The brown variant in Figure 7, is *eech* [e:x] in Figure 8 and shows dark red colours throughout the southern areas and up alongside the western border of Fryslân. Variant *each* [i.əx] (blue in Figure 7) variant has the highest proportion in the central and western part of the north of Fryslân also in Figure 9. Finally, the *éch* [ɛ:x] variant (green in Figure 7) shows the highest proportion in the north-eastern area in Figure 10. The dialect quiz data, then, shows a regional distribution of variants that follow the same pattern as that in findings of investigation using more traditional dialectological methodologies.

<sup>1</sup>I would like to extend a heartfelt thanks to Herbert Kruitbosch for creating the interactive maps of the responses in the dialect quiz. They can be found on [stommen.nl/uitspraakkaarten](http://stommen.nl/uitspraakkaarten)

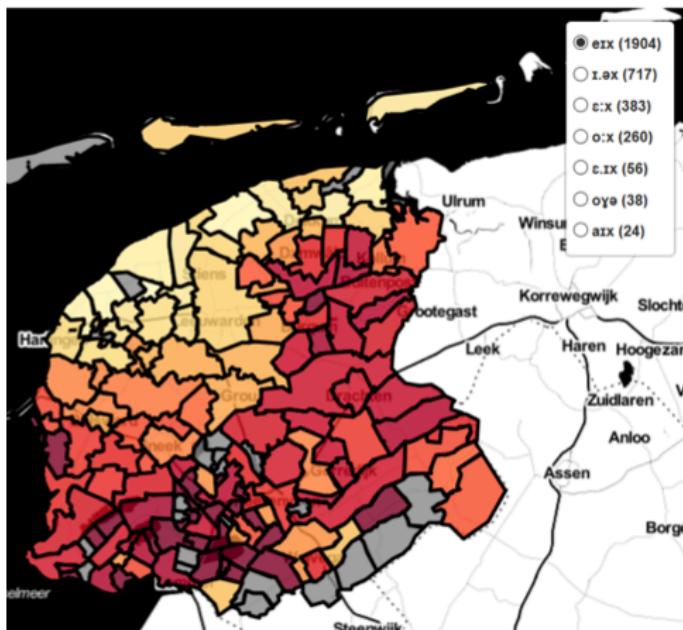


Figure 8: The regional distribution of the eech [e:x] response in the Stimmen Dialect Quiz, corresponding to the brown variant in Figure 7. Darker colours indicate higher reported usage frequency, grey areas indicate no reported usage.

#### 4.3 Variation in (r)

The last sociolinguistic variable for which we can compare previous data with that from the Stimmen project is (r). A highly noticeable sound change that has taken place in Dutch in recent decades has been the innovation and spread of variant [ɹ], the approximant bunched realisation of post-vocalic coda /r/. This is a feature popularly referred to as a ‘Gooise r’. In a large-scale study Sebregts (2015) concludes that age, region as well as gender predict the use of the approximant variant in Dutch, with female speakers using the highest proportion of the [ɹ], a finding supported by evidence in Van Bezooijen (2005). One interesting research question is whether, in highly bilingual populations, sociolinguistic variants are taken on also in a second, closely related language. This is what Van Bezooijen (2009) considers, studying pronunciation in a picture naming task eliciting coda /r/. In the data collected from 26 Frisian-speaking and 30 Town-Frisian subjects, there is no sign of convergence towards variation patterns found in Dutch. No approximant codas are found at all in her analysis of the realisations of ‘boer’ *farmer* and ‘gieter’ *water can*. The main variant of /r/ used in Fryslân is alveolar, with uvular variants present in respondents who speak Town Frisian.

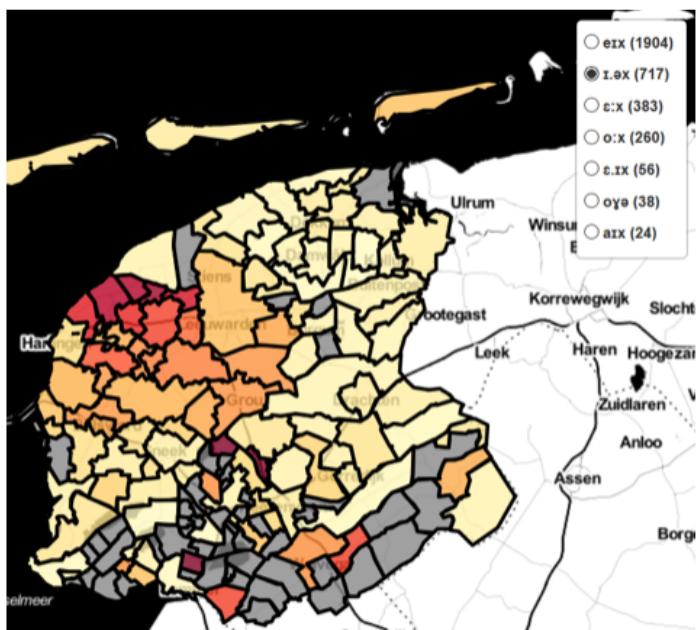


Figure 9: The regional distribution of each [i.eɪx] response in the Stimmen Dialect Quiz, corresponding to the blue variant in Figure 7. Darker colours indicate higher reported usage frequency, grey areas indicate no reported usage.

Bakker (2018) uses the spoken data collected in the Stimmen project labelled ‘Frisian’ to consider whether [ɿ] has gained ground in Fryslân a decade after Van Bezooijen (2009). Using auditory analysis and an additional coder he considers variation in the recordings of ‘ear’ *ear* from the Stimmen picture task and concludes that the usage of the approximant variant is very rare (1.6% of the cases), as can be seen in Table 3–3. He calls for more research, as he finds that the approximant variants are produced by young female speakers and discusses whether the Stimmen data could show the first stage of introduction of the variant to Frisian. For the purpose of the aim of the current paper, however, it is fair to conclude that a percentage of 1.6% is comparable to the finding that Van Bezooijen made, of negligible usage of the approximant variant in coda position in Frisian speech.

Note that Bakker (2018) attests a larger proportion of uvular variants (7.3%) in the Frisian Stimmen data than that attested in Van Bezooijen (2009; 0%). The difference between the two findings might be explained by the fact that the Stimmen data includes recordings from a much larger regional area than that included by Van Bezooijen (2009). In the Stimmen data, respondents living in areas in which

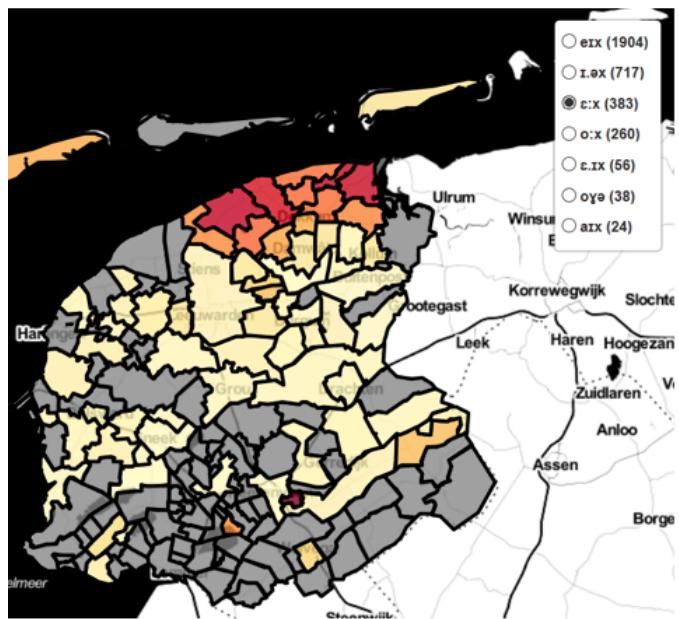


Figure 10: The regional distribution of the éch [ɛ:x] response in the Stimmen Dialect Quiz, corresponding to the green variant in Figure 7. Darker colours indicate higher reported usage frequency, grey areas indicate no reported usage.

Table 6: Variants of /r/ produced (as coded by Bakker 2018) in the Stimmen picture naming task ('ear'), and in the data reported by Van Bezooijen (2009).

	Bakker (2018) 'ear' in Stimmen data	Van Bezooijen (2009) 'boer' and 'gieter' in Frisian	Van Bezooijen (2009) 'boer' and 'gieter' in Town Frisian
alveolar	323 (87.5%)	52 (100%)	48 (80%)
uvular	27 (7.3%)	0 (0%)	12 (20%)
approximant	6 (1.6%)	0 (0%)	0 (0%)
'other'	13 (3.6%)	0 (0%)	0 (0%)
Total	369 (100%)	52 (100%)	60 (100%)

Town Frisian and Bildts are spoken (and where the uvular variant is widespread) feature in the data set. This could possibly have resulted in a higher proportion of uvular /r/ recordings, not only for the data labelled as “Town Frisian”, but also for the data labelled as “Frisian”. Bakker (2018) concludes that the most uvular variants in the Stimmen data are produced by subjects who state they come from Leeuwarden, the Sneek, or Franeker areas, which are all areas in which Town Frisian is spoken alongside Frisian and Dutch. The development of Town Frisian and its relationship to Frisian on a phonological level is a topic worthy of further research, on the basis of the findings in Bakker (2018).

## **5 Discussion and conclusion**

The aim of this paper has been to validate the use of crowd sourced language data from minority languages for linguistic research. Crowd sourcing is a possible means to increase available research data from such languages. Currently the amount of available data for research from such languages is minimal, something that was also true for the minoritised varieties in the north of the Netherlands before the initiation of the Stimmen project. Overall, Stimmen has been a successful attempt in gathering large amounts of speech data from minority language users, but primarily those in the Netherlands. The data collected in the project is predominantly from younger users but is well-balanced in terms of gender representation.

Previous considerations of validity of crowd sourced data indicates that quality can be comparable to that collected with traditional methods, as long as the task for the user is not too complex. In the Stimmen application, users were asked to record their own words for pictures in the Picture Naming Task, or to indicate which variant they use in a Dialect Quiz, made especially for inhabitants of Fryslân. When considering validity of a method one should ideally test the outcomes from the methodology against a larger test battery. However, the amount of research data available from minority communities, including the community in question in this paper, Frisian, gives a rather small basis of comparison. Yet, the relatively recent sociolinguistic work that is available for comparison, Hilton & Weening (2014), Stefan, Klinkenberg, and Versloot (fc.) and Van Bezooijen (2009) all display very similar results to those found in analyzes of the data from Stimmen. This indicates that the data collected through the Stimmen application is valuable resource for investigating linguistic variation.

The finding that crowd sourced data from minority languages can be used for research of variation in speech is welcome. In a world in which travel, and

contact, restrictions have become more common-place the opportunity to collect data remotely is a game changer. Mobile phones and other portable technology that allows contact with a speech community within a simple click should be taken into use as much as possible. The inclusion of data collected using smartphones and virtual games can lead to much wider representation, not only of different linguistic varieties, but also of more participants, in research. Future work in this field should focus on whether remote efforts also increases the participation of contested language users, such as new speakers, and how one can ensure that a representative sample of all age groups is included in remote studies.

## References

- Bakker, Wiljo. 2018. *The way you /r/*. Groningen: University of Groningen. () .
- Bonney, Rick, Tina B. Phillips, Heidi L. Ballard & Jody W. Enck. 2016. Can citizen science enhance public understanding of science? *Public Understanding of Science* 25(1). 2–16.
- Breuker, Pieter. 1992. Taalkundige skaaimerken fan normearre fryske. *Us Wurk* 41(1-2). 1–58.
- Brunelle, Marc. 2009. Diglossia and monosyllabization in eastern cham: A sociolinguistic study. *Variation in indigenous minority languages*. 47–75.
- Centraal Bureau voor de Statistiek. 2018. *CBS StatLine - regionale kerncijfers Nederland*. <https://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=70072ned&D1=0-88,292-293&D2=6,20-22&D3=21-23&VW=T> (1 January, 2019).
- De Decker, Paul & Jennifer Nycz. 2011. For the record: Which digital media can be used for sociophonetic analysis? *University of Pennsylvania Working Papers in Linguistics* 17(2). 51–59.
- De Haan, Germen J. 1997. Contact-induced changes in modern west frisian. *Us Wurk* 46(1-4). 61–89.
- De Vries, Nic J., Marelle H. Davel, Jaco Badenhorst, Willem D. Basson, Febe De Wet, Etienne Barnard & Alta De Waal. 2014. A smartphone-based asr data collection tool for under-resourced languages. *Speech communication* 56. 119–131.
- Dorian, Nancy. 1978. The fate of morphological complexity in language death: Evidence from East Gaelic. *Language* 54(3). 590–609.
- Feitsma, Tony, Els van der Geest, Frits van der Kuip & Irénke Meekma. 1987. Variations and development in frisian sandhi phenomena.

- Goeman, Ton, Johan Taeldeman & Piet van Reenen. 2017. *Mand/fand/gtrp-database, dialecttranscripts 1980–1995*. <http://www.meertens.knaw.nl/mand/database/>.
- Hilton, Nanna Haug. 2021. Stimmen: A citizen science approach to minority language sociolinguistics. *Linguistics Vanguard* 7(s1). 1–15.
- Hilton, Nanna Haug & Joke Weening. 2014. Poster presented at sociolinguistics circle – university of groningen. In *Is this standardisation? Variation in (sk) in Frisian*.
- Hoekstra, Erik & Marjo van Koppen. 2000. Het bildts als resultaat van fries-hollands taalcontact. *Amsterdamse Beiträge zur Älteren Germanistik* 54. 89.
- Horn, Alexander. 2019. Can the online crowd match real expert judgments? how task complexity and coder location affect the validity of crowd-coded data. *European Journal of Political Research* 58(1). 236–247.
- Kolly, Marie-José & Adrian Leemann. 2015. Dialäkt äpp: Communicating dialectology to the public—crowdsourcing dialects from the public. *Trends in phonetics and phonology. Studies from German-speaking Europe*. 271–285.
- Leemann, Adrian & Marie-José Kolly. 2013. *Dialäkt äpp*. <https://itunes.%20apple.%20com/ch/app/dialakt-app/id606559705> (16 January, 2014).
- Leemann, Adrian, Marie-José Kolly, Ross Purves, David Britain & Elvira Glaser. 2016. Crowdsourcing language change with smartphone applications. *PLoS one* 11(1). 1–25.
- Lind, Fabienne, Maria Gruber & Hajo G. Boomgaarden. 2017. Content analysis by the crowd: Assessing the usability of crowdsourcing for coding latent constructs. *Communication methods and measures* 11(3). 191–209.
- Möller, Robert & Stephan Elspäß. 2015. *21. atlas zur deutschen alltagssprache (ada)*.
- Nagy, Naomi. 2009. The challenges of less commonly studied languages. *Variation in indigenous minority languages, Impact: Studies in language and society* 25. 397–417.
- Nagy, Naomi & Miriam Meyerhoff. 2008. Poster presented at NWA2008. In *The love that dare not speak its name: The fascination with monolingual speech*.
- Nicholas, Joe. 1988. British language diversity surveys (1977–87): A critical examination. *Language and Education* 2(1). 15–33.
- Noglo, Kossi. 2009. Sociophonetic variation in urban Ewe. *Variation in indigenous minority languages*. 229–244.
- Nota, Amber, Nanna Haug Hilton & Matt Coler. 2016. Word and phrasal stress disentangled: Pitch peak alignment in Frisian and Dutch declarative structures. *Speech Prosody* 2016. 464–468.

- O'Rourke, Bernadette & Fernando Ramallo. 2013. Competing ideologies of linguistic authority amongst new speakers in contemporary galicia. *Language in Society* 42(3). 287–305.
- O'Shannessy, Carmel. 2009. Language variation and change in a north australian indigenous community. *Variationist approaches to indigenous minority languages*. 419–439.
- Sebregts, Koen. 2015. *The sociophonetics and phonology of Dutch r*. Leiden: LOT.
- Shameem, Nikhat. 1998. Validating self-reported language proficiency by testing performance in an immigrant fijians. *Language Testing* 15(1). 86–108.
- Shing, Han-Chin, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III & Philip Resnik. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 25–36.
- Sjölin, Bo. 1976. “*min Frysk*”: Een onderzoek naar het ontstaan van transfer en “code-switching” in gesproken Fries. Fryske Ynstitut aan de Ryksuniversiteit te Grins.
- Stanford, James N. 2016. A call for more diverse sources of data: Variationist approaches in non- English contexts. *Journal of Sociolinguistics* 20. 525–541.
- Stanford, James N. & Dennis Preston. 2009. *Variation in indigenous minority languages* (IMPACT: Studies in Language, Culture and Society 25). Amsterdam: John Benjamins Publishing.
- Steemers, Patrick, Marc Beijn, Dirk Reuser, Remco Gaykema, Erik Bookholt, Joel van Veen & Mike Odenhoven. 2017. *Deloitte global mobile consumer survey 2017: The Dutch edition*. (2 July, 2018).
- Stefan, Nika, Edwin Klinkenberg, Arjen Pieter Versloot, et al. 2014. Frisian sociological language survey goes linguistic: Introduction to a new research component. *Philologia Frisica Anno*. 240–257.
- TaalAtlas. 2015. *De Fryske taalatlas 2015, Fryskje taal yn byld*. Leeuwarden.
- Van Bezooijen, Renee. 2005. Approximant/r/in Dutch: Routes and feelings. *Speech communication* 47(1-2). 15–31.
- Van Bezooijen, Renee. 2009. The pronunciation of/r/in frisian. *Variation in Indigenous Minority Languages* 25. 299.
- van Bree, Cor. 1994. Het probleem van het ontstaan van het ‘stadsfries’ in verband met nieuwe talen in contact-theorieën. In *Predota, stanislaw (red.). handelingen regionaal colloquium neerlandicum wroclaw*, vol. 1993, 43–66.
- Vaux, Bert. 2004. Let's go usa: 2004. In chap. American dialects. Let's Go Publications.
- Weinreich, Uriel, William Labov & Marvin Herzog. 1968. *Empirical foundations for a theory of language change*. Vol. 58. University of Texas Press Austin.

Wenker, Georg. 1881. *Sprach-Atlas von Nord-und Mitteldeutschland: Text und Einleitung: Auf Grund von systematisch mit Hilfe der Volksschullehrer gesammeltem Material aus circa 30000 Orten.* Strassburg: Trübner.

# Chapter 8

## Complexity of endangered minority languages: The sound system of Wymysiöeryś

Alexander Andrason<sup>a</sup>

<sup>a</sup>Stellenbosch University

This paper demonstrates that Wymysiöeryś – a severely endangered Germanic minority language – exhibits remarkable complexity despite its moribund status. By analyzing twelve phonetic/phonological properties, the author concludes that the complexity of Wymysiöeryś is greater, both locally and globally, than that of two control languages: Middle High German and Modern Standard German. In most cases, the surplus of complexity attested is attributed to contact with the dominant language, Polish.

### 1 Introduction

Language endangerment, language shift, language obsolescence, and language death all have “considerable impact” on the *structure* of the languages affected (Palosaari & Campbell 2011: 110).<sup>1</sup> The most pervasive form of impact is the apparent simplification and impoverishment of the *grammar* of endangered and moribund languages (see Dorian 1973: 590–591, 1980: 85, Silva-Corvalán 1995: 9, Mesthrie et al. 2009: 256, Palosaari & Campbell 2011: 110–117, Sallabank 2012: 101, 111, 118, 2013: 126, Filipović & Pütz 2016: 2, Aikhenvald 2007: 43, Meakins et al. 2019: 294, 297) – whether it is phonetics/phonology, morphology, syntax,

---

<sup>1</sup>The present article emerged as a result of my PhD dissertation, *Polish borrowings in Wymysiöeryś: A formal linguistic analysis of Germano-Slavonic language contact in Wilamowice* (Andrason 2021).



or vocabulary (Dorian 1978: 591, 1980: 85, Palosaari & Campbell 2011: 110, 113–115).<sup>2</sup> Even though these reductive processes are especially patent and the most rampant in the varieties used by semi-speakers or rusty speakers, who do not learn the language fully in an intergenerational transmission and/or do not use it for the greater parts of their lives (see Palosaari & Campbell 2011, Grinevald & Bert 2011), simplification and impoverishment also seem to affect speakers whose language acquisition was uninterrupted and/or who have spoken the language relatively continuously. Overall, an endangered or dying language viewed as a *holistic* (though not uniform) *linguistic phenomenon* apparently reduces its complexity generation after generation (Austin 1986: 203; see also Dorian 1978, 1980, Swiggers 2007, Sallabank 2012, 2013, Palosaari & Campbell 2011, Filipović & Pütz 2016).<sup>3</sup> This especially occurs in destabilized or unbalanced types of language contact in which the endangered language is gradually displaced by the dominant code (Aikhenvald 2007: 47, Meakins et al. 2019).<sup>4</sup>

Simplification is a common phenomenon in language contact (McWhorter 1998, 2001, Matras & Sakel 2007). It is typical of pidgins (Mühlhäusler 1977, 1986, Trudgill 1986, 1996, 2010: 306–308, Kusters 2003, Siegel 2008: 190, Juvonen 2008: 321–322) and, albeit to a lesser extent and not without contest (DeGraff 2003, 2005, Ansaldi & Matthews 2007: 12–14, Hammarström 2008: 300, Bakker et al. 2011) of creoles (McWhorter 1998, 2001, 2005, Seuren & Wekker 2001, Parkvall 2008). The noticeable exceptions are mixed languages, which may maintain or even increase complexity (Matras 2009: 288, 305, Meakins et al. 2013, Velupillai

<sup>2</sup>Simplification typically implies reduction or loss of marked features (Palosaari & Campbell 2011: 113, Sallabank 2013: 126) due to regularization and overgeneralization (Silva-Corvalán 1995: 9–10, Palosaari & Campbell 2011: 113, Sallabank 2013: 126, Filipović & Pütz 2016: 2). The features that tend to be reduced or lost involve: phonological contrasts (Dorian 1980: 85, Palosaari & Campbell 2011: 113), morphological marking and distinctions (Dorian 1980: 85, Palosaari & Campbell 2011: 115, Meakins et al. 2019: 297), synthetic structures (which are replaced by analytic constructions) (Silva-Corvalán 1995: 10, Palosaari & Campbell 2011: 115, Meakins et al. 2019: 297), syntactic (Dorian 1980: 85, Palosaari & Campbell 2011: 115) and stylistic patterns (Dorian 1980: 85, Palosaari & Campbell 2011: 115), as well as vocabulary (Sallabank 2012: 118).

<sup>3</sup>It should be noted that many scholars speak about the simplification of endangered and dying languages *in general* (Dorian 1980: 85, Austin 1986: 203, Swiggers 2007: 24, Mesthrie et al. 2009: 256, Palosaari & Campbell 2011: 110, 112, Sallabank 2013: 118) rather than referring to the variety that is *only* used by semi-speakers. After all, it would not be surprising that, similar to imperfect L2 speakers, semi-speakers would not make use of the entire linguistic repertoire available in the language.

<sup>4</sup>Certainly, language endangerment and language contact are not the same phenomenon. However, although a language may die “without language shift” (Austin 1986: 201), most cases of language endangerment and language death “involve language *replacement* or shift” (Austin 1986: 201). This presupposes contact between the languages involved and bilingualism of endangered language speakers, at least at the population level.

2015: 301–307, 329–330, 402, Meakins et al. 2019: 296–297, 326–327) and layered languages, which may exhibit traces of both simplification and complexification (Aikhenvald 2007: 42–43).<sup>5</sup>

The present paper examines whether the severe endangerment of a language – specifically, a minority language that has nearly been replaced by another code and drifts towards a seemingly imminent death – is correlated with structural simplicity; or, inversely, whether severely endangered languages may exhibit remarkable complexity due to the transfer of elements from the dominant code. The language system under analysis is Wymysiöeryś – a colonial East Central German variety that has interacted with the dominant Polish language for more than seven centuries (Putschke 1980: 498, Wiesinger 1980: 497–498, Wicherkie-wicz 2003) and that, due to increasingly aggressive Polonization, has been, more or less, endangered since the end of World War I in 1918 (Neels 2016), currently finding itself on the verge of total extinction (Andrason & Król 2016).<sup>6</sup> I will study the linguistic repertoires of fluent speakers – the remaining 65 native Wymysiöeryś speakers born between 1913 and 1993 who have acquired the language in uninterrupted intergenerational transmission and have spoken it relatively continuously.<sup>7</sup> Inversely, the idiolects of semi-speakers – who do exhibit radical simplification and impoverishment processes but have no bearing on the transmission of Wymysiöeryś to the younger generations and thus the structure of the language as such – are not taken into consideration.

This study centers on the idea of absolute complexity (Kusters 2008, Dahl 2004, 2009, Miestamo 2008, 2009). To calculate absolute complexity, I deploy the concept of effective complexity (Gell-Mann 1995, Gell-Mann & Lloyd 2004) and take into account two main criteria: distinctiveness and economy (Miestamo 2006a, 2008, Sinnemäki 2008, 2009, 2011, Parkvall 2008). I focus on local complexities

---

<sup>5</sup> Although pre-pidgins, stabilized pidgins, expanded pidgins, creoles, and post-creole varieties are often simpler than the feeding languages (Velupillai 2015), they tend to – in the above order – gradually increase their complexity (Mühlhäusler 1986: 5–11, Parkvall 2008: 281, Trudgill 2010: 306, Velupillai 2015). Thus, both simplification and complexification are important factors in the development of pidgins and creoles (Heine & Kuteva 2005: 258, Trudgill 2010: 309), operating with distinct intensity at different stages of the pidgin-creole life cycle.

<sup>6</sup> Wymysiöeryś is classified as “nearly extinct”, in the Extended Graded Intergenerational Disruption Scale and “moribund” or “severely endangered” by Moseley & Nicholas (2010). The language is currently used by less than fifty elderly native speakers (Ritchie 2016: 73, Chromik 2016: 91) whose number rapidly decreases every year (Andrason & Król 2016, Andrason 2021).

<sup>7</sup> The number of these speakers was 65 at the beginning of the 21<sup>st</sup> century, when I started my research on Wymysiöeryś. Unfortunately, many of them have passed away in the interim (cf. footnote 6 above). For the list of these speakers, their names, and dates of birth see Andrason (2021).

pertaining to twelve distinct phonetic/phonological features, subsequently combining them into a global value representing the complexity of the entire sound-system module (Miestamo 2006a,b, 2009, Deutscher 2009, Sinnemäki 2014). The complexity of Wymysiöeryś, both feature-locally and module-globally, will not be quantified autonomously, but will rather be narratively estimated in relation to two control systems: (a mother-system) Middle High German and (a sister-system) Modern Standard German (cf. Deutscher 2009 and Dahl 2009). In instances where Wymysiöeryś exhibits a surplus of information – i.e. a positive difference in complexity when compared to the control languages – I will examine whether this surplus can be attributed to Polish influence, either in whole or in part.

The article will be organized as follows: in Section 2, I will explain the theoretical framework underlying my study or the manner of complexity measurement adopted. In Section 3, I will compare the complexities of Wymysiöeryś with those of Middle High German and Modern Standard German – first locally and next globally. In Section 4, I will verify whether, and how intensively, the surplus of information attested in Wymysiöeryś draws on Polish – again, first locally and next globally. In Section 5, I will draw conclusions and propose lines of future research.

## 2 Framework – measuring language complexity

The approach adopted in this paper is one of the most common and theoretically least problematic manners of analyzing the complexity of natural languages. It draws on the idea of complexity that is: (a) epistemologically absolute, (b) computationally effective, (c) built around the criteria of distinctiveness and economy, (d) relational, and (e) primarily local. Below, I explain these five ideas in detail.

1. From an epistemological perspective, the type of complexity analyzed in this research is absolute. This complexity type pertains to the system viewed as “an autonomous entity” in disconnection from the observer (Kusters 2008: 4). Absolute complexity is therefore the “objective property of the [language] system” (Miestamo 2008: 23, Dahl 2004), regardless of the characteristics of its users, whether speakers, hearers, first-language or second-language learners.<sup>8</sup>

---

<sup>8</sup>Inversely, I am not concerned with relative complexity or complexity experienced by users of a given language (see Kusters 2003, Miestamo 2008: 23, 2009: 81–82). Some scholars do not regard relative complexity as complexity *sensu stricto* (Dahl 2004: 39–40) instead preferring the terms “difficulty” and “cost” (Miestamo 2008: 27, Lindström 2008).

2. To compute absolute complexity, I will deploy the concept of effective complexity otherwise referred to as Gell-Mann complexity (Gell-Mann 1995, Gell-Mann & Lloyd 2004). This complexity measure refers to the amount of information necessary to describe non-randomness within a system – the longer the description of regularities is, the more complex a system is (Mitchell 2009: 99). In linguistics, this corresponds to the amount of information required to describe rules governing a language (McWhorter 2005, 2007, 2009, Miestamo 2008: 25, 2009: 81–82, Sinnemäki 2008, 2009, 2011, Parkvall 2008).<sup>9</sup>
3. In the quantification of effective complexity, I will take into consideration two main criteria: a distinctiveness criterion and an economy criterion (Miestamo 2006a,b, 2008, Sinnemäki 2008, 2009, 2011, Parkvall 2008). According to the distinctiveness criterion, the complexity of a language increases with a greater categorical diversity, i.e. with more distinctions being made and more domains being overtly specified. According to the economy criterion, complexity increases with a greater formal diversity, i.e. with more manners of encoding of a given category or distinction.<sup>10</sup>
4. The degree of complexity of Wymysiöeryś will be estimated in relation to the complexities of two diachronic and dialectal control-systems (cf. Deutscher 2009 and Dahl 2009): Middle High German and Modern Standard German. Middle High German is regarded as a non-distant ancestor of Wymysiöeryś or its mother-language (Kleczkowski 1920, Wiesinger 1980: 496, 498, 1983: 911, Morciniec 1984, Zieniukowa & Wicherkiewicz 1997: 308, 2001: 492–493, Wicherkiewicz 1998: 200, Wicherkiewicz & Olko 2016: 19, Źak 2016: 132).<sup>11</sup> Modern Standard German is a closely related West Germanic sister-language (Chromik & Dolatowski 2013). Neither of the control systems has experienced intense influence from Polish or other Slavonic languages.

<sup>9</sup>The other common manner of quantifying complexity is algorithmic complexity, also referred to as Kolmogorov complexity. This manner of quantification calculates disorder or randomness, e.g. the “amount of surprise” contained in a message (Mitchell 2009: 97–98) or the “difficulty of description” (Shalizi 2006: 52).

<sup>10</sup>Thus, synonymy, redundancy, allomorphy, and free variations increase complexity (McWhorter 2007, 2008). Similarly, exceptions contribute to the increase in complexity as they constitute additional rules (cf. Hammarström 2008: 29; see also McWhorter 2007, 2008).

<sup>11</sup>The selection of Middle High German and Modern Standard German as pre-contact and non-contact “control” languages, also stems from the availability of extensive and detailed grammatical studies dedicated to these languages, which render the comparison fully operational.

5. The estimation of effective complexity will mainly be conducted at a local level. I will determine how many of the selected phonetic/phonological features (categories) are instantiated in the tested languages (distinctiveness) and how many expression manners of each feature there are (economy).<sup>12</sup> With regard to each feature, the analysis will be captured in a narrative form “drawing on the concepts of (approximate) equality  $\approx$  ( $x \approx y$  means that  $x$  is approximately equal to  $y$ ), inequality  $\leq$  ( $x \leq y$  means that  $x$  is less than or equal to  $y$ ), [and] strict inequality  $<$  ( $x < y$  means that  $x$  is less than  $y$ )” (Andrason et al. forthcoming).<sup>13</sup> However, the analysis of complexity will not be limited to separated local domains. On the contrary, I will combine the local complexities into a global value that indicates the complexity of the entire sound-system module of each of the three languages.

The approach used in this paper has its limitations, which are inherent to and unavoidable in complexity studies. To begin, contemporary science lacks a single, comprehensive, all-purpose complexity measure. Instead, a variety of measurement methods – at least 48 as observed by Edmonds (1999) – coexist simultaneously, differing in what is calculated and how the calculation is executed. Crucially, the various methods yield different complexity results (for an overview consult Peliti & Vulpiani 1988, Badii & Politi 1997, Rescher 1998, Edmonds 1999, Shalizi 2006, and Mitchell 2009). Similarly, linguists have not reached an agreement as for how language complexity should be measured and compared (Newmeyer & Preston 2014: 7). Virtually every scholar develops an at least minimally different method of measurement. However, neither this disagreement nor the excessive proliferation of measurement techniques is surprising. They rather reflect the fact that natural languages constitute genuine com-

---

<sup>12</sup>The features analyzed in here are principally phonetic features. Therefore, I will use square bracket notation when writing about the sounds of Wymysiöeryś and the two control languages. However, I will occasionally refer to phonology as well. This phonetic orientation is motivated. First, scholarship still lacks a phonological analysis of the Wymysiöeryś language – all descriptions being virtually phonetic. Second, more generally, phonology is much more theory-dependent than phonetics (see that, in Polish, under certain theoretical premises, [i] and [ɔ/i] are treated as a single phoneme despite being clearly distinct from an articulatory perspective and being indeed perceived as two distinct vowels by native speakers). Accordingly, all the inventories from other authors are phonetic (for instance, with respect to Standard Modern German see Johnson & Braber 2008, Fagan 2009, O’Brien & Fagan 2016; see also Eisenberg 1994, Dodd et al. 2003: 352) unless stated otherwise (for Modern High German see Russ 1994, Wiese 1996, and Fox 2005).

<sup>13</sup>Given the narrative approach adopted, the symbols  $\approx$ ,  $\leq$ , and  $<$  reflect three types of relationships between  $x$  and  $y$ : similarity, minimal difference, and substantial difference (cf. Andrason et al. forthcoming).

plex systems, which renders any modeling fragmentary, provisional, and tentative, irrespective of how potent and sophisticated it is (Cilliers et al. 2013: 3, Andrason & Król 2016). With regard to the method adopted in this paper, a number of objections could be raised. First, absolute complexity is theory-oriented and, perhaps, heavily theory-dependent (Miestamo 2008: 24, Kusters 2008: 5, 8, Hammarström 2008: 289). It depends on the theories of language in which the description of grammar and its analysis are conducted (Miestamo 2008: 27, Kusters 2008: 5, 7–8, Hammarström 2008: 289). Second, local complexity allows for various manners of granularity. That is, each local complexity can be fragmentized into more atomic types, which inversely means that each local complexity constitutes global complexity from the perspective of more fragmentary levels of analysis. As a result, local but modular complexities are not free from three further problems typical of global complexity: sampling problems, commensurability problems, and modularity problems. Third, given the quantitative depth of a language, the account of all the details included in a single module and their quantification are unfeasible, both theoretically and practically (Miestamo 2006b: 30, 2008). Linguists rather determine a more or less restrictive sample of studied phenomena, limiting themselves to analyzing a set of distinctions or categories – rather than all of them. No universal methods of such delimitation exist and often detail-ness is determined by utilitarian aspects such as the aim of the analysis to be undertaken, the (maximal) feasibility of the research, and the scientist's theoretical paradigm (Deutscher 2009: 248). Fourth, even for modular complexity, it is uncertain how to commensurate the different features found in a single module and represent them in identical numerical terms, since each such term refers to qualitatively different phenomena and encapsulates, in principle, incomparable properties (see Miestamo 2006a,b, 2008: 30, 2009: 83). Fifth, dividing the module into distinct features or categories presupposes the highly problematic division of language into separate parts – for which local complexities are subsequently calculated. All those limitations – of which linguists are well aware – render the measurement of local complexities and that of modular complexity (i.e. global complexity at a module level), as well as their comparison across languages extremely difficult, if not elusive. My method partially responds to these problems. To mitigate theory-dependence, I am eclectic with regard to theory underlying the description of features. By taking into account a number of studies that follow different theoretical principles, I purposefully average the complexities inferred from the available descriptions. To mitigate granularity and sampling problems, the categories selected coincide with categories distinguished in general phonetic/phonological studies and with phonetic/phonological features usually described in comparative works on the Germanic and Slavonic language

families (Rothstein et al. 1993, Jacobs et al. 1994, Sussex & Cubberley 2006, Harbert 2007). Lastly, to mitigate the commensurability problem, I use a narrative method instead of a strictly numerical one. The problem of modularity cannot be mitigated, as the division of the sound-system module into separated units is presupposed by the method used in my study.

Additionally, there are two problems related to the control systems used in estimating the changes in complexity of Wymysiöeryś. First, the data related to the sound systems of Middle High German and Modern Standard German are secondary and draw on the studies presented by other scholars. Although all these studies are generally recognized in Germanic scholarship as authoritative, the information presented in them need not be exhaustive. This especially applies to Middle High German as this language may have contained some features that have not been reported thus far. The results of my comparison of Wymysiöeryś with Middle High German are inevitably contingent on the other linguists' views of Middle High German and their interpretation of direct textual evidence. This type of risk is unavoidable in typological and diachronic studies when one must rely on others' data and analyses. Second, Middle High German itself is an umbrella term that encompasses a number of German varieties that (a) were spoken between the 11<sup>th</sup>/12<sup>th</sup> and 14<sup>th</sup>/15<sup>th</sup> centuries; (b) were successors of Old High German; and (c) entirely or partially underwent the second consonant shift. This means that the grammatical repertoire of Middle High German is possibly richer than the repertoire of any single German variety that was spoken between the 11<sup>th</sup>/12<sup>th</sup> and 14<sup>th</sup>/15<sup>th</sup> centuries. To put it simply, while Wymysiöeryś is a single variety of East Middle German, Middle High German is a conglomerate of many varieties. Again, this problem is typical of studies that compare modern languages with old, classical, and extinct languages (see Andrason et al. forthcoming).

### 3 Evidence

The Wymysiöeryś language described below is a heterogenous and internally diversified linguistic system. Most sources employed make reference to Wymysiöeryś spoken at the time of its near extinction, i.e. at the end of the 20<sup>th</sup> and the beginning of the 21<sup>st</sup> century (Lasatowicz 1994, Wicherkiewicz 1998, 2003, Zieciukowa & Wicherkiewicz 2001, Ritchie 2012, Weckwerth 2015, Źak 2016, 2019). When describing this modern Wymysiöeryś variety, I will widely draw on an original database developed over the course of more than fifteen years of field-work activities conducted by Tymoteusz Król and myself. This primary evidence

has been the foundation of several of the papers that I have published alone or in collaboration with Król (in particular Andrason 2014b,a, 2015, Andrason & Król 2016). I will refer extensively to the findings of those studies.<sup>14</sup> Additionally, the data presented will draw on three works dedicated to pre-war Wymysiöeryś – the stage at which true endangerment had already begun, even though it was still less pronounced than it is currently (Kleczkowski 1920, 1921, Mojmir 1930–1936, and Wicherkiewicz 2003, who describes the language used by Florian Biesik in his poems, the majority of which were written between 1920 and 1924). The evidence related to Middle High German and Modern Standard German is secondary and draws on canonical studies dedicated to these two languages and the (West) Germanic linguistic family. In particular, regarding Middle High German: Wright (1917), de Boor & Wisniewski (1973), Simmler (1985), Paul (2007), Hennings (2012), and Hall (2017). Regarding Modern Standard German: Hall (1992, 2000), Russ (1994), Eisenberg (1994), Wiese (1996), Dodd et al. (2003), Fox (2005), Johnson & Braber (2008), Fagan (2009), Caratini (2009), and O'Brien & Fagan (2016), and generally (West) Germanic: Iverson & Salmons (1995, 1999, 2003, 2008), Goblirsch (1997, 2018), Harbert (2007), van der Hoek (2010), and van Oostendorp (2019).

In this section, I will first determine the value of the twelve local complexities of Wymysiöeryś in relation to the two control systems (Section 3.1 – Section 3.12). Next, I will estimate the relational complexity of the three languages at the global level of the sound-system module (Section 3.13).

### 3.1 Monophthongs

Depending on the study, the number of monophthongs in Wymysiöeryś varies. The highest number attested is 15. The lowest number is 9. Most analyses distinguish well above ten vowels. Maximal systems have been proposed by Andrason & Król (2016), Kleczkowski (1920) and Mojmir (1930–1936). Andrason & Król (2016: 20) identify fifteen core vowels: [i], [ɪ], [e], [ɛ], [a], [ɑ], [o], [ɔ], [u], [y], [v], [ø], [ø], [œ], and [ə]. Kleczkowski (1920: 11–12, 171) and Mojmir (1930–1936: xii–xiii) recognize thirteen vowels: [i], [ɪ] [e], [ɛ], [a], [ɑ], [o], [ɔ], [u], [v] [y], [ø], [ə].<sup>15</sup> Systems of twelve monophthongs are proposed by Lasatowicz (1994), Zie-

<sup>14</sup>The informants who have participated in the empirical research conducted by Król and myself, and whose language is reflected in previously published articles as well as in this paper, are (or were) fluent native speakers of Wymysiöeryś (see Section 1).

<sup>15</sup>The phonetic interpretation in IPA terminology is mine. Kleczkowski (1920) offers detailed descriptions which allow for such an interpretation. Given the limitations in space, I will not provide examples of phonetic/phonological features in words and/or constructions. These may be found in the works referred to in each section.

niukowa & Wicherkiewicz (2001), Wicherkiewicz (2003), and Andrason (2014a). Lasatowicz (1994: 32–41) distinguishes: [i], [e], [ɛ], [a], [ɑ], [o], [ɔ], [u], [y], [v], [ø], [ə]. Zieniukowa & Wicherkiewicz (2001: 499–500) distinguish: [i], [e], [ɛ], [a], [ɑ], [o], [u], [y], [v], [ø], [ə], and [i]. Wicherkiewicz (2003: 407) distinguishes the same set of vowels merely replacing [o] with [ɔ]. Andrason (2014a: 126–127) distinguishes: [i], [ɪ], [e], [ɛ], [a], [ɑ], [o], [ɔ], [u], [y], [v], and [ø]. The most reduced systems of monophthongs are postulated by Weckwerth (2015) and Ritchie (2012) who discern nine vowels: [i], [ɪ], [e], [ɛ], [ø], [a], [ɑ], [ɔ], and [u].<sup>16</sup>

The vocalic inventory of Middle High German consists of at least nine basic short monophthongs: *a* [a], *e* [e], *ē* (transcribed as [ē]), *ä* [ɛ], *i* [i], *o* [o], *ö* [ø], *u* [u], and *ü* [y] (Wright 1917: 2–5, Simmler 1985: 1131, 1133, de Boor & Wisniewski 1973: 36–37, 41, Paul 2007: 62–63, 87–97, Hall 2017: 9, Schmidt 2017: 69).<sup>17</sup> Additionally, the reduced vowel *e* [ə] was used in unaccented syllables (Wright 1917: 3, Simmler 1985: 1133, Hall 2017: 9). The language also had long vowels (see Section 3.5 below), of which some may have exhibited slightly different qualities when compared to their short counterparts, apart from distinctive quantity (Wright 1917: 3, Caratini 2009: 185; cf. Hall 2017: 9 and Schmidt 2017: 69). This would increase the total number of monophthongs to maximally eighteen vowels. The vocalic system of Modern Standard German includes 15 vowels of different qualities: [i:], [ɪ], [y:], [v], [e:], [ø:], [ɛ(:)], [œ], [u:], [ʊ], [o:], [ɔ], [a(:)], [ə] and [ɐ] (Fagan 2009: 7, 17, Johnson & Braber 2008: 109–110, O'Brien & Fagan 2016: 17–19; see also Russ 1994: 119, Wiese 1996: 19–21, and Fox 2005: 35, 41 who analyze the vocalic phonemes of Modern Standard German). In some studies, additional sounds [ɑ] (Eisenberg 1994: 350, Dodd et al. 2003: 3, Johnson & Braber 2008: 110, Caratini 2009: 71) and [æ:], as different in quality from [ɛ], are distinguished (Fox 2005: 35–36, 38).<sup>18</sup>

Overall, monophthongs (*M*) are typical features of both Wymysiöeryś and the two control languages. Quantitatively, the number of monophthongs in Wymysiöeryś is (with minor disturbances) similar to that of Modern Standard German and Middle High German. Therefore, the complexity of monophthongs in the three languages can be viewed as approximately equal:  $M^W \approx M^{MHG}$  and  $M^W \approx M^{MSG}$ .<sup>19</sup>

<sup>16</sup>Lasatowicz (1994) limits his set of vowels to eight sounds of an uncertain phonetic interpretation: *i*, *e*, *o*, *ö*, *ø*, *u*, *a*, *y*. Given the lack of linguistic training of its author, this system cannot be regarded as trustful (see a similar observation in Wicherkiewicz 2003).

<sup>17</sup>On the status of the vowels *e* (closed), *ē* (mid-open), and *ä* (open), consult Simmler (1985: 1132, 1134; see also Wright 1917: 2–5, de Boor & Wisniewski 1973: 36–37, 41, Paul 2007: 87–91). In a few studies, the number of vowels is reduced to seven (Caratini 2009: 184–185).

<sup>18</sup>Note that Fox (2005) analyzes phonemes. Sporadically, in unassimilated English borrowings, one finds two additional vowels, namely [æ] and [ʌ] (Fox 2005: 53, Caratini 2009: 73).

<sup>19</sup>The following abbreviations will be used in this paper: *W* – Wymysiöeryś; *MHG* – Middle High German; *MSG* – Modern Standard German.

### 3.2 Diphthongs

The number of Wymysioeryś diphthongs varies between nine and six.<sup>20</sup> Kleczkowski (1920: 12–13) and Mojmir (1930–1936: xiii–xiv) distinguish nine diphthongs: [ie:], [iɛ:], [aj], [oj], [ə(:)j], [ou], [au], [uø:] / [u:ø], and [u:ə].<sup>21</sup> Systems of eight diphthongs are proposed by Lasatowicz (1994), Wicherkiewicz (2003), and Zieniukowa & Wicherkiewicz (2001). Specifically, Lasatowicz (1994: 33, 40–41) distinguishes: [i(:)ɔ̃], [ẽi], [æ̃], [aõ], [ɔ̃ø̃], [ø(:)ɔ̃], [y(:)ɔ̃], and [u(:)ɔ̃]. Wicherkiewicz (2003: 407–408) distinguishes: [i(:)ɔ̃], [ẽi], [aẽ], [aõ], [ɔ̃ø̃], [ø(:)ɔ̃], [y(:)ɔ̃] and [uɔ̃]. Zieniukowa & Wicherkiewicz (2001: 499–500) distinguish: [aĩ], [yẽ], [ĩĩ], [ĩj̃], [aũ], [ø̃ø̃], and [uẽ]. Systems of six diphthongs are formulated by Andrason & Król (2016) and Weckwerth (2015). Andrason & Król's system (2016: 21) contains [aĩ], [eĩ], [ɔ̃ĩ], [œ̃ỹ], [ĩø̃] and [ĩɔ̃], while Weckwerth's system (2015: 1) includes [aĩ], [eĩ], [ɔ̃ỹ], [ø̃ø̃], [ỹø̃] and [ĩø̃].

The system of diphthongs in Middle High German tends to consist of six sounds: *ei*, *ou*, *öu* (*öi*, *öü*), *ie*, *uo*, and *üe* (Wright 1917: 2–3, 5–6, 17, de Boor & Wisniewski 1973: 39–41, 45, Simmler 1985: 1133, Paul 2007: 62–63, 103–108, Hall 2017: 9, Schmidt 2017: 69). Their phonetic interpretation is most likely [ẽi], [oũ], [ø̃ỹ], [ĩẽ], [ũõ], and [ỹẽ], respectively (Paul 2007: 103–108, Hall 2017: 9). Some linguists expand this system to nine sounds, adding the diphthongs *au* [aʊ̃], *eu* [œ̃ỹ]/[ɔ̃ĩ], and *ui* [ur̃] (Caratini 2009: 185). In Modern Standard German, the number of genuine diphthongs has decreased to three: [aĩ], [aʊ̃], and [ɔ̃ĩ] (Eisenberg 1994: 354, Dodd et al. 2003: 4, Johnson & Braber 2008: 112–113, Fagan 2009: 9, O'Brien & Fagan 2016: 19). However, a new wave of diphthongs has also emerged due to the vocalic pronunciation of *r* as [r̃] and the assimilation of loanwords ending in *-ion* and *-ation*. As a result, the inventory of diphthongs has been enriched by [ĩã], [ũã], [ẽã], and [ĩõ] (Eisenberg 1994: 354, Fox 2005: 53–54, Fagan 2009: 9).<sup>22</sup>

Overall, both Wymysioeryś and the two control languages contain diphthongs (*D*). Moreover, quantitatively, the number of diphthongs found in Wymysioeryś, on the one hand, and those present in Middle High German and Modern Standard German, on the other hand, are similar. As a result, the complexity of diphthongs is approximately equal in the three languages:  $D^W \approx D^{MHG}$  and  $D^W \approx D^{MSG}$ .

<sup>20</sup>The largest set of fourteen diphthongs is posited by Latosiński (1909: 270): *ia*, *iu*, *iy*, *ie*, *uö*, *uy*, *oe*, *oi*, *ou*, *ei*, *ae*, *aei*, *au*, *yi*. (cf. Wicherkiewicz 2003: 411). As in Section 3.1, I will disregard this system in my discussion.

<sup>21</sup>This system could be extended to ten sounds if [uø:] and [u:ø] are viewed as different diphthongs.

<sup>22</sup>Two additional diphthongs, i.e. [er̃] and [oʊ̃], may appear in unadjusted English loanwords (Caratini 2009: 73).

### 3.3 Triphthongs

The Wymysioeryś vocalic system contains one true triphthong, i.e. a sequence of three vocalic elements (usually glide, vowel, and glide) used within the same nucleus. This triphthong is [χøœ], noted alternatively as [χ'øœ] or [χ'øə] (Andrason 2014a: 126, Andrason & Król 2016: 23).

True triphthongs are absent in Middle High German (Hall 2017: 10). Neither Wright (1917), de Boor & Wisniewski (1973), Simmler (1985), Paul (2007), nor Schmidt (2017) mention them in their grammatical analyses. Similarly, there are no true triphthongs in Modern Standard German. Accordingly, they do not feature in works dedicated to German phonetics (see Eisenberg 1994, Johnson & Braber 2008, Fagan 2009, and O'Brien & Fagan 2016) and phonology (see Russ 1994, Wiese 1996, and Fox 2005).

Overall, the category of triphthongs ( $T$ ) is only instantiated in Wymysioeryś. Hence, by definition, the complexity of triphthongs in the two control languages is strictly lower than in Wymysioeryś:  $T^W > T^{MHG}$  and  $T^W > T^{MSG}$ .

### 3.4 Vocalic sonorants

A syllable – and in particular its nucleus – can be formed in Wymysioeryś not only by genuine vowels, but by sonorants as well. Five sonorants can be syllabic or vocalic: [l], [r], [n], [m], and [ŋ] (Kleczkowski 1920: 12, Mojmir 1930–1936: xiii). By far, the most common are [l] and [n] (Andrason 2021).

The sound system of Middle High German most likely lacked syllabic or vocalic sonorants, as the Proto-Indo-European [ɳ], [ɳ̥], [l], and [r̥] developed a full vocalic component, namely *u*, *o*, *ü*, or *ö* (from Old High German *o* and *u*, and an earlier Germanic \**u*) (Paul 2007: 62, 66, 146–150; see also de Boor & Wisniewski 1973: 48). A possible case, where the presence of syllabic sonorants has been hypothesized, involves the metathesis of *ɔr* into *rɔ* through an ‘*r*’ stage (Paul 2007: 147). Even if a few (highly debatable) instances of syllabic sonorants could be hypothesized, the relevance of such sounds was minimal for the vocalic system of Middle High German. In contrast, syllabic liquids ([l] and [r̥]) and syllabic nasals ([n], [m], [ŋ]) are a common feature of Modern Standard German, where they arose due to the reduction of *shwa* (Fagan 2009: 24, 32, Johnson & Braber 2008: 130–131, O'Brien & Fagan 2016: 18).

Overall, the category of syllabic/vocalic sonorants ( $S$ ) is instantiated in Wymysioeryś and one of the control languages. It is present in Modern High German but absent in Middle High German. The number of syllabic/vocalic sonorants

found in Wymysioeryś and Modern Standard German are quantitatively similar. Therefore, the complexity of syllabic/vocalic sonorants is approximately equal in these two languages:  $S^W \approx S^{MSG}$ . In contrast, the complexity relationship between Wymysioeryś and Middle High German is that of strict inequality:  $S^W > S^{MHG}$ .

### 3.5 Vocalic length

Most studies recognize the presence of long vowels and the significance of vocalic length in Wymysioeryś. Their most fervent advocates are Kleczkowski (1920) and Mojmir (1930–1936). Kleczkowski (1920: 174) argued that, contrary to contemporary Polish, length was a part of the vocalic system of Wymysioeryś, and that the opposition between short and long vowels was fundamental from a systemic perspective (Kleczkowski 1920: 11–12, 26–27). In fact, Kleczkowski (1920: 26) distinguishes four grades of length, splitting long and short vowels into two subclasses each: extra-long and long, on the one hand; and middle and short, on the other hand. In his model, the following monophthongs can be long or extra-long: [i:], [e:], [ɛ:], [a:], [ɑ:], [ɔ:], [ʊ:], [y:] (Kleczkowski 1920: 11–12, 26–27). More reduced systems of long vowels are discerned by Lasatowicz (1994), Zieniukowa & Wicherkiewicz (2001), and Wicherkiewicz (2003). Lasatowicz (1994: 40–41) distinguishes seven long vowels: [i:], [a:], [e:], [o:], [y:], [ø:], [u:]. Zieniukowa & Wicherkiewicz (2001: 499–500) distinguish six long vowels: [i:], [e:], [a:], [u:], [y:], and [ø:]. Wicherkiewicz (2003: 405–407) distinguishes five long vowels: [i:], [a:], [e:], [u:], [y:], and [ø:]. The relevance of vocalic length is also acknowledged by Andrason & Król (2016: 27–28). In contrast, Weckwerth (2015: 1–2) suggests that vocalic length, even though present, has a limited distinctive role, as the main contrast between vowels involves quality rather than quantity.<sup>23</sup>

Length was a relevant and contrastive feature of the vocalic system of Middle High German (Simmller 1985: 1133, Seiler 2005). In addition to short singletons (see Section 3.1), the language possessed eight long vowels: â [a:], æ [æ:], ê (ie) [e:], î [i:], ô (uo) [o:], oe (üe) [ø:], û [u:], and iu [y:] (Wright 1917: 2–7, de Boor & Wisniewski 1973: 37–38, 41, Simmler 1985: 1133, Paul 2007: 62–63, 97–100, Hall 2017: 9, Schmidt 2017: 69). As mentioned in Section 3.1, the short and long vowels likely exhibited at least minimal qualitative differences (Wright 1917: 3,

---

<sup>23</sup>My recent studies and fieldwork in Wilamowice show that long vowels are a constant feature of Wymysioeryś, although contrast between short and long monophthongs often involves at least minimal changes in quality. See words like *fool* [fo:w] ‘grey, gray-haired’, or the pair *hoon* [ho:n] ‘rooster’ versus *hon* [hən]/[hon] ‘roosters’.

Caratini 2009: 185, Hall 2017: 9, and Schmidt 2017: 69). Length also plays an important role in Modern Standard German, where vowels differ both in quality and quantity (Fagan 2009: 7–9, O’Brien & Fagan 2016: 17, 19; see also the short and long phonemes in Russ 1994: 118–119 and Wiese 1996, Fox 2005: 39–41).<sup>24</sup> Length is responsible for the systemic division of all vowels into ‘lax’ and ‘tense’. Lax vowels are always short, while tense vowels are long in a stressed position. The tense long vowels comprise eight sounds: [i:], [e:], [ɛ:], [a:/ɑ:], [o:], [u:], [y:], and [ø:] (Eisenberg 1994: 350–352, Johnson & Braber 2008: 109–110, Fagan 2009: 7–9, O’Brien & Fagan 2016: 17, 19; see also Russ 1994: 118–119, Wiese 1996: 19–21, Fox 2005: 38–39, 41, and their discussion of long vowel phonemes) sometimes expanded by an additional vowel [æ:] (Fox 2005: 35–36, 38).<sup>25</sup>

Overall, the category of vocalic length (VL) is instantiated in both Wymysiöeryś and the two control systems. The number of long vowels in the two groups is similar. As a result, the complexity of vocalic length in Wymysiöeryś, Middle High German, and Modern Standard German is approximately equal:  $VL^W \approx VL^{MHG}$  and  $VL^W \approx VL^{MSG}$ .

### 3.6 Nasality

The Wymysiöeryś vocalic system is characterized by the presence of nasal and/or nasalized vowels. The most common nasal sounds are the vowels [ɔ] and ɛ [ɛ] and three nasal approximants [w̩], [ɥ̩], [j̩] that can accompany oral vowels (Andrason 2021). Sporadically, vowels [ã/ɑ] and [é] are used (Andrason 2021). The actual nasal feature, and thus the nasalization of a vowel, may vary from strong (a genuine nasal vowel) to weak (a slightly nasalized oral vowel with a nasal consonant; see Kleczkowski 1920: 12, Mojmir 1930–1936: xiii).

Nasal vowels were absent in Middle High German (Wright 1917, Paul 2007, de Boor & Wisniewski 1973). Their position in Modern Standard German is similarly weak. In general, “German vowels are oral [and no] nasal vowel belongs to the core vocalic system” (Caratini 2009: 71). Nasalized vowels – [ɛ(:)], [ɔ(:)], [ɑ(:)], [œ(:)] – are only found in loanwords from French (Fagan 2009: 9–10, Caratini 2009: 51, 73–74, O’Brien & Fagan 2016: 22). Even there, however, a pronunciation with an oral vowel and a nasal consonant is grammatical (Russ 1994: 78, Fox 2005: 53, Fagan 2009: 9, O’Brien & Fagan 2016: 22). Being “unstable” and restricted to a small number of words of foreign origin, the role of nasal vowels in the vowel

<sup>24</sup>That is, both quality and length are used to differentiate sounds and ultimately lexemes (Eisenberg 1994: 352, Dodd et al. 2003: 3, Johnson & Braber 2008: 109, Fagan 2009: 7–9, O’Brien & Fagan 2016: 17–19).

<sup>25</sup>As always, Fox’s (2005) analysis concerns phonology.

system of Modern Standard German is marginal (Fox 2005: 53, Fagan 2009: 10, Johnson & Braber 2008: 90).

Overall, the category of nasality ( $N$ ) is instantiated in Wymysiöeryś and one of the control languages, i.e. Modern Standard German. However, the number of nasal vowels and their systemic relevance is greater in Wymysiöeryś than in Modern Standard German.<sup>26</sup> Hence, the complexity relationship between these two languages is of a strict-inequality type, i.e.  $N^W > N^{MSG}$ . This strict inequality is even more evident if Wymysiöeryś is compared to Middle High German:  $N^W > N^{MHG}$ .

### 3.7 Consonants

Both quantitatively and qualitatively, the consonantal system of Wymysiöeryś is remarkable. The size of the fundamental system of consonants oscillates between thirty and thirty-nine sounds. Due to palatalizing processes (see Section 3.9), this system may be expanded to more than fifty. As far as “non-palatalized” models are considered, the maximal system is posited by Andrason & Król (2016: 17–18). To be exact, Andrason & Król (2016) distinguish thirty-nine sounds: (a) plosives: [p], [b], [t], [d], [c], [ɟ], [k], [g], and [?]; (b) fricatives: [f], [v], [s], [z], [ç], [z], [s], [z], [ʃ], [ʒ], [x], and [h]; (c) nasals: [m], [n], [n̄], and [ŋ]; (d) liquids [l] and [r]; (e) affricates: [ts], [dz], [tç], [dʒ], [ts], [dz], [tʃ], [dʒ]; and approximants: [w] and [j].<sup>27</sup> A similarly abundant system of thirty-four consonants was proposed by Kleczkowski (1920: 13–14) and Mojmir (1930–1936: xiv–xv). The main difference consists in the presence of [l] instead of [w]<sup>28</sup> and the absence of the distinction between the postalveolars [s], [z], [ts], [dz] and the palatalo-alveolars [ʃ], [ʒ], [tʃ], [dʒ]. More reduced consonantal systems are proposed by Wicheriewicz (2003) and Lasatowicz (1994). Wicheriewicz (2003: 406–409) distinguishes thirty consonants. He expands the set of consonants by [pf] and [ç], on the one hand, but prescinds from [l], [ʃ], [?] and the distinction between the postalveolars [s], [z], [ts], [dz] and the palatalo-alveolars [ʃ], [ʒ], [tʃ], [dʒ], on the one hand.<sup>29</sup> Lasatowicz (1994: 36, 42, 52) also distinguishes thirty consonants. Similar to Wicheriewicz (2003), Lasatowicz’s system contains the sound [ç] but

<sup>26</sup>That is, fewer sounds can be nasal and the feature of nasality is generally less common.

<sup>27</sup>The set of approximants may be extended to [w] which occurs before nasals (Andrason & Król 2016: 17–18, Andrason 2021).

<sup>28</sup>At the beginning of the 20<sup>th</sup> century, *t* was still pronounced as a velarized alveolar lateral approximant [ɫ] rather than approximant [w] as it is the rule currently (Kleczkowski 1920: 13, 121–126, Mojmir 1930–1936: xiv, Wicheriewicz 2003: 406, Źak 2019, see Section 4.4).

<sup>29</sup>Note that [dz], [dʒ], [dʒ], [z] are unattested in Biesik’s poems (Wicheriewicz 2003: 408–409).

fails to include consonants [c], [f], [?], [n] and to make the distinction between the postalveolars [s], [z], [ts], [dz] and the palatalo-alveolars [ʃ], [ʒ], [tʃ], [dʒ].<sup>30</sup> Since nearly all consonants (except [?] and [h]) have their palatal(ized) variants (see Section 3.9; see also Kleczkowski 1920: 15, Mojmir 1930–1936, Andrason & Król 2016), the system of consonants may be extended to fifty-one sounds – or even more – by incorporating, for instance, [p<sup>j</sup>], [b<sup>j</sup>], [t<sup>j</sup>], [d<sup>j</sup>], [m<sup>j</sup>], [n<sup>j</sup>], [f<sup>j</sup>], [v<sup>j</sup>], [l<sup>j</sup>]/[ɿ], [x<sup>j</sup>]/[ç], [r<sup>j</sup>] and [w<sup>j</sup>] (Andrason 2021).

The system of consonants exhibited by Middle High German most likely contained twenty-four sounds. Specifically, (a) plosives: [p], [b], [t], [d], [k], [g]; (b) fricatives: [f], [v], [s], [ʃ], [z] (or [z]), [j], [ç], [x], and [h]; (c) nasals: [m], [n], [ŋ]; (d) liquids: [l], [r] (Paul 2007: 141, 169–171); (e) affricates [pf], [ts], [kx] (de Boor & Wisniewski 1973: 18–19, Simmler 1985: 1135, Paul 2007: 141); and possibly an approximant [w] (Wright 1917: 24, Simmler 1985: 1135). Additionally, some studies expand this system by [β] (de Boor & Wisniewski 1973: 18). The consonantal inventory of Modern Standard German includes the following twenty-seven sounds: (a) plosives: [p], [b], [t], [d], [k], [g], [?]; (b) fricatives: [f], [v], [s], [z], [ʃ], [ç] (the last sound is often treated as peripheral), [j] (also defined as an approximant), [x], [ç], [h]; (c) nasals: [m], [n], [ŋ]; and (d) liquids: [l], [r], [ɹ]; and (e) affricates: [pf], [ts], [tʃ] (Eisenberg 1994: 353–355, Johnson & Braber 2008: 92, 99–101, 104, Fagan 2009: 10–14, O'Brien & Fagan 2016: 13–16). If the affricate [dʒ] (O'Brien & Fagan 2016: 16) and the nasal [m̩] (Hall 2000) are included, the number of consonants increases to twenty-nine “basic” sounds (see also the more phonologically oriented analyses offered by Russ 1994: 112–115, 121–122; Wiese 1996: 22–26, Hall 2000: 31, and Fox 2005: 35–37).

Overall, the category of consonants ( $C$ ) is instantiated in both Wymysiöeryś and the two control languages. However, whether maximal (extended) or minimal (basic), the number of Wymysiöeryś consonants is always larger, in certain models radically, than the number of consonants found in Middle High German and Modern Standard German. As a result, the complexity of consonants in Wymysiöeryś is substantially greater than in the control languages, i.e.  $C^W > C^{MHG}$  and  $C^W > C^{MSG}$ .

### 3.8 Consonantal length

Consonantal length belongs to the phonetic repertoire of Wymysiöeryś, although different studies ascribe it a distinct systemic relevance. According to Kleczkowski (1920: 15) and Mojmir (1930–1936: xv), although attested, long consonants are not particularly common. In contrast, Wicherkiewicz (2003: 405–407) identifies

<sup>30</sup>Latosiński (1909: 270) distinguishes twenty-six consonants.

a number of long consonants in Biesik's poems (namely, [m:], [f:], [p:], [k:], [t:], [ts:], and [l:]) and notes that they are still pronounced at least "slightly longer" than their short counterparts by modern speakers Wicherkiewicz (2003: 407). I have detected a relatively large set of long consonants in my own field work, namely: nasals [n:], [n̩:], [m]; fricatives: [s:], [z:]; stops: [p:], [t:], [k:]; and affricates [ts:] and [dʒ/dʒ:]; as well as [r:] (Andrason 2021). Although long consonants also allow for a shortened pronunciation as singletons, consonantal length seems to be a regular feature of the Wymysiöeryś sound system.

Long or geminated consonants were a typical component of Middle High German (Wright 1917: 25, 27–28, 30–31, Simmler 1985: 1134–1135, Goblirsch 1997, 2018, Jessen 1998: 334, Paul 2007: 141), where the length played, to an extent, a phonemic function (Fourquet 1963: 85–88, Moosmüller & Brandstätter 2015). The following long consonantal sounds are usually identified for Middle High German: *pp* [p:], *bb* [b:], *tt* [t:]; *gg* [g:], *ff* [f:], *ss* [s:], *mm* [m:], *nn* [n:], *ll* [l:], and *rr* [r:] (Wright 1917: 25, Paul 2007: 141). The set of long consonants is often extended by [ʃ:], [x:], and [k:] (Simmler 1985: 1135, Paul 2007: 142, 171, for a discussion consult Goblirsch 1997, 2018 and Paul 2007: 141–175). Modern Standard German has no geminate or long consonants "at the phonetic level" (Caratini 2009: 70, Goblirsch 2018). The only consonantal sounds present in the language are thus singletons (Caratini 2009: 70, Fagan 2009) and spelling them with double consonants generally indicates that a preceding vowel is short (Russ 1994: 118, 140).

Overall, the category of consonantal length (*CL*) is instantiated in Wymysiöeryś and one of the control languages, i.e. Middle High German. It is absent in Modern Standard German. The number of long consonants in Wymysiöeryś and Middle High German is relatively similar and, thus, their respective complexities may be viewed as approximately equal, i.e.  $CL^W \approx CL^{MHG}$ . The comparison between Wymysiöeryś and Modern Standard German reveals the relation of strict inequality, i.e.  $CL^W > CL^{MSG}$ .

### 3.9 Palatalization

A palatalization-based opposition between the so-called hard (non-palatal(ized)) and soft (palatal(ized)) consonants is a pervasive and essential component of the Wymysiöeryś sound system (Kleczkowski 1920, Anders 1933, Wicherkiewicz 2003: 402, 405–409, Andrason 2021). Virtually every non-palatal(ized) hard sound possesses its palatal(ized) soft counterpart (Kleczkowski 1920: 15, Mojmir 1930–1936: xv, Andrason & Król 2016). Apart from the alveolo-palatal (or palatalized postalveolar) sounds (i.e. the fricatives [ç] and [z], the affricates [tç] and [dʒ], and the nasal [n]), Wymysiöeryś exhibits the following palatal(ized) consonants,

each contrastive with a hard equivalent: [p<sup>j</sup>] – [p], [b<sup>j</sup>] – [b], [v<sup>j</sup>] – [t], [d<sup>j</sup>] – [d], [k<sup>j</sup>]/[c] – [k], [g<sup>j</sup>]/[f] – [g], [m<sup>j</sup>] – [m], [ŋ<sup>j</sup>] – [ŋ], [f<sup>j</sup>] – [f], [v<sup>j</sup>] – [v], [l<sup>j</sup>]/[k] – [l], [x<sup>j</sup>]/[ç] – [x], [r<sup>j</sup>] – [r], and [w<sup>j</sup>] – [w]. Palatalization also plays a relevant role in the morphology of Wymysiöeryś, for instance, triggering the use of an allomorphic ending *-ja* instead of *-a* in inflectional forms of nouns and adjectives, e.g. *ryk* [rək] ‘back’ – dat.pl. *rykja* [rəc(:)a] (Andrason 2014a, 2015, Andrason & Król 2016). This large number of palatal(ized) consonants and their extensive use in the lexicon and grammar give the Wymysiöeryś language a soft resonance, fully comparable to Polish, but noticeably distinct from German (Kleczkowski 1920; see also Latosiński 1909: 271–272).<sup>31</sup>

Palatalization operated in Middle High German residually. The most evident palatalizing process affected the consonant *s* that was softened to [ʃ] before the consonants *k*, *l*, *m*, *n*, *p*, *t*, *w* (Paul 2007, Fagan 2009: 196, 209, Hennings 2012: 41). Additionally, *g* was palatalized to *j* (Paul 2007: 37). Modern Standard German fails to exploit palatalization and palatal(ized) consonants to an extent that would be comparable to that attested in Wymysiöeryś. The most evident case of palatalization is the softening of [x] to [ç] (Fagan 2009: 26–27, O’Brien & Fagan 2016: 45). The language also contains other palatal sounds, namely [j], [ʃ] and [ʒ] (Johnson & Braber 2008: 92, 95, 104, Fagan 2009: 9, 12–13, van der Hoek 2010, O’Brien & Fagan 2016: 14, 16).

Overall, the category of palatalization (*P*) is instantiated in Wymysiöeryś and the two control languages. However, the number of palatalized consonants and the systemic effects of palatalization are far more significant in Wymysiöeryś than in Middle High German and Modern Standard German. Hence, palatalization is substantially more complex in Wymysiöeryś than in the control languages:  $P^W > P^{MHG}$  and  $P^W > P^{MSG}$ .

### 3.10 Aspiration

In Wymysiöeryś, voiceless plosive consonants /p/, /t/, /k/ fail to be aspirated in a word-initial position, and in all other positions (however, see further below), thus appearing as [p], [t], [k] (Kleczkowski 1920: 14–15, 28, Mojmir 1930–1936: xiv–xv, Andrason & Król 2016: 17–19). This means that they contrast with /b/, /d/, /g/ in terms of voicing as in Polish (*jak w polskiem*) (Kleczkowski 1920: 28), the latter series being pronounced voiced in word-initial and word-internal position (Andrason & Król 2016: 17–19).<sup>32</sup> An analogous situation is found in Biesik’s

<sup>31</sup>The perception of Wymysiöeryś as a soft language – and in that regard equal to Polish – is a usual reaction when a non-Wymysiöeryś speaker who is familiar with German (and Polish) is exposed to the Wymysiöeryś language.

<sup>32</sup>In a final position, /b/, /d/, and /g/ are generally devoiced, and thus their opposition with /p/, /t/, and /k/ is neutralized.

poems where /p/, /t/, /k/ were unaspirated, and the opposition with /b/, /d/, /g/ involved the feature of voicing only (Wicherkiewicz 2003: 399–409). Aspiration is also absent in the Wymysiöery's variety described by Lasatowicz (1994: 42).<sup>33</sup> Only sporadically, a soft aspiration of /p/, /t/, and /k/ is audible in a word-final position (Andrason & Król 2016: 19).

In studies on Middle High German, the opposition between /p/, /t/, /k/ and /b/, /d/, /g/ is generally viewed in terms of tenseness (Goblirsch 1997, 2018, Jessen 1998) – alternatively referred to as “spread glottis” (Harbert 2007: 44, Iverson & Salmons 2003: 44, 2008) – that is, *fortis* versus *lenis* (Simmler 1985: 1134, Weddige 2007: 18, Hennings 2012: 8–10, Paul 2007: 131, 141, Moosmüller & Brandstätter 2015). However, the determination of the precise phonetic nature of this opposition is elusive. Most likely, the contrast translated onto a set of phenomena, such as force, quantity, voicing, and aspiration, all of them characterized by distinct degrees of relevance (Simmler 1985: 1133–1135). The most relevant of them were articulatory force (Wright 1917: 22–23, Weddige 2007: 18) and quantitative augmentation (Goblirsch 1997, 2018, Jessen 1998: 334; see also Simmler 1985: 1135) – both related to intensity. Often, voicing is considered the third crucial property correlated with tenseness (de Boor & Wisniewski 1973: 18, Simmler 1985: 1133, Weddige 2007: 19, Seiler 2009, Hennings 2012: 8–10). In contrast, even though aspiration is apparently “inherent” to the Germanic family (Iverson & Salmons 2003: 44, Iverson & Salmons 2008: 2), its role in Middle High German was secondary and the three *fortes* consonants were most likely not (fully) aspirated (Simmler 1985: 1133, Hennings 2012: 9; see also Paul 2007: 133). In Modern Standard German, the system of plosives is also based on the tenseness or spread-glottis contrast, such that the opposition between /p/, /t/, /k/ and /b/, /d/, /g/ is generally explained as *fortis* versus *lenis* (Russ 1994: 115, Wiese 1996, Jessen 1998: 22, 136, 142–143, Fox 2005: 42, Iverson & Salmons 2008: 3, Caratini 2009). The feature of tenseness is correlated primarily with aspiration, with /p/, /t/, /k/ being “heavily aspirated in prosodically prominent positions” (Iverson & Salmons 2008: 3), e.g. word-initially (Russ 1994: 115, 117, 121, Iverson & Salmons 2003, 2008, Fox 2005: 42, Caratini 2009).<sup>34</sup> In this system, voicing is viewed as a secondary feature (Jessen 1998: 334).

<sup>33</sup>Lasatowicz (1994: 42) argues for a *fortis-lenis* distinction between /p/, /t/, /k/ and /b/, /d/, /g/, without explaining how this contrast should be understood in phonetic terms.

<sup>34</sup>Although most scholars reject voice as a distinctive feature in Modern Standard German, its relevance is also acknowledged (Wiese 1996: 169), since the series /b/, /d/, /g/ surfaces not only as unaspirated but also as partially voiced (Iverson & Salmons 2008: 3). Overall, the plosives contrast in both aspiration (primarily) and voicing (secondarily). The tense consonants /p/, /t/, /k/ are aspirated and/or unvoiced. The lax consonants /b/, /d/, /g/ are unaspirated and/or voiced (Jessen 1998: 43–44; see also Caratini 2009: 70).

Overall, the category of aspiration (*A*) is not instantiated (or very poorly instantiated) in Wymysiöeryś and Middle High German while it is present and central in Modern Standard German. Therefore, the complexity relationship between Wymysiöeryś and the control languages is as follows:  $A^W \approx A^{MHG}$  and  $A^W < A^{MSG}$ .

### 3.11 Onset clusters

Wymysiöeryś tolerates elaborated consonant onsets (Andrason 2015: 71). Mono-segmental onsets may exhibit all consonants except [ŋ].<sup>35</sup> Bisegmental onsets exhibit a considerable variety, tolerating the following clusters: (a) stop + liquid/nasal/fricative/approximant; (b) fricative/liquid/nasal/fricative/stop/approximant/affricate; and (c) affricate + liquid/nasal/fricative (Andrason 2021). Contrary to many West Germanic languages (see Section 4.7), Wymysiöeryś tolerates bisegmental onsets such as [tl] and [dl], onsets with a glide as their second segment [j/w], and onsets with a voiced sibilant as their first element, e.g. [zm], [zm], [zv]. Additionally, it contains other onsets rare in German and its relatives: [kf], [tf], [tx], [ps], [pf], [bz], [gʒ], [tʃf], and [ʃf]. Three consonant onsets are also common in Wymysiöeryś, and a wide range of combinations are possible: stop + fricative + nasal (e.g. [bʒn]); fricative + fricative + stop (e.g. [fsp], [fst], [vzd], [vzg]); fricative + fricative + nasal (e.g. [vzm]); fricative + stop + fricative (e.g. [fkf], [stf]); fricative + stop + liquid (e.g. [skr], [spr], [str], [skn]; and [ʃkl], [ʃkr], [ʃpr], [ʃtr]). Although not particularly frequent, a few onsets composed of four consonants are attested, e.g. [vskf] and [pstr]. In addition to the quantitative complexity outlined above, Wymysiöeryś attests a significant qualitative variety of onset clusters. This is evident in the fact that the language tolerates not only sequences that conform to the sonority scale, but also those that violate it.<sup>36</sup> Apart from clusters in which the first element is a sibilant [s] or [ʃ] (as is common in

<sup>35</sup>This means that contrary to Modern Standard German (see below), [ʃ] and [x] may form monosegmental onsets.

<sup>36</sup>The sonority scale depicts the increase in the relative sonority of sounds and their “vowel-likeness” (Foley 1977, Clements 1990). Generally, the sonority increases from obstruents to vowels, via sonorants. A more fine-grained representation of the scale is as follows: voiceless stops > voiced stops > voiceless fricatives > voiced fricatives > nasals > 1 > r > glides / high vowels > low vowels (Harbert 2007: 65). This scale imposes sonority restrictions whereby, in onsets, consonants placed higher on the sonority scale may not occur before consonants placed lower on the sonority scale (Harbert 2007: 66, 68). In other words, the sonority of a syllable may not decrease from the left edge to its nucleus, but rather increases (Harbert 2007: 66, 68). Inversely, elements in codas must “decline in sonority toward the right edge of the syllable” (Harbert 2007: 73).

Germanic languages; see below), a large number of such “ill-formed” sequences is tolerated, especially: (a) fricative + stop + fricative (e.g. [fkʃ]) and (b) fricative + fricative + stop (e.g. [fsp], [fst], [vzd], [vgz]).

Middle High German maximally contains three consonants in onsets, which is typical of the Germanic family in general (Harbert 2007: 66, van Oostendorp 2019: 34). As elsewhere in the Germanic family (van Oostendorp 2019: 35), monosegmental onsets exhibit very few restrictions, e.g. [x], [ç], and [ŋ]. In biconsonantal onsets, sequences composed of an obstruent and a liquid are allowed with the exception of [tl] and [dl] (van Oostendorp 2019: 36, compare with Wymysiöeryś above). Other types of onsets are more restricted, with a number of combinations being disallowed, e.g. onsets with a glide as second segment [j/u/w] and onsets composed of sibilants and voiced obstruents (cf. van Oostendorp 2019: 36–38). In general, conforming to the behavior exhibited by West Germanic languages, Middle High German complies with the sonority-based constraints (Harbert 2007: 68, 73), van Oostendorp 2019: 36) to a much larger extent than is the case in Wymysiöeryś. In Modern Standard German, monosegmental onsets tolerate most consonants with the exception of [x], [ŋ], and [ʃ] (Fox 2005: 58).<sup>37</sup> For complex onsets, only doubles are relatively common. Two basic types can be discerned: obstruent (plosive, fricative, affricate) + liquid ([r/l]) and fricative (mostly, [ʃ]) + C (Fox 2005: 58). Specifically, the following combinations are grammatical: stop + liquid/nasal/fricative; fricative + liquid/nasal; and only for ſ, fricative + fricative/stop (Veith et al. 1980: 133, Eisenberg 1994: 356, Russ 1994: 120). The only common second segments are sonorants (Fagan 2009: 35; see also O’Brien & Fagan 2016: 66, Hall 2000: 231). Additionally, affricates may combine with a liquid or a fricative (Fagan 2009: 35, 58). Triple onsets are scarce and highly restrictive, both qualitatively and quantitatively (Kučera & Monroe 1968: 50). Only five permutations are grammatical (i.e. [skl], [skr], [spl], [spr], [str]), all of them of the type [s,ʃ] + stop + liquid (Fox 2005: 58, Fagan 2009: 36, Hall 1992: 69, 2000). Quadruple onsets are disallowed (Kučera & Monroe 1968: 50, Eisenberg 1994: 355, Fox 2005: 55; see also Wiese 1996). Onsets largely comply with the sonority scale principle (Wiese 1996: 260, Fox 2005: 60, van Oostendorp 2019: 36–37). The only common exceptions involve [ʃ] and [s] which may occur before stops (Fox 2005: 60, van Oostendorp 2019: 39–40).

Overall, consonant clusters (*OC*) in onsets are instantiated in Wymysiöeryś and the two control languages. However, onset clusters are significantly longer

---

<sup>37</sup>When describing the consonant clusters in Modern Standard German in Section 3.11 and Section 3.12, I will often refer to Russ (1994), Wiese (1996), Hall (2000), Fox (2005), and O’Brien & Fagan (2016). It should be noted that, in their discussion of phonotactics, these authors are principally concerned with phonemes.

(i.e. larger sequences are tolerated) and more varied (i.e. a more diverse set of combinations is grammatical) in Wymysiöeryś than in Middle High German and Modern Standard German. This yields the following complexity relationships between Wymysiöeryś and the control languages:  $OC^W > OC^{MHG}$  and  $OC^W > OC^{MSG}$ .

### 3.12 Coda clusters

Wymysiöeryś allows for elaborated and qualitatively diversified consonant codas. All consonants are allowed in monosegmental codas with the exception of [?] and [h], as well as voiced plosives, due to devoicing processes operating in a word-final position. Diverse combinations are also tolerated in codas composed of two consonants, including a large set of clusters containing the affricates [ts] and [tʃ] and their variants (also resulting from the devoicing of /dʒ/ and /dʒ/).<sup>38</sup> Nevertheless, certain constraints operate, of which the most pervasive is the ungrammaticality of the final [j], [h], and [?], as well as the avoidance of final voiced obstruents. Three-segment codas are also widely attested; however, they are only common in morphologically complex forms. Their presence in mono-morphemic (i.e. non-inflected) lexemes is in contrast limited. The most common mono-morphemic three-segment codas are: [nft], [mpt], as well as [mpl] and [ndl] in cases where the final sonorants are not syllabic. The most common combinations found in pluri-morphemic words involve [Cst] and [CCt], which tend to arise in inflected forms of verbs (e.g. in the present tense and the preterite) and adjectives (e.g. superlative), e.g. [tst], [kst], [fst], [hst], [mst], [nst], [ŋst], [wst], [ŋkt], [wtst]. Four-consonant coda are also grammatical being attested exclusively in inflected forms of verbs and adjectives. The first element is always a sonorant, the second is a stop, while the third and fourth segments are usually filled by the cluster [st], e.g. [wdst], [lkst], [ntst], and [nkst]. Five-segment codas are generally avoided in Wymysiöeryś. Codas typically comply with the sonority scale principle. Common exceptions are clusters ending in a plosive and a sibilant, e.g. [ps] or [ks].

As is typical of West Germanic languages (van Oostendorp 2019: 40), in Middle High German, [h] and [?] were disallowed in mono-segmental codas in a word-final position (Paul 2007: 161), as is, most likely, also true of the voiced plosives due to their final devoicing (Paul 2007: 19). Bi-segmental codas were common, and a diversified set of combinations was tolerated, e.g. liquid + obstruent (including [lt]) or nasal + obstruent (cf. van Oostendorp 2019: 41–42). Codas longer

<sup>38</sup>For example, one finds [tʃt] and [tst] as the first segment, and [ʃtʃ], [ltʃ], [ntʃ], [rtʃ], [wts], [lts], [nts], [rts], [ndz] as the second segment (Andrason 2021).

than two consonants – i.e. triplets and quadruplets – usually emerge in forms that exhibit complex morphology, e.g. verbal, nominal, and adjectival inflections. The typical word-final element in complex codas are voiceless coronal obstruents as is the rule in West Germanic languages in general (van Oostendorp 2019: 43–44; for various examples in Middle High German see Paul 2007: 146–175). In Modern Standard German, all consonants, except [b], [d], [g], [v], [z], [j], [h], and [?], are grammatical in monosegmental codas (Fox 2005: 59, van Oostendorp 2019: 41). Two-consonant codas also show minimal restrictions and a large variety of combinations are grammatical (Eisenberg 1994: 356, Russ 1994: 124–125). Specifically, bisegmental codas cannot end in voiced obstruents, [j], [h], [?], and voiced plosives. The allowed clusters are mostly of the four types: [l/r] + obstruent or nasal; nasal + obstruent; obstruent + [t]; and plosive + fricative (Fox 2005: 59). Codas composed of three segments are common, especially due to the presence of inflectional endings (Russ 1994: 125). However, certain important restrictions on their combinatory freedom operate as well (Fox 2005: 59, Fagan 2009: 37). The most pervasive of them is the presence of *t* or *s* in the final position (Fox 2005: 59). Four-segment codas invariably have a sonorant as their first element, and [st<sup>h</sup>] or [t<sup>h</sup>s] as their third and fourth elements (Fagan 2009: 37–38). Their presence within a single morpheme is extremely rare – only two lexemes are attested (Fox 2005: 59, Fagan 2009: 3; see also Hall 1992: 121). Inversely, they tend to appear in multi-morphemic forms, e.g. in inflected nouns, adjectives, and verbs (Hall 1992: 121, Russ 1994: 125, Fagan 2009: 37–38, van Oostendorp 2019: 43–44). Five-segment codas are highly problematic. There are only three inflectional forms that exhibit such combinations of sounds (Hall 1992: 121, Russ 1994: 125). For some scholars, these clusters are “not well formed” (Wiese 1996: 48) being indeed “unpronounceable for many Germans” (Fagan 2009: 51 referring to Hall 1992: 121). As was the case of onsets (see Section 3.11), codas generally conform to the sonority scale principle, a common exception being [ps] and [t<sup>h</sup>s] (Wiese 1996: 260, Fox 2005: 60, Fagan 2009: 37).

Overall, complex codas (*CC*) are instantiated both in Wymysiöeryś and the two control systems. In all three languages, coda clusters exhibit similar length and variety. These degrees of complexity of Wymysiöeryś, Middle High German, and Modern Standard German are thus approximately equal:  $CC^W \approx CC^{MHG}$  and  $CC^W \approx CC^{MSG}$ .

### 3.13 Module-global complexity

The relational complexities of Wymysiöeryś, Middle High German, and Modern Standard German determined locally for the twelve phonetic/phonological fea-

tures are recapitulated in Table 1 below. For each feature studied, relational complexity was determined in terms of equality  $\approx$  (similar), inequality  $\leq / \geq$  (minimally lower/greater), and strict inequality  $< / >$  (substantially lower/greater). The evidence shows that the complexity of Wymysiöeryś is substantially greater than the complexity exhibited by Middle High German and Modern Standard German with regard to the features of triphthongs, nasality, consonants, palatalization, and onset clusters. With regard to the feature of vocalic sonorants and consonantal length, the complexity of Wymysiöeryś is substantially lower than the complexities of Middle High German and Modern Standard German, respectively. Lastly, with regard to aspiration, the complexity of Wymysiöeryś is substantially lower than the complexity exhibited by Modern Standard German. In the remaining cases, the complexity of Wymysiöeryś is equal to the complexities of Middle High German and Modern Standard German.

Table 1: Local complexity of Wymysiöeryś in relation to Middle High German and Modern Standard German

Features	Wymysiöeryś versus Middle High German	Wymysiöeryś versus Modern Standard German
Monophthongs	W $\approx$ MHG	W $\approx$ MSG
Diphthongs	W $\approx$ MHG	W $\approx$ MSG
Triphthongs	W $>$ MHG	W $>$ MSG
Vocalic sonorants	W $>$ MHG	W $\approx$ MSG
Vocalic length	W $\approx$ MHG	W $\approx$ MSG
Nasality	W $>$ MHG	W $>$ MSG
Consonants	W $>$ MHG	W $>$ MSG
Consonantal length	W $\approx$ MHG	W $>$ MSG
Palatalization	W $>$ MHG	W $>$ MSG
Aspiration	(?) W $\approx$ MHG	W $<$ MSG
Onset clusters	W $>$ MHG	W $>$ MSG
Coda clusters	W $\approx$ MHG	W $\approx$ MSG

When analyzed globally from the perspective of the entire sound-system module, Wymysiöeryś exhibits greater complexity than the two control systems. In comparison to Middle High German, Wymysiöeryś exhibits greater complexity in half of the features – in the other half, the complexity of the two languages is equal. In comparison to Modern Standard German, Wymysiöeryś also exhibits greater complexity in six features; in one feature, the complexity of Modern Standard German surpasses that of Wymysiöeryś; in the remaining five features, the

complexities of the two languages are equal. If one allocates 1 for being substantially greater (>), 0 for equality ( $\approx$ ), and -1 for being substantially lower (<), the relational global sound-system complexity values (*SS-COMPL*) are the following: Wymysiöeryś (+6) versus Middle High German (-6); and Wymysiöeryś (+5) and Modern Standard German (-4). Accordingly,  $SS\text{-}COMPL^W > SS\text{-}COMPL^{MHG}$  and  $SS\text{-}COMPL^W > SS\text{-}COMPL^{MSG}$ .

It is also possible to relate the complexities of the three languages simultaneously, and thus take into account complexity relationship linking Middle High German and Modern Standard German in addition to relationships involving Wymysiöeryś. The scale in Figure 1 below represents the relative position of the three languages. The scale is limited by two extremes: maximal score +24 (the complexity of language *x* is substantially greater than the complexity of languages *y* and *z* in all features) and minimal score -24 (the complexity of language *x* is substantially lower than the complexity of languages *y* and *z* in all features). Out of possible 24 points (12 for each comparative analysis with one of the other systems), Wymysiöeryś scores +11 (12 >; 11  $\approx$ ; 1 <). Middle High German scores -7 (1 >; 15  $\approx$ ; 8 <). Modern Standard German scores -4 (3 >; 14  $\approx$ ; 7 <). This again demonstrates the greater module-global complexity of Wymysiöeryś if compared to the two control languages.

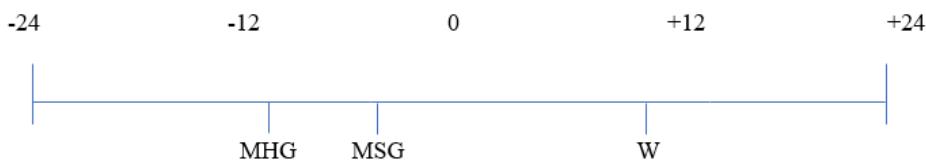


Figure 1: Relative complexity of Wymysiöeryś, Middle High German and Modern Standard German

For some features (i.e. triphthongs, vocalic sonorants, nasal vowels, and consonantal length) the contrast between Wymysiöeryś, Middle High German, and Modern Standard German concerns the presence of a category (distinctiveness), not only the number of the category's expression manners (economy). That is, in one language (or two languages) a certain category is instantiated, while in the remaining one(s), it is not instantiated at all. In total, Wymysiöeryś instantiates eleven categories (only the category of aspiration is not expressed); Middle High German instantiates eight categories (the absent categories are triphthongs, vocalic sonorants, nasality, and most likely aspiration); Modern Standard German instantiates ten categories (the absent categories are triphthongs and consonantal length). If, for the instantiation of each category, a language is allocated 1

point, Wymysiöeryś scores 11, Middle High German 8, and Modern Standard German 10 – Wymysiöeryś being thus more complex than the control languages.

As explained in Section 2, the estimation of module-global complexity is always problematic due to the issue of commensuration and the availability of various manners of combining local complexities. Therefore, the converging results of the three complexity measurements of the sound-system module used in this section – which all identify Wymysiöeryś as the most complex among the analyzed languages – demonstrate that Wymysiöeryś' greater complexity score is not accidental or theory-driven. It is thus very likely that the sound system of Wymysiöeryś is *objectively* more complex than the two control languages.<sup>39</sup>

## 4 The origin of the surplus

Having established that as far as the sound system is concerned, the global complexity of Wymysiöeryś is greater than the complexity of Middle High German and Modern Standard German, I will determine whether this complexity surplus exhibited by Wymysiöeryś is attributable to contact with Polish. First, the source of the complexity surplus found in seven features will be analyzed: triphthongs, vocalic sonorants, nasality, consonants, consonantal length, palatalization, and onset clusters (Section 4.1–Section 4.7). This will subsequently allow me to determine the surplus' origin from a global perspective (Section 4.8).

### 4.1 Triphthongs

Currently, the vocalic system of Polish contains only short monophthongs.<sup>40</sup> Inversely, syllables may not exhibit complex nuclei, and thus long vowels, diphthongs, and triphthongs (Gussmann 2007: 181; see also Strutyński 1998: 59–60, 72, 74, Jassem 2003: 105–106, Sussex & Cubberley 2006: 154, 156, see however Wągiel 2016). At previous diachronic stages, specifically between the 10<sup>th</sup> and the 15<sup>th</sup> century, Polish did exhibit long vowels. Vocalic length was however lost in the 16<sup>th</sup> century (Długosz-Kurczabowa & Dubisz 2006: 132, 136–137). Although absent at a phonemic level, diphthongs emerge in Polish due to nasalization processes (Wągiel 2016: 53, 62–64, 83)<sup>41</sup> and also exist in dialects (Dejna 1973, Bać

<sup>39</sup>Since the absence of a category has more far-reaching systemic effects than the distinct numbers of encoding manners, one could arguably give even more weight to the complexity of those languages where a category is present. Since the complexity hierarchy of the three languages would remain the same, this would have no important bearings on the result of my study.

<sup>40</sup>These vowels are /i/, /i/ø/, /u/, /e/, /a/, /o/. Often, [i/ø] and [i] are considered allophones of a single phoneme (Sussex & Cubberley 2006: 154, 156).

<sup>41</sup>Alternatively, the emergent nasal components are analyzed as approximants (Gussmann 2007).

1997). Sometimes, sequences composed of a vowel and the approximants [j] or [w] are analyzed as diphthongs (Jassem 1973, Demenko 1999, Wągiel 2016: 81–83). In any case, neither presently nor at its previous developmental stages, Polish possessed true triphthongs. As a result, the presence of triphthongs in Wymysioeryś cannot result from contact with Polish.<sup>42</sup>

#### 4.2 Vocalic sonorant

Although Polish admits sequences with “trapped” sonorants, in which a sonorant is enclosed between two elements of lower sonority, typically two obstruents, e.g. [drg] *drgać* ‘vibrate’ (Kijak 2008: 62, 66), it fails to possess true syllabic sonorants. This contrasts with the situation attested in other (neighboring) Slavonic languages where sonorants used in the above-mentioned sequences tend to exhibit a syllabic status (Sussex & Cubberley 2006, Kijak 2008: 66, see the absence of syllabic sonorants in discussions of Polish phonetics and phonology, e.g. Strutyński 1998, Jassem 2003, Gussmann 2007, and Wągiel 2016). As a result, the presence of vocalic sonorants in Wymysioeryś cannot be attributed to Polish influence.<sup>43</sup>

#### 4.3 Nasality

Nasality is a prominent feature of the Polish sound system. Polish has two nasal phonemes *q* /ɔ/ and *ɛ* /ɛ/ (Urbańczyk 1991: 297–298, Rothstein et al. 1993: 659, Bloch-Rozmej 1997, Gussmann 2007, Wągiel 2016: 88, 100). In addition to those two phonemes, which are usually realized as [ɔ] and [ɛ], Polish contains a large number of nasal vowels at a phonetic level, e.g. [i̯], [ã], [ũ], and [i/ɛ] (Bloch-Rozmej 1997: 95, Strutyński 1998: 58–59, 61, 72). Overall, for every oral vowel, there is a nasal equivalent used in certain environments (Urbańczyk 1991: 298).<sup>44</sup> As a result, nasality is viewed as a key phonological and phonetic category in Polish (Bąk 1997, Strutyński 1998: 77, Gussmann 2007: 269–287, Wągiel 2016).<sup>45</sup>

<sup>42</sup>One should note that, although absent in Modern Standard German (see Section 3.3), triphthongs are attested in the West Germanic family, e.g. in Bavarian German and High Alemannic varieties.

<sup>43</sup>On the other hand, syllabic sonorants are also relatively common in various varieties of (Modern Standard) German (see Section 3.4) and in other Germanic languages.

<sup>44</sup>Such environments are: /n/ + /fricative/ and /m/ + /f, v/ (Urbańczyk 1991: 298).

<sup>45</sup>In a careful Standard Polish speech, the realization of nasality is asynchronous (Urbańczyk 1991: 297–298, Bąk 1997). This gives rise to the emergence of nasal approximants such as [w̄], [ū], and [j̄] (Rothstein et al. 1993: 660, Gussmann 2007: 270–271). In colloquial speech, nasal vowels often resolve into oral vowels and nasal consonants (Rothstein et al. 1993: 659, Bąk 1997, Rubach 1977, Rowicka & van de Weijer 1992, Bloch-Rozmej 1997: 84–86, Gussmann 2007: 271).

Contrary to Polish, nasal vowels do not constitute a prominent feature in the phonetics and phonology of continental Germanic languages (see their absence in general works on the Germanic family, e.g. Harbert 2007 and Jacobs et al. 1994).<sup>46</sup> In West Germanic and German varieties, nasal vowels are generally restricted to loanwords, often allowing for an alternative oral pronunciation (Russ 1994: 78, 108, Fagan 2009: 9, Caratini 2009: 51, 73–74). German dialects in which nasality is more prominent are: Swabian (an Upper German, Alemannic dialect), Pfaelzisch (*Pfälzisch*) or Palatine German (van Ness 1994: 423, Stevenson 1997: 71, Niebaum & Macha 1999: 197), and the dialect of Luzern (Bacher 1905: 179). Secondary nasal vowels are also found in Yiddish (Weinreich 2008: 583–585, Addendum 606, Herzog et al. 1992: 19–20, 41, Jacobs et al. 2005: 97–99) and Frisian (Hoekstra & Tierstna 1994: 508).<sup>47</sup>

Although nasality is present in the Germanic family, being a common phonetic process from a cross-linguistic perspective, its emergence in Wymysiöeryś most likely stems from Polish influence. Indeed, nasal vowels appear most commonly and most consistently in Polish loans. This complies with the origin of nasality in Yiddish where its presence is attributed to Slavonic influence (Weinreich 2008: 583–585).<sup>48</sup>

#### 4.4 Consonants

Polish has a large and diversified set of consonants. The basic consonantal inventory consists of 41 sounds: 38 consonants and 3 approximants [j, w, w<sup>j</sup>] (Bąk 1997, Strutyński 1998: 74, Jassem 2003, Gussmann 2007: 3–8). This set is often expanded to nearly fifty due to the inclusion of voiceless sonorants [m̥, n̥, l̥, r̥] and a voiced velar [ɣ] (Gussmann 2007: 4). With the incorporation of palatalized consonants and the approximant [ɥ], the maximal system ascends to nearly seventy consonants (Strutyński 1998: 54, 72–73). Crucially, the consonants that are absent in Middle High German and Modern German but that currently feature in Wymysiöeryś are all found in Polish too. This includes: (a) laminal alveolo-palatal ([ç], [z], [tç], and [dʒ]) and postalveolar sibilants and affricates ([s], [z], [ts], [dʒ]) (Hamann 2003, 2004; cf. Karaś & Madejowa 1977 and Gussmann 2007:

<sup>46</sup>The exception is a chapter dedicated to Old Icelandic (Práinsson 1994: 147).

<sup>47</sup>Nasality is more consistently present in peripheral languages: Surinam Dutch (DeSchutter 1994: 444), Afrikaans (Donaldson 1994: 481), and – albeit rather as an archaism used by older speakers – Pennsylvania German (van Ness 1994: 423).

<sup>48</sup>Note also that continental German varieties where nasality is more visible (e.g. Pfaelzisch/Palatine and Luzern) are usually spoken in areas adjacent to languages containing prominent nasal vowels, in particular French.

75–78);<sup>49</sup> (b) alveolo-palatal nasal [n] (Jassem 2003: 104, alternatively transcribed as a palatal [ɲ]); (c) a series of other palatal(ized) consonants (see Section 4.7, Rothstein et al. 1993: 687–690, Strutyński 1998: 38, 42–44, 54, Sussex & Cubberley 2006: 165–166, Gussmann 2007: 4–7); and (d) the labialized velar approximant [w] (Jassem 2003, Strutyński 1998, Gussmann 2007).

With regard to soft and hard sibilants and affricates, contact with Polish seems to be the direct and sole factor responsible for their introduction to Wymysiöeryś (Żak 2016, Andrason 2014b,a, 2015, 2021). This can be inferred from the absence of those two series in West Germanic languages, on the one hand, and their particular stability in Wymysiöeryś in lexical borrowings from Polish, on the other hand.<sup>50</sup> For the remaining types of consonantal surplus, Polish seems to have (significantly) strengthened and/or accelerated tendencies that are typologically common and that had operated (at least marginally) language- or family-internally.

Although the presence of the alveolo-palatal nasal [n] may partially be attributed to the Polish influence, being evident in a large number of Polish loanwords, in which the original sound [n] is rendered as such, it seems to coincide with language-internal processes. In genuine Wymysiöeryś vocabulary, [n] typically derived from ĥ [ŋ̊] (itself a reflex of an original cluster *ng/nc* [ŋ]) or arose in cases where *n* was followed by palatal sounds *i*, *j*, or *ć* (Andrason 2021). The palatalization of the velar nasal [ŋ] to [ŋ̊] constitutes a recurrent cross-linguistic tendency. It occurred in the Szynwald/Bojków (Schönwald) dialect, closely related to Wymysiöeryś, which suggests a dialectal – family-internal – development (cf. Gusinde 1911: 98–99). Even though articulatory proximity may motivate the development from [ŋ̊] to [n], this change occurred only after World War II, coinciding with the increased presence of the Polish language in Wilamowice (see that it was still written ĥ by Kleczkowski 1920 and Mojmir 1930–1936). The palatalization of *n* in palatal contexts, which had already taken place before

<sup>49</sup>The sounds of the “hard” series are defined – especially by Polish scholars – as postalveolars and represented by [ʃ], [ʒ], [tʃ], [dʒ] (cf. Biedrzycki 1974, Spencer 1986, Dogil 1990, Jassem 2003, and Gussmann 2007; see also Stieber 1958, Rospond 1971, Wierzchowska 1980). The same class has also been viewed – mostly by Anglo-Saxon and German researchers – as retroflex, the respective sounds being transcribed as [ʂ], [ʐ], [ʈ], and [ɖ] (cf. Keating 1991, Ladefoged & Maddieson 1996, Padgett & Zygis 2003, Hamann 2003 and 2004). While the former notation suggests a partially palatalized sound, the latter implies that the tongue shape is concave and apical or subapical. The actual realization of these consonants is, however, neither palatal(ized) nor fully retroflex, but rather laminal and flat – their closest IPA equivalents being [ʂ], [ʐ], [ʈs], and [ɖz] (cf. Hamann 2003).

<sup>50</sup>However, the two series of sibilants and affricates are not restricted to the vocabulary borrowed from Polish. They can also be used in genuine Germanic lexemes.

the war (Kleczkowski 1920), is a common cross-linguistic phenomenon. Strong palatalization tendencies affecting *n* operated in Eastern diphthongized Silesian dialects, sometimes even more widely than in Wymysioeryś (Waniek 1880: 32, 41, von Unwerth 1908: 39–40, Gusinde 1911: 96–98, 115, 144, Andrason 2021).<sup>51</sup>

Similarly, the presence of other palatal sounds in Wymysioeryś may be attributed to Polish influence as well as language- and family-internal processes. That is, Polish might have fortified palatalizing tendencies that were already operating in the Wymysioeryś language and its Silesian relatives. As a result, the visibility of palatal(ized) consonants was intensified, their central status in the phonetic and phonological system was established, and new palatalization rules were introduced to those already operating (see Section 4.6).

The development of the labialized velar approximant [w] from the velarized alveolar lateral approximant [ɬ] has also resulted from two drifts: language-external and language-internal (Andrason 2014b, 2015, 2021, Źak 2019). Polish has significantly intensified and perhaps accelerated the process whose foundations were already in place (for a similar view consult Selmer 1933: 234). On the one hand, the change seems to imitate an analogous development operating in Polish, i.e. the replacement of [ɬ] by [w], known under the term *walczanie*. The process appeared in Polish dialects in the 16<sup>th</sup> and 17<sup>th</sup> century. At the turn of the 19<sup>th</sup> and the 20<sup>th</sup> century, it spread beyond dialects to the standard language, where it became the norm in the second half of the 20<sup>th</sup> century (Urbańczyk 1991: 372, Gussmann 2007: 28).<sup>52</sup> Chronologically, the change of [ɬ] to [w] in Wymysioeryś coincides with the period of the full generalization of [w] in Standard Polish, which is also the time where the Polonization of Wilamowice increased substantially. On the other hand, the development of [ɬ] to [w] is found in other Central East (colonial) German varieties. It was, for example, attested in Lower Silesian and diphthongized Silesian dialects (von Unwerth 1908: 35, Gusinde 1911: 105, Selmer 1933: 233–234). In the Szynwałd/Bojków dialect, it had been established by the beginning of the 20<sup>th</sup> century (Gusinde 1911: 104–105, Kleczkowski 1920: 125, 161–162).<sup>53</sup> The same process could thus have been carried on in Wy-

<sup>51</sup>It also occurred in contexts where *n* appeared after a short vowel and before dental consonants. Compare *k'eńt* ‘children’ in Szynwałd/Bojków with *kynt* in Wymysioeryś (Gusinde 1911: 98, Kleczkowski 1920: 116) – contrary to Polish.

<sup>52</sup>Currently, the pronunciation of *l* as [ɬ] is perceived as “an affectation” (Gussmann 2007: 28). More regularly, it occurs only in east-southern dialects (Dubisz et al. 1995: 146; see also Nitsch 1957: 46–47, Źak 2019).

<sup>53</sup>However, as in Wymysioeryś, the change that took place in Szynwałd/Bojków is attributed to Polish influence; specifically, to the Polish Silesian variety used in the Upper Silesian coal basin and industrial region, where [ɬ] had earlier developed into [w] (Nitsch 1909: 156, Gusinde 1911: 104–105, Kleczkowski 1920: 126).

mysiöeryś. Furthermore, the development of [l] into [w] has occurred in other members of the Germanic family: varieties of Swiss German dialects, Thuringian dialects, Lusatian dialects, East-Low German dialects, Franconian dialects, and Low Franconian dialects (Selmer 1933, Besch et al. 1983: 1111–1112, Leemann et al. 2014).<sup>54</sup> It is indeed common from a cross-linguistic perspective, featuring not only in Slavonic and Germanic, but also in Romance and other language phyla (Żak 2019).

#### 4.5 Consonantal length

Polish contains geminated or long consonants. They occur in an intervocalic and word-initial position (Gussmann 2007: 241, Wągiel 2016: 82). Since a number of minimal pairs may be identified, geminated consonants play a phonemic role, at least peripherally (Wągiel 2016: 82).

Length is a pervasive – both synchronically and diachronically – feature of Germanic languages (Harbert 2007: 74–79). Geminate consonants arose in old and medieval Germanic languages, both in the Northern and Western branches, where they acquired a systemic relevance (Harbert 2007: 74–75). Subsequently, various languages underwent changes and long consonants have often been simplified (Lass 1992, Harbert 2007: 75–78)). This degemination is visible in the development from Middle High German to Modern Standard German and many other West Germanic languages (Harbert 2007: 76–78, Schmidt 2017). In modern languages, only North Germanic exhibits genuine long consonants (Harbert 2007: 78–79).

Rather than deriving directly from contact with Polish, the consonantal length in Wymysiöeryś most likely constitutes an inherited Germanic property, as it existed in Middle High German – the diachronic source of Wymysiöeryś. The Polish language could however have contributed to the maintenance of long consonants, thus preventing the developments that have taken place in many other modern West Germanic languages and German varieties.

#### 4.6 Palatalization

Polish exhibits various types of palatalizing processes and a wide range of palatalization-based oppositions. Polish has been viewed as one of “the most highly palatalized” languages in the entire Slavonic branch (Sussex & Cubberley 2006: 165), the one that attests to “a more advanced state of [...] palatalization than

---

<sup>54</sup>Often, however, the vocalic pronunciation of *l* in German varieties is regarded as influenced by Romance and Slavonic languages (Selmer 1933: 235–238, 243).

any of the other” members of this language family (Sussex & Cubberley 2006: 165). Indeed, the contrast between palatal(ized) consonants and non-palatal(ized) consonants – generally referred to as “soft” and “hard” respectively (Urbańczyk 1991: 244, Strutyński 1998: 43–44) – underpins not only the sound system of Polish but also the language’s morphology. Crucially, for all consonants, there is a corresponding palatal(ized) consonant, either at a phonemic or a phonetic level (Rothstein et al. 1993: 687–690, Sussex & Cubberley 2006: 165–166, Gussmann 2007: 4–7).<sup>55</sup>

Although certain types of palatalization have operated in the Germanic family, and palatal(ized) sounds feature relatively prominently in Dutch, Frisian, and Afrikaans (Hoekstra & Tierstna 1994: 529, Donaldson 1994: 482, van der Hoek 2010), as well as in Icelandic and Faroese (Barnes & Weyhe 2013: 193–195, Harbert 2007: 48–49), palatalization is not as essential a component of the Germanic sound system as, for example, aspiration. Its role in the phonetics and phonology of West Germanic languages is certainly less fundamental than is the case of Slavonic languages (see Harbert 2007: 48–49). Crucially, German (see Section 3.9) and most of its dialects fail to exploit palatalization and palatal(ized) consonants to an extent that would be comparable to that attested in Polish (and in *Wymysiöeryś*). As attested at the beginning of the 20<sup>th</sup> century, German dialects exhibited a slightly more palatalization-oriented character than Standard Modern German (von Unwerth 1908: 38–40, 53–54, 60, 71).<sup>56</sup>

Given the peripheral status of palatalization in German varieties and West Germanic languages in contrast to its central position in Polish and Slavonic languages, it is highly probable that the extensive use of palatal(ized) consonants in *Wymysiöeryś* and the central position of palatalization in its phonetic and phonological system, may be attributed to contact with Polish (see a similar conclusion in Kleczkowski 1920: 15 and Żak 2016: 136). This proposal is consistent with the scenario posited for Yiddish which acquired a wide array of palatal(ized) consonants most likely under the influence of Slavonic languages (Jacobs et al. 1994: 394, Harbert 2007: 26). Furthermore, two types of palatalizing processes seem to have been transferred directly from Polish, being absent in other colonial Central East German varieties: (a) regressive palatalization, i.e. a palatal(ized) pronunciation of consonants due to the presence of subsequent front vowels (contrary

<sup>55</sup>For an exhaustive list of “soft” and “hard” consonants consult Strutyński (1998). The phonemic status of palatal(ized) consonants is related to the status of the vowels *i* and *y* (Strutyński 1998: 77–78; Sussex & Cubberley 2006: 167). Regarding phonological and morphophonemic aspects of palatalization in Polish see Gussmann (2007).

<sup>56</sup>Apparently, the strongest palatal effects were found in diphthongized dialects, to which *Wymysiöeryś* belonged.

to the progressive palatalization typical of Silesian German; see next paragraph) and (b) a palatalizing process analogous to the so-called fourth palatalization, i.e. the development of *ky/ke* [k] and *gy/ge* [g] into *ki/kje* [c] and *gi/gje* [ʃ], respectively (Żak 2016: 136, see Dejna 1973: 124–129, Urbańczyk 1991: 244, Długosz-Kurczabowa & Dubisz 2006: 146–147).

However, although Wymysioeryś and Polish currently exhibit similar sets of palatal(ized) consonants and regressive palatalization operates both in Wymysioeryś and Polish, the two systems are not identical. The most relevant difference pertains to the manner with which various palatal(ized) consonants emerged. In genuine Wymysioeryś vocabulary, the palata(ized) realization of the consonant was – and still often is – conditioned by the vowel that precedes it (Kleczkowski 1920: 125) rather than by the vowel that follows, which is typical of Polish. The same principle governed palatalization in all Silesian German dialects thus revealing a firm family-internal tendency (von Unwerth 1908: 71).<sup>57</sup>

Overall, Polish might have fortified palatalizing tendencies that were already operating in the Wymysioeryś language and its Silesian relatives. As a result, the visibility of palatal(ized) consonants was intensified, their central status in the phonetic and phonological system was established, and new palatalization rules were added to those already existing.

#### 4.7 Onset clusters

Polish exhibits rich phonotactics, tolerating complex consonant clusters in onset positions (Gussmann 2007, Zydorowicz 2010: 567, Dziubalska-Kołaczyk & Zydorowicz 2014, Zydorowicz & Orzechowska 2017: 101) – a property that is characteristic of the Slavonic family, in general (Sussex & Cubberley 2006). Given that both the length of clusters and the number of combinations is “impressive” (Zydorowicz & Orzechowska 2017: 101), Polish is considered as “one of the most permissive languages” as far as phonotactics are concerned (Kijak 2008: 62). With regard to length, onset clusters tolerate maximally four elements (Zydorowicz 2010: 565, Dziubalska-Kołaczyk & Zydorowicz 2014, Zydorowicz & Orzechowska 2017: 98). With regard to combinatorics, 231 types of doubles, 165 triples, and 15 quadruples are found in onsets in Polish (Bargielówna 1950, Zydorowicz 2010: 565–567, Zydorowicz & Orzechowska 2017: 107–108). The richness of Polish phonotactics is not only quantitative but also concerns the qualitative properties of clusters. That is, Polish allows for onset clusters that exhibit falling sonority profiles (e.g.

<sup>57</sup>In a further contrast to Polish, in Silesian German – including the variety of Szynwald/Bojków – palatalization operated spontaneously before a dental consonant, either plosive, nasal, or lateral (von Unwerth 1908: 38–39, 68–69, Gusinde 1911: 98).

[rt]) and clusters with unchanged sonority values (the so-called plateau clusters; e.g. [fsx-]) in addition to those with rising sonority (e.g. [tr]) (Dukiewicz 1980, Zydorowicz & Orzechowska 2017: 104). Accordingly, sequences that are “ill-formed” from the perspective of the sonority scale (e.g. [rt], [rdz], [pstr]) are often tolerated (Zydorowicz & Orzechowska 2017: 104).

Even though onset clusters in Germanic languages can be complex (Harbert 2007, van Oostendorp 2019: 33), their complexity is lower than in Polish and Slavonic languages (cf. Kučera & Monroe 1968 who contrast German with Russian and Czech). Overall, only bi- and tri-segmental clusters are allowed in onsets (van Oostendorp 2019: 34–36). Even biconsonantal onsets, the most permissive ones, exhibit various combinatory restrictions (Harbert 2007, van Oostendorp 2019). For instance, onsets with a glide as their second element, onsets composed of sibilants and voiced obstruents, and the clusters [tl] and [dl] are generally disallowed (van Oostendorp 2019).<sup>58</sup> The most permissive language as far as phonotactics are concerned, is Yiddish (van Oostendorp 2019) – likely due to Slavonic influence. Three-segmental onsets are even more restricted and mainly appear with [s] and [ʃ] as the first element. With a few exceptions involving [s] and [ʃ], two- and three-consonant onsets must comply with sonority hierarchy (Harbert 2007: 68, 73, van Oostendorp 2019). This compliance is larger than what one observes in Polish.

The greater qualitative and quantitative restrictions exhibited by onsets in Germanic languages than is the case in Polish, as well as the fact that the most complex Wymysiöeryś onsets appear in Polish loanwords suggests the contact-induced increase in the complexity of onsets in Wymysiöeryś.

#### 4.8 Module-global perspective

The discussion in Section 4.1–Section 4.7 suggests that most of the surplus of complexity exhibited by Wymysiöeryś in the sound-system module can be attributed to contact with Polish. In case of four features (i.e. nasalization, consonants, palatalization, and onset clusters) contact with Polish is the principle reason for the complexity attested, although in some instances, enhancing the (more or less visible) tendencies already operating at a language- or family-internal level. In case of one feature (i.e. consonantal length), the Polish influence is secondary – it is the family-internal genetic drift that is the primary factor motivating the complexity surplus observed. Lastly, in case of two features (i.e. triphthongs and vocalic sonorants), Polish has not contributed, even minimally, to the complexity surplus exhibited by Wymysiöeryś.

---

<sup>58</sup>In contrast, all of these onset clusters are allowed in Polish.

Although contact with Polish has most likely contributed to the complexification of Wymysiöeryś, it is also responsible for its simplification in certain aspects. As explained in Section 3.10, in Germanic, the distinction between /p/, /t/, and /k/ and /b/, /d/, and /g/ involves primarily the feature of tenseness (Jessen 1998) or spread glottis (Harbert 2007: 44). Its typical acoustic effect is aspiration (Iverson & Salmons 1995, 1999, 2003, 2008, Harbert 2007: 44). Indeed, the spread-glottis principle – referred to as Germanic enhancement (Iverson & Salmons 2003: 44) – with aspiration effects one of the fundamental and “inherent” rules governing the sound system of Germanic languages. It has been operating since the development of the proto-language, being responsible for a series of changes and developments (Iverson & Salmons 2003: 44, Iverson & Salmons 2008: 2–4). The spread-glottis principle and aspiration are absent in Wymysiöeryś. This absence is most likely due to interaction with Polish where tenseness and aspiration have never operated. Overall, however, the contribution of contact with Polish to the complexity of Wymysiöeryś is by far more positive than negative.

## 5 Conclusion

This contribution demonstrates that Wymysiöeryś – a severely endangered moribund minority language – exhibits remarkable complexity. Therefore, its severe endangerment and moribund status are not correlated with structural simplicity – at least, in the variety used by fluent speakers.<sup>59</sup> The surplus of complexity is largely attributable to the transfer of elements from the dominant code, Polish. This confirms the view of language contact as not only having simplifying effects on languages, but also as contributing to their complexification – even in the situation of seemingly imminent language death.

The analysis of local complexities pertaining to diverse phonetic/phonological features (monophthongs, diphthongs, triphthongs, vocalic length, vocalic sonorants, nasality, consonants, consonantal length, palatalization, aspiration, onset and coda clusters) and their subsequent combination into a global relational value demonstrate the following: (a) locally, the complexity of Wymysiöeryś is typically superior or equal to that of Middle High German and Modern Standard German; (b) module-globally, the complexity of Wymysiöeryś is greater than

---

<sup>59</sup>As explained in Section 1, the semi-speakers of Wymysiöeryś, who did not learn the language properly in intergenerational transmission and rarely (if ever) use it, exhibit radical simplification and impoverishment. However, they have no bearing on general language use, the transmission of Wymysiöeryś to the younger generations, and the language’s structure overall.

the overall complexity of Middle High German and Modern Standard German; (c) both locally and globally, the surplus of information exhibited by Wymysiöeryś – and, thus, the positive difference in complexity when compared with the two control languages – can, in its largest part, be attributed to the contact with Polish. That is, by assimilating various Slavonic properties, and simultaneously maintaining its inherited or internally developed Germanic traits, the sound system of Wymysiöeryś is richer than the systems of its mother and at least some of its sister languages.

While this research only demonstrates the contact-induced complexification of the sound-system module of Wymysiöeryś, it is likely that a similar increase in complexity would be observed in other modules, whether morphology, syntax, or vocabulary. The likelihood of such complexifications is motivated by the general trend exhibited by Wymysiöeryś, namely the simultaneous maintenance of the Germanic foundation and its enhancement by Polish elements – a trend that goes beyond accidental complexity fluctuations (Andrason 2021). However, since in any given language, the complexities of different modules generally need not coincide, this hypothesis must be verified in future studies.

## References

- Aikhenvald, Alexandra. 2007. Grammars in contact: A cross-linguistic perspective. In Alexandra Aikhenvald & Robert Dixon (eds.), *Grammars in contact: A cross-linguistic typology*, 1–66. Oxford: Oxford University Press.
- Anders, Heinrich. 1933. *Gedichte von Florian Biesik in der Mundart von Wilamowice*. Kraków/Poznań: Uniwersytet Poznański.
- Andrason, Alexander. 2014a. Szkic gramatyczny języka wilamowskiego u progu jego wyginięcia. *Małopolska* 16. 119–135.
- Andrason, Alexander. 2014b. The Polish component in the Vilamovicean language. *Glossos* 12. [https://slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/media\\_items\\_files/andrason.original.pdf](https://slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/media_items_files/andrason.original.pdf). (25 June, 2022).
- Andrason, Alexander. 2015. Vilamovicean: A Germanic-Slavic mixed language? *Studies in Polish Linguistics* 10(2). 57–85.
- Andrason, Alexander. 2021. *Polish borrowings in Wymysorys: A formal linguistic analysis of Germano-Slavonic language contact in Wilamowice*. Reykjavík: University of Iceland. (Doctoral dissertation).
- Andrason, Alexander & Tymoteusz Król. 2016. *A grammar of Wymysorys*. Durham: Duke University, SEELRC. [https://slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/media\\_items\\_files/wymysorys-grammar.original.pdf](https://slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/media_items_files/wymysorys-grammar.original.pdf) (8 May, 2019).

- Andrason, Alexander, John Sullivan & Justyna Olko. Forthcoming. Language simplification in endangered languages? Inflectional categories of nouns and verbs in Eastern Huasteca Nahuatl and Wymysorys. *Studies in Language*.
- Ansaldo, Umberto & Stephen Matthews. 2007. Deconstructing creole: The rationale. In Umberto Ansaldi & Stephen Matthews (eds.), *Deconstructing creole*, 1–18. Amsterdam: John Benjamins Publishing.
- Austin, Peter. 1986. Structural change in language obsolescence: Some Eastern Australian examples. *Australian Journal of Linguistics* 6. 201–230.
- Bacher, Josef. 1905. *Die deutsche Sprachinsel Lusern*. Innsbruck: Wagner'sche Universitätsbuchhandlung.
- Badii, Remo & Antonio Politi. 1997. *Complexity: Hierarchical structures and scaling in physics*. Cambridge: Cambridge University Press.
- Bąk, Piotr. 1997. *Gramatyka języka polskiego*. Lublin: PWZN.
- Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall & Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26. 5–42.
- Bargiełówna, Maria. 1950. Grupy fonemów spółgłoskowych współczesnej polszczyzny kulturalnej. *Biuletyn Polskiego Towarzystwa Językoznawczego* 10. 1–25.
- Barnes, Michael & Eivind Weyhe. 2013. Faroese. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 190–218. London: Routledge.
- Besch, Werner, Ulrich Knoop, Wolfgang Putschke & Herbert Wiegand (eds.). 1983. *Dialektologie: Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Vol. 1.2. Berlin: De Gruyter. 807–1714.
- Biedrzycki, Leszek. 1974. *Abriß der polnischen Phonetik*. Warszawa: Wiedza Powszechna.
- Bloch-Rozmaj, Anna. 1997. Nasal vowels in Polish. *Papers and Studies in Contrastive Linguistics* 32. 81–103.
- Caratinini, Emilie. 2009. *Vocalic and consonantal quantity in German: Synchronic and diachronic perspectives*. Nice: Université de Nice-Sophia Antipolis & Universität Leipzig. (Doctoral dissertation).
- Chromik, Bartłomiej. 2016. From Wilamowice to the philosophy of science and back again: Three levels of linguistic ideologies. In Justyna Olko, Tomasz Wiślickiewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 91–109. Warszawa: Uniwersytet Warszawski.
- Chromik, Bartholomiej & Marek Dolatowski. 2013. *Halcnovian and Bielsko-Biała enclave*. <http://inne-języki.amu.edu.pl/Frontend/Language/Details/11> (8 May, 2019).

- Cilliers, Paul, Harry Biggs, Sonja Blignaut, Aiden Choles, Jan-Hendrik Hofmeyer, Graham Jewitt & Dirk Roux. 2013. Complexity, modeling, and natural resource management. *Ecology and Society* 18(3). 1–12.
- Clements, Nick. 1990. The role of sonority cycle in core syllabification. In J. Kingston & M. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 233–283. New York: Cambridge University Press.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins Publishing.
- Dahl, Östen. 2009. Testing the assumption of complexity invariance: The case of Elfdalian and Swedish. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 50–63. Oxford: Oxford University Press.
- de Boor, Helmut & Roswitha Wisniewski. 1973. *Mittelhochdeutsche Grammatik*. Berlin: De Gruyter Mouton.
- DeGraff, Michel. 2003. Against creole exceptionalism. *Language* 79(2). 391–410.
- DeGraff, Michel. 2005. Linguists' most dangerous myth: The fallacy of creole exceptionalism. *Language in Society* 34. 533–591.
- Dejna, Karol. 1973. *Dialekty polskie*. Wrocław: Ossolineum.
- Demenko, Grażyna. 1999. *Analiza cech suprasegmentalnych języka polskiego na potrzeby technologii mowy*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu.
- DeSchutter, Georges. 1994. Dutch. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 439–477. London: Routledge.
- Deutscher, Guy. 2009. Overall complexity: A wild goose chase? In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 243–251. Oxford: Oxford University Press.
- Długosz-Kurczabowa, Krystyna & Stanisław Dubisz. 2006. *Gramatyka historyczna języka polskiego*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Dodd, Bill, Christine Eckhard-Black, John Klapper & Ruth Whittle. 2003. *Modern German grammar*. London: Routledge.
- Dogil, Grzegorz. 1990. Hissing and hushing fricatives: A comment on non-anterior spirants in Polish. Unpublished manuscript, Stuttgart University. Stuttgart.
- Donaldson, Bruce. 1994. Afrikaans. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 478–504. London: Routledge.
- Dorian, Nancy. 1973. Grammatical change in a dying dialect. *Language* 49. 413–438.

- Dorian, Nancy. 1978. The fate of morphological complexity in language death: Evidence from East Gaelic. *Language* 54(3). 590–609.
- Dorian, Nancy. 1980. Language shift in community and individual: The phenomenon of the laggard semi-speaker. *International Journal of the Sociology of Language* 25. 85–94.
- Dubisz, Stanisław, Halina Karaś & Nijola Kolis. 1995. *Dialekty i gwary polskie*. Warszawa: Wiedza Powszechna.
- Dukiewicz, Leokadia. 1980. Rodzaje i częstość występowania nagłosowych grup fonemów typu CC(V)-, CCC(V)- i CCCC(V)- w polskich tekstach prasowych. *Lětopis Instituta za Serbski Ludospyt* 27(2). 142–157.
- Dziubalska-Kołaczyk, Katarzyna & Paulina Zydrorowicz. 2014. The production of high-frequency clusters by native and non-native users of Polish. *Concordia Working Papers in Applied Linguistics* 5. 130–144.
- Edmonds, Bruce. 1999. *Syntactic measures of complexity*. Manchester: University of Manchester.
- Eisenberg, Pater. 1994. German. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 349–387. London: Routledge.
- Fagan, Sarah. 2009. *German: A linguistic introduction*. Cambridge: Cambridge University Press.
- Filipović, Luna & Martin Pütz. 2016. Introduction: Endangered languages and languages in danger. In Luna Filipović & Martin Pütz (eds.), *Endangered languages and languages in danger: Issues of documentation, policy, and language rights*, 1–22. Amsterdam: John Benjamins Publishing.
- Foley, James. 1977. *Foundations of theoretical phonology*. Cambridge: Cambridge University Press.
- Fourquet, Jean. 1963. Einige unklare Punkte der deutschen Lautgeschichte in phonologischer Sicht. In Siegfried Gutenbrunner, Hugo Moser, Walter Rehm & Heinz Rupp (eds.), *Die Wissenschaft von deutscher Sprache und Dichtung. Methoden, Probleme, Aufgaben. Festschrift für Friedrich Maurer, zum 65. Geburtstag am 5. Januar 1963*, 84–90. Stuttgart: Ernst Klett Verlag.
- Fox, Anthony. 2005. *The structure of German*. Oxford: Oxford University Press.
- Gell-Mann, Murray. 1995. What is complexity? *Complexity* 1(1). 16–19.
- Gell-Mann, Murray & Seth Lloyd. 2004. Effective complexity. In Murray Gell-Mann & Constantino Tsallis (eds.), *Nonextensive entropy: Interdisciplinary applications*, 387–398. Oxford: Oxford University Press.
- Goblirsch, Kurt. 1997. From voice to length in High German consonants. *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* 2. 257–280.

- Goblirsch, Kurt. 2018. *Lenition and vowel lengthening in the Germanic languages: On the history of quantity in Germanic*. Cambridge: Cambridge University Press.
- Grinevald, Colette & Michel Bert. 2011. Speakers and communities. In Peter Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 45–65. Cambridge University Press.
- Gusinde, Konrad. 1911. *Eine vergessene deutsche Sprachinsel im polnischen Oberschlesien (die Mundart von Schönwald bei Gleiwitz)*. Breslau: M. & M. Marcus.
- Gussmann, Edmund. 2007. *The phonology of Polish*. Oxford: Oxford University Press.
- Hall, Tracy. 2017. Underlying and derived glides in Middle High German. *Glossa: A Journal of General Linguistics* 2(1). 1–31.
- Hall, Tracy Alan. 1992. *Syllable structure and syllable-related processes in German*. Tübingen: Niemeyer.
- Hall, Tracy Alan. 2000. *Phonologie: Eine Einführung*. Berlin: De Gruyter.
- Hamann, Silke. 2003. *The phonetics and phonology of retroflexes*. Utrecht: LOT.
- Hamann, Silke. 2004. Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association* 34. 53–67.
- Hammarström, Harald. 2008. Complexity in numeral systems with an investigation into pidgins and creoles. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 287–304. Amsterdam: John Benjamins Publishing.
- Harbert, Wayne. 2007. *The Germanic languages*. Cambridge: Cambridge University Press.
- Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Hennings, Thordis. 2012. *Einführung in das Mittelhochdeutsche*. Berlin: De Gruyter.
- Herzog, Marvin, Ulrike Kiefer, Robert Neumann, Wolfgang Putschke, Andrew Sunshine, Vera Baviskar & Uriel Weinreich. 1992. *The language and culture atlas of Ashkenazic Jewry: Historical and theoretical foundations*. Vol. 1. Tübingen: Max Niemeyer Verlag & YIVO Institute for Jewish Research.
- Hoekstra, Jarich & Peter M. Tierstna. 1994. Frisian. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 505–531. London: Routledge.
- Iverson, Gregory & Joseph Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12. 369–396.
- Iverson, Gregory & Joseph Salmons. 1999. Glottal spreading bias in Germanic. *Linguistische Berichte* 178. 135–151.

- Iverson, Gregory & Joseph Salmons. 2003. Laryngeal enhancement in early Germanic. *Phonology* 20. 43–72.
- Iverson, Gregory & Joseph Salmons. 2008. Germanic aspiration: Phonetic enhancement and language contact. *Sprachwissenschaft* 33. 257–278.
- Jacobs, Neil G., Ellen Prince & Johan van der Auwera. 1994. Yiddish. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 388–419. London: Routledge.
- Jacobs, Neil G., Ellen Prince & Johan van der Auwera. 2005. *Yiddish: A linguistic introduction*. Cambridge: Cambridge University Press.
- Jassem, Wiktor. 1973. *Podstawy fonetyki akustycznej*. Warszawa: PWN.
- Jassem, Wiktor. 2003. Polish. *Journal of the International Phonetic Association* 33. 103–108.
- Jessen, Michael. 1998. *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins Publishing.
- Johnson, Sally & Natalie Braber. 2008. *Exploring the German language*. Cambridge: Cambridge University Press.
- Juvonen, Päivi. 2008. Complexity and simplicity in minimal lexica. The lexicon of Chinook Jargon. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 321–340. Amsterdam: John Benjamins Publishing.
- Karaś, Mieczysław & Maria Madejowa. 1977. *Słownik wymowy polskiej*. Warszawa: PWN.
- Keating, Patricia. 1991. Coronal places of articulation. In Carole Paradis & Jean-François Prunet (eds.), *The special status of coronals: Internal and external evidence*, 29–48. San Diego: Academic Press.
- Kijak, Artur. 2008. *Polish and English consonantal clusters: A contrastive analysis within the strict CV framework*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Kleczkowski, Adam. 1920. *Dialekt Wilamowic w zachodniej Galicji: Fonetyka i fлексja*. Kraków: Polska Akademia Umiejętności.
- Kleczkowski, Adam. 1921. *Dialekt Wilamowic w zachodniej Galicji: Składnia i szkwyrazów*. Poznań: Uniwersytet Poznański.
- Kučera, Henry & George K. Monroe. 1968. *A comparative quantitative phonology of Russian, Czech, and German*. New York: Elsevier.
- Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht: LOT.
- Kusters, Wouter. 2008. Complexity in linguistic theory, language learning and language change. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.),

- Language complexity: Typology, contact, change*, 3–22. Amsterdam: John Benjamins Publishing.
- Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.
- Lasatowicz, Maria K. 1994. *Die deutsche Mundart von Wilamowice zwischen 1920 und 1987*. Opole: Wyższa Szkoła Pedagogiczna.
- Lass, Roger. 1992. Phonology and morphology. In Richard Hogg & David Denison (eds.), *The Cambridge history of the English language*, vol. 2, 23–155. Cambridge: Cambridge University Press.
- Latosiński, Józef. 1909. *Monografia Miasteczka Wilamowic: Na podstawie źródeł autentycznych*. Kraków: Drukarnia Literacka.
- Leemann, Adrian, Marie-José Kolly, David Britain, Iwar Werlen & Dieter Studer-Joho. 2014. The diffusion of /l/-vocalization in Swiss German. *Language Variation and Change* 26(2). 191–218.
- Lindström, Eva. 2008. Language complexity and interlinguistic difficulty. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 217–235. Amsterdam: John Benjamins Publishing.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Matras, Yaron & Jeanette Sakel (eds.). 2007. *Grammatical borrowing in cross-linguistic perspective*. Berlin: De Gruyter Mouton.
- McWhorter, John. 1998. Identifying the creole prototype: Vindicating a typological claim. *Language* 74. 788–818.
- McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5. 125–166.
- McWhorter, John. 2005. *Defining creole*. Oxford: Oxford University Press.
- McWhorter, John. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press.
- McWhorter, John. 2008. Why does a language undress? Strange cases in Indonesia. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 167–190. Amsterdam: John Benjamins Publishing.
- McWhorter, John. 2009. Oh no!: A bewilderingly multifunctional Saramaccan word teaches us how a creole language develops complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 141–163. Oxford: Oxford University Press.
- Meakins, Felicity, Xia Hua, Cassandra Algy & Lindell Bromham. 2013. Mixed languages. In Peter Bakker & Yaron Matras (eds.), *Contact languages: A comprehensive guide*, 159–228. Berlin: De Gruyter Mouton.

- Meakins, Felicity, Xia Hua, Cassandra Algy & Lindell Bromham. 2019. Birth of a contact language did not favor simplification. *Language* 95(2). 294–332.
- Mesthrie, Rajend, Joan Swann, Ana Deumert & William L. Leap. 2009. *Introducing sociolinguistics*. Edinburgh: Edinburgh University Press.
- Miestamo, Matti. 2006a. On the complexity of standard negation. In Mickael Suominen, Anu Airola Arppe, Orvokki Heinämäki, Matti Miestamo, Urho Määttä, Jussi Niemi, Kari K. Pitkänen & Kaius Sinnemäki (eds.), *A man of measure: Festschrift in honour of Fred Karlsson on his 60th birthday*, 345–356. Turku: The Linguistic Association of Finland.
- Miestamo, Matti. 2006b. On the feasibility of complexity metrics. In Krista Kerge & Maria-Maren Sepper (eds.), *Finest linguistics: Proceedings of the annual Finnish and Estonian conference of linguistics, Tallinn, May 6–7, 2004*. Tallinn: Tallinn University Press.
- Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins Publishing.
- Miestamo, Matti. 2009. Implicational hierarchies and grammatical complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins Publishing.
- Mitchell, Melanie. 2009. *Complexity: A Guided Tour*. Oxford: Oxford University Press.
- Mojmir, Herman. 1930–1936. *Wörterbuch der deutschen Mundart von Wilamowice*. Kraków: Polska Akademia Umiejętności.
- Moosmüller, Sylvia & Julia Brandstätter. 2015. Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect. *Language Science* 46. 84–95.
- Morciniec, Norbert. 1984. Die flämische Ostkolonisation und der Dialekt von Wilamowice in Südpolen. *Slavica Gandensia* 11. 7–18.
- Moseley, Christopher & Alexandre Nicholas. 2010. *Atlas of the world's languages in danger*. Paris: UNESCO Publishing.
- Mühlhäusler, Peter. 1977. *Pidginisation and simplification of language*. Canberra: Pacific Linguistics.
- Mühlhäusler, Peter. 1986. *Pidgin and creole linguistics*. Oxford: Blackwell.
- Neels, Rinaldo. 2016. Language and identity in Wilamowice: A complex history of language choices and language attitudes. In Justyna Olko, Tomasz Wicheriewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 111–130. Warszawa: Uniwersytet Warszawski.

- Newmeyer, Frederick J. & Laurel B. Preston. 2014. Introduction. In *Measuring Grammatical Complexity*, 1–13. Oxford: Oxford University Press.
- Niebaum, Hermann & Jürgen Macha. 1999. *Einführung in die Dialektologie des Deutschen*. Tübingen: Max Niemeyer Verlag.
- Nitsch, Kazimierz. 1909. Dialekty polskie śląska. *Materiały i Prace Komisji Językowej Akademii Umiejętności w Krakowie* 4. 85–356.
- Nitsch, Kazimierz. 1957. *Dialekty języka polskiego*. Wrocław: Zakład Narodowy imienia Ossolińskich.
- O'Brien, Mary & Sarah Fagan. 2016. *German phonetics and phonology: Theory and practice*. New Haven: Yale University Press.
- Padgett, Jay & Marzena Zygis. 2003. The evolution of sibilants in Polish and Russian. *ZAS Papers in Linguistics* 32. 155–174.
- Palosaari, Naomi & Lyle Campbell. 2011. Structural aspects of language endangerment. In Peter Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 100–119. Cambridge: Cambridge University Press.
- Parkvall, Mikael. 2008. Language complexity: Typology, contact, change. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), 265–285. Amsterdam: John Benjamins Publishing.
- Paul, Hermann. 2007. *Mittelhochdeutsche Grammatik*. 25th edn. Tübingen: Max Niemeyer Verlag.
- Peliti, Luca & Angelo Vulpiani (eds.). 1988. *Measures of Complexity: Proceedings of a Conference Held in Rome, September 30–October 2, 1987*. Berlin: Springer.
- Putschke, Wolfgang. 1980. Ostmitteldeutsch. In Hans Peter Althaus, Helmut Henne & Herbert Wiegand (eds.), *Lexikon der germanistischen Linguistik*, 474–479. Tübingen: Max Niemeyer Verlag.
- Rescher, Nicholas. 1998. *Complexity: A philosophical overview*. New Brunswick: Transaction Publishers.
- Ritchie, Carlo J. W. 2012. *Some considerations on the origins of Wymysorys*. [https://www.academia.edu/8290384/Some\\_Considerations\\_on\\_the\\_Origins\\_of\\_Wymysorys](https://www.academia.edu/8290384/Some_Considerations_on_the_Origins_of_Wymysorys). Sydney: University of Sydney. (BA Thesis). (1 January, 2021).
- Ritchie, Carlo J. W. 2016. Language visibility and Wymysorys. In Justyna Olko, Tomasz Wicherkiewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 65–79. Warszawa: Uniwersytet Warszawski.
- Rospond, Stanisława. 1971. *Gramatyka historyczna języka polskiego*. Warszawa: PAN.
- Polish. 1993. In Robert Rothstein, Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 686–758. London: Routledge.

- Rowicka, Grace J. & Joost M. van de Weijer. 1992. Nasal vowels in Polish. In Lambertus C. J. Barbiers, Marcel den Dikken & Claartje Levelt (eds.), *Proceedings of the Leiden conference for junior linguists*, vol. 3, 219–231. Leiden: Leiden University Press.
- Rubach, Jerzy. 1977. Nasalization in Polish. *Journal of Phonetics* 5. 17–25.
- Russ, Charles V. J. 1994. *The German language today: A linguistic introduction*. London: Routledge.
- Sallabank, Julia. 2012. Diversity and language policy for endangered languages. In Bernard Spolsky (ed.), *The Cambridge handbook of language policy*, 100–123. Cambridge: Cambridge University Press.
- Sallabank, Julia. 2013. *Attitudes to endangered languages: Identities and policies*. Cambridge: Cambridge University Press.
- Schmidt, Hans. 2017. *Einführung in die deutsche Sprachgeschichte*. Stuttgart: J.B. Metzler.
- Seiler, Guido. 2005. Wie verlaufen syntaktische Isoglossen, und welche Konsequenzen sind daraus zu ziehen? In Eckhard Eggars, Jürgen Erich Schmidt & Dieter Stellmacher (eds.), *Zeitschrift für Dialektologie und Linguistik. Beihefte*, vol. 130, 313–341. Stuttgart: Franz Steiner Verlag.
- Seiler, Guido. 2009. Sound change or analogy? Monosyllabic lengthening in German and some of its consequences. *Journal of Comparative German Linguistics* 12. 229–272.
- Selmer, Carl. 1933. Velarization and *u*-vocalization of *l* in German dialects. *Transactions and Proceedings of the Modern Language Association of America* 48(1). 220–244.
- Seuren, Pieter A. M. & Herman C. Wekker. 2001. Semantic transparency as a factor in creole genesis. In Pieter A. M. Seuren (ed.), *A view of language*, 421–431. Oxford: Oxford University Press.
- Shalizi, Cosma. 2006. Methods and techniques of complex science: An overview. In Thomas S. Deisboeck & Yasha Kresh (eds.), *Complex systems science in biomedicine*, 33–114. New York: Springer.
- Siegel, Jeff. 2008. Pidgins/creoles and second language acquisition. In Silvia Kouwenberg & John V. Singler (eds.), *The handbook of pidgin and creole studies*, 189–218. Oxford: Wiley-Blackwell.
- Silva-Corvalán, Carmen. 1995. The study of language contact: An overview of the issues. In Carmen Silva-Corvalán (ed.), *Spanish in four continents: Studies in language contact and bilingualism*, 1–13. Washington: Georgetown University Press.

- Simmler, Franz. 1985. Phonetik und Phonologie, Graphetik und Graphemik des Mittelhochdeutschen. In Werner Besch, Oskar Reichmann & Stefan Sonderegger (eds.), *Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*, vol. 2, 1129–1138. Berlin: De Gruyter.
- Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 67–88. Amsterdam: John Benjamins Publishing.
- Sinnemäki, Kaius. 2009. Complexity in core argument marking and population size. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 126–140. Oxford: Oxford University Press.
- Sinnemäki, Kaius. 2011. *Language universals and linguistic complexity: Three case studies in core argument marking*. University of Helsinki: University of Helsinki. (Doctoral dissertation).
- Sinnemäki, Kaius. 2014. Complexity trade-offs: A case study. In Frederick J. Newmeyer & Laurel B. Preston (eds.), *Measuring grammatical complexity*, 179–201. Oxford: Oxford University Press.
- Spencer, Andrew. 1986. A non-linear analysis of vowel-zero alternations in Polish. *Journal of Linguistics* 22. 249–280.
- Stevenson, Patrick. 1997. *The German-speaking world: A practical introduction to sociolinguistic issues*. London: Routledge.
- Stieber, Zdzisław. 1958. *Rozwój fonologiczny języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Strutyński, Janusz. 1998. *Gramatyka polska: Wprowadzenie, fonetyka, fonologia, morfologia, składnia*. Kraków: Tomasz Strutyński.
- Sussex, Roland & Paul Cubberley. 2006. *The Slavic languages*. Cambridge: Cambridge University Press.
- Swiggers, Pierce. 2007. Two key concepts of language endangerment: Language obsolescence and language death. *Linguistica* 47. 21–33.
- Práinsson, Höskuldur. 1994. Icelandic. In Ekkehard König & Johan van der Auwerda (eds.), *The Germanic languages*, 142–189. London: Routledge.
- Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.
- Trudgill, Peter. 1996. Dual-source pidgins and reverse creoloids: Northern perspectives on language contact. In Ernst H. Jahr & Ingvild Broch (eds.), *Language contact in the Arctic: Northern pidgins and contact languages*, 5–14. Berlin: De Gruyter Mouton.
- Trudgill, Peter. 2010. Contact and sociolinguistic typology. In Raymond Hickey (ed.), *The handbook of language contact*, 299–319. Oxford: Wiley-Blackwell.
- Urbańczyk, Stanisław. 1991. *Encyklopedia języka polskiego*. Wrocław: Ossolineum.

- van der Hoek, Michel. 2010. *Palatalization in West Germanic*. University of Minnesota. (Doctoral dissertation).
- van Ness, Silke. 1994. Pennsylvania German. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 420–438. London: Routledge.
- van Oostendorp, Marc. 2019. Germanic syllable structure. In Michael T. Putnam & Richard Page (eds.), *Cambridge handbook of Germanic linguistics*. Cambridge: Cambridge University Press.
- Phonetik. 1980. In Werner Veith, Hans-Peter Althaus, Helmut Henne & Herbert E. Wiegand (eds.), *Lexikon der germanistischen Linguistik*, 129–137. Tübingen: Max Niemeyer Verlag.
- Velupillai, Viveka. 2015. *Pidgins, creoles and mixed languages. An introduction*. Amsterdam: John Benjamins Publishing.
- von Unwerth, Wolf. 1908. *Die schlesische Mundart in ihren Lautverhältnissen grammatisch und geographisch dargestellt*. Breslau: M. & M. Marcus.
- Wągiel, Marcin. 2016. *Fonematyka języka polskiego w ujęciu funkcjonalizmu ak-sjomatycznego*. Ołomuniec: Uniwersystet im. Palckiego w Ołomuńcu.
- Waniek, Gustav. 1880. *Zum Vocalismus der schlesischen Mundart*. Bielitz: W. Fröhlich.
- Weckwerth, Jarosław. 2015. *Poland's Linguistic Heritage Documentation Database for Endangered Languages*. <http://www.inne-jezyki.amu.edu.pl> (15 May, 2019).
- Weddige, Hilkert. 2007. *Mittelhochdeutsch: Eine Einführung*. München: C.H.Beck.
- Weinreich, Max. 2008. *History of the Yiddish language*. Vol. 2. London: Yale University Press.
- Wicherkiewicz, Tomasz. 1998. Wir kuza ny mejr wymysojerysz. Stan etnolektu miasteczka Wilamowic u końca XX wieku. In Zdzisława Staszewska (ed.), *Miasto. Teren koegzystencji pokoleń*, 236–242. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Wicherkiewicz, Tomasz. 2003. *The making of a language: The case of the idiom of Wilamowice, Southern Poland*. Berlin: De Gruyter Mouton.
- Wicherkiewicz, Tomasz & Justyna Olko. 2016. Researching, documenting and reviving Wymysioeryś: A historical outline. In Justyna Olko, Tomasz Wicherkiewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 17–53. Warszawa: Uniwersytet Warszawski.
- Wierzchowska, Bożena. 1980. *Fonetika i fonologia języka polskiego*. Wrocław: Osolineum.
- Wiese, Richard. 1996. *The phonology of German*. Oxford: Oxford University Press.
- Wiesinger, Peter. 1980. Deutsche Sprachinseln. In Hans-Peter Althaus, Helmut Henne & Herbert Wiegand (eds.), *Lexikon der germanistischen Linguistik*, 491–500. Tübingen: Max Niemeyer Verlag.

- Wiesinger, Peter. 1983. Deutsche Dialektgebiete außerhalb des deutschen Sprachgebiets: Mittel-, Südost-, und Osteuropa (mit einem Anhang von Heinz Kloss). In Werner Besch, Ulrich Knoop, Wolfgang Putschke & Herbert E. Wiegand (eds.), *Dialektologie: Ein Handbuch zur deutschen und allgemeinen Dialektforschung*, vol. 2, 900–929. Berlin: De Gruyter Mouton.
- Wright, Joseph. 1917. *Middle High German Primer*. Oxford: Clarendon Press.
- Żak, Andrzej. 2016. The influence of the Polish language upon Wymysorys in light of the revitalization process. In Justyna Olko, Tomasz Wicherkiewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 131–149. Warszawa: Uniwersytet Warszawski.
- Żak, Andrzej. 2019. O pewnym szczególe z wilamowskiej fonetyki – próba analizy w kontekście arealnym. *Studia z Filologii Polskiej i Słowiańskiej* (54). 1–15.
- Zieniukowa, Jadwiga & Tomasz Wicherkiewicz. 1997. Do-Pyjter-Jańska i Cym-Hala-Mockja oraz inne zjawiska z antroponimii Wilamowic w dawnym Księstwie Oświęcimskim. In Hanna Popowska-Taborska & Jerzy Duma (eds.), *Onomastyka i dialektologia: Prace dedykowane pani Profesor Ewie Rzetelskiej-Feleszko*, 307–316. Warszawa: Instytut Sławistyki PAN.
- Zieniukowa, Jadwiga & Tomasz Wicherkiewicz. 2001. Sutuacja etnolektu wilamowskiego jako enklawy językowej. In Antoni Barciak (ed.), *Wilamowice. Przyroda, historia, język, kultura, oraz społeczeństwo miasta i gminy*, 489–519. Wilamowice: Urząd Gminy.
- Zydrorowicz, Paulina. 2010. Consonant clusters across morpheme boundaries: Polish morphonotactic inventory and its acquisition. *Poznań Studies in Contemporary Linguistics* 46(4). 565–588.
- Zydrorowicz, Paulina & Paula Orzechowska. 2017. The study of Polish phonotactics: Measures of phonotactic preferability. *Studies in Polish Linguistics* 12(2). 97–121.

# Chapter 9

## Language variation, language myths, and language ideology as constructive elements of the Wymysiöeryś ethnolinguistic identity

Tomasz Wicherkiewicz<sup>a</sup>

<sup>a</sup>Adam Mickiewicz University Poznań

Wymysiöeryś is a highly endangered language spoken today by the last few speakers and a growing number of new speakers in the small town of Wilamowice. After two generations of language decay and attrition, the language (community) has recently undergone intensive revitalisation processes.

In the past, the town Wilamowice formed a part of the so-called Bielsko-Biała linguistic enclave (*Bielitz-Bialaer Sprachinsel*), which had its roots in the massive German colonisation in the 12/13<sup>th</sup> century and created and/or populated several villages and towns in the border areas of Silesia and Galicia. In modern times, Wilamowice has constituted either a peripheral or an entirely distinct enclave (out) of the *Sprachinsel*.

The Wymysiöeryś language has been classified as a colonial variety of East Central German. Both sociolinguistic research and historical records indicate, however, that at various periods, different ideas on the origin and identity of the community have been shared and uttered by and on the Wilamowiceans and Wymysiöeryś. Such ethnotheories of provenance, including some folk linguistic evidence and myths, referred to various Germanic countries as places of origin of the first settlers.

From a contact linguistic perspective, the microlect of Wilamowice has certainly undergone interactions of various types and intensities with Polish (and its varieties) and standard High German. The evidence of such contacts, shift or even hybridisation can be found in all subsystems of the microlanguage; however, it is a rather perceptual dialectology and ethnoscience perspective of language variation which has been adopted in the present outline.

Tomasz Wicherkiewicz. 2022. Language variation, language myths, and language ideology as constructive elements of the Wymysiöeryś ethnolinguistic identity. In Matt Coler & Andrew Nevins (eds.), *Contemporary research in minoritized and diaspora languages of Europe*, 261–280. Berlin: Language Science Press. DOI: 10.5281/zenodo.7446971



## 1 Introduction

### 1.1 Concepts

Wilamowice is the unique home to the speech community of Wymysioeryś (ISO *wym*),<sup>1</sup> a severely endangered Germanic microlanguage, now spoken by several tens of users, both the local last speakers and a growing group of new speakers (in a 2009 UNESCO report, Wymysioeryś was referred to as "seriously endangered" and almost extinct).

Usually, the term *microlanguage* describes a language variety used by a numerically small and geographically or culturally isolated speech community. The geographical isolation may result from, e.g., peripheral, transborder, or insular location, while the cultural one from religious, ethnic, or social factors. In the case of many *speech microcommunities*, the criteria of territorial and social separation have been of decisive importance for both the internal and external identity of the group.

The term *microlanguage* is used here in its most neutral possible sense, although relating to two other terms used in the Slavic and Germanic sociolinguistic tradition: *literary microlanguage* (Russian: *микроязык*) and *language enclave* (German: *Sprachinsel*) respectively.

The former, introduced by Aleksandr Dulichenko (e.g. 2006: 27), refers to "such a form of an existing language (or dialect), which has been used in written texts and is characterised by normalizing trends emerging as a result of the functioning of the literary writing form within the framework of a more or less organised literary and linguistic process". Slavic linguistics uses the term mainly regarding those literary and linguistic forms that exist alongside the major Slavic languages of historically prominent nations and often possess a written form, with a certain degree of standardisation, and are used in a limited variety of circumstances and always alongside a national literary language. Genetically, Slavic literary microlanguage refers genetically to one of the major Slavic languages. According to Balowska (1999: 43), among the key criteria for distinguishing micro- from macrolanguages are the number of users, the degree of normalisation achieved by the language code, as well as the extent of its polyvalence. Some Slavic linguists,

---

<sup>1</sup>At the request of Tymoteusz Król, a young language activist from Wilamowice, the US Library of Congress added Wymysioeryś to the register of languages in July 2007. Later, the International Organisation for Standardisation registered it under the ISO 639-3 code *wym*. At the turn of the 20<sup>th</sup> century, efforts to obtain ISO codes became one of the procedures aimed at recognizing the linguistic status of language varieties.

however, question the sense of distinguishing such a category from among minority languages (Stern 2018).

The category of literary microlanguages has become a significant (research) domain in Slavic linguistics, and recently also in other areas of minority studies. That growing attention and applicability is probably a direct result of the growing role of written language standardisation for minority language preservation, maintenance, and revitalisation strategies. Quite underresearched remains, however, the development and course of language contacts between such *literary microlanguages* and their roofing *national* literary (standard) languages. Yet, in the case of close genetic linguistic relationship between the microlanguage and its *Dachsprache*,<sup>2</sup> language contacts have often been neglected by researchers. On the other hand, contacts between a microlanguage and its roofing literary national standard might be of a completely different range when the two lects are not close genetic cognates, as is the case between the substantially Germanic Wymysiöeryś and its modern *Dachsprache* Polish, especially when considering other language contacts of the microlanguage in the past (with Silesian German, standard German, Polish dialects, etc.)

Admittedly, Wymysiöeryś has not been extensively studied as a *literary microlanguage sensu stricto*, yet the role of Wilamowicean literature and literacy in its modern history (and historical sociolinguistics) has been quite crucial. It is the growing corpora of Wymysiöeryś literary texts and on Wilamowice speech community that keep developing into the main source of language material for comparative or contact linguistics, as well as sociolinguistics of the speech community, including the historical analysis of language ideologies and their impact upon the actual language system of Wymysiöeryś.<sup>3</sup>

Another keyword essential for researching and understanding the role of language variation in the development of Wilamowice's language ideology is *Sprachinsel* (German for 'language enclave'). However, the term and concept of *Sprachinsel* is still problematic, mainly due to its strongly ideological contents and methodological context in the past (cf. below). Numerous attempts which have been made to devise a definition are closely connected with certain research traditions on the one hand, and with the history of individual *Sprachinseln* on the other. A widely accepted definition among German linguists was suggested by Wiesinger (1983: 901), who asserted that language islands were either discrete or

---

<sup>2</sup>*Dachsprache* (German for 'roofing language', or 'umbrella language') is a term proposed by Heinz Kloss in reference to a language form that serves as a standard language in a country (or another polity with the same standard literary language) above other lects/idioms, regardless of their genetic affiliation or position in the actual dialect continuum (cf. e.g. Kloss 1967).

<sup>3</sup>For an outline of literary output in Wymysiöeryś, cf. Wicherkiewicz 2019.

areal, relatively small, enclosed linguistic settlements situated within a relatively large area where another language is used. Other definitions have stressed the aspects of isolation and *roofing* by a majority language, as, e.g., Protze (1995: 55) “language islands are mainly characterised by minimal contact with not only the motherland but also the surrounding speech community. The minority language is roofed by the standard variety of the majority language and the communities are linguistically and culturally isolated”.

As summed up in Claudia M. Riehl’s (2010: 333) outline on *Sprachinseln* (interpreted as *discontinuous language spaces*), “difficulties in defining a Sprachinsel are primarily due to the fact that the term cannot be readily detached from the historical-political constellations and social changes connected with the emergence of language islands (i.e., the colonisation of Eastern Europe and the New World)”. Nonetheless, the concept of *Sprachinsel* might be indispensable for a better discernment of both the non-existent *Bielitz-Bialer Sprachinsel* macro-enclave and its erstwhile exclave of Wilamowice.

A combined, thus interdisciplinary, definition by Klaus Mattheier (1996) may be quoted here:

“A language island is a linguistic community which develops as a result of disrupted or delayed linguistic cultural assimilation. Surrounded or enclosed by a linguistic and/or ethnic dominant culture, such an island is made up of a linguistic minority that has become separated from its original roots; and, as a function of its unique socio-psychological disposition (i.e., its “island mentality”) is held separate and apart from the majority culture with which it maintains tangential contact (...)"

The main characteristic of the *delayed assimilation* is based, among others, on the speakers’ attitude toward the linguistic majority. Therefore, language islands exist “rather in the minds of speech islanders than on the landscape” (Mattheier 1994: 106).

## 1.2 Location

The town of Wilamowice (*Wymysoü* in *Wymysiöeryś*) is currently situated in the Southern Poland province of Silesia, county of Bielsko-Biała, and inhabited by ca. 3 thousand people. The town is also the administrative center for the eponymous commune with the total population of ca. 18 thousand. Worth stressing, however, is the actual lack of cultural and linguistic continuity with the other settlements in the municipality, with one exception the village of Stara Wieś (*Wymysdiöf*), referred to as a medieval sister settlement of Wilamowice.

The community was founded most probably around 1300 as a result (of multiple waves) of the German eastward expansion, which could have also included migrants from Germanic-speaking lands of Flanders, Friesland, or even Romance Wallonia. The colonists established a cluster of settlements circumjacent to what was to become the towns of Bielitz/Bielsko and Biała/Biala, on both sides of the river Biała/Bialka, which constituted a centuries-old bunch of boundaries between various polities or macroregional districts, as duchies, dioceses, kingdoms, crown-lands, or state borders between partitioners of Poland.<sup>4</sup>

The colonies developed into what used to be called the *Bielitz-Bialaer Sprachinsel* in the German historical dialectology, i.e., a mixed urban-rural complex with its own cultural profile and a dialect-cluster consisting of several subvarieties as markers of both extra-and intra-group identities. The enclave broke up into the *Bielitz-Biala Sprachinsel* proper and Wilamowice (as a secondary sub-exclave) as a consequence of Polonisation of some villages such as Pisarzowice, which has since separated Wilamowice from the rest of the Sprachinsel until the 1940s. The majority populations of those two towns were German-speaking, as were the inhabitants of all villages which formed the proper *Sprachinsel*. However, most of those who had inhabited the Bielsko-Biała language enclave were aware that the language varieties they spoke were different from the standard *Hochdeutsch*. The varieties shared the endolinguonym *Pauerisch/Päuersch*, referring to the rural, *peasant* character of the ring of villages which surrounded the urban multilingual core of Bielsko and Biala (those two towns were merged twice in their history: 1941-1945 by the Nazi German occupants, and in 1951 now by the Polish administration). Always a microcosm in itself, either as an autonomous duchy or located at the boundaries of Prussia and Austria, who had partitioned Poland in the 18<sup>th</sup> century, Bielsko-Biała *sensu largo* was long considered an archetypic paradigm of a German *Sprachinsel*. In fact, the varieties under concern formed a local dialectal continuum, with the northeasternmost extreme in Wilamowice. When the continuum was broken, and Wilamowice got separated territorially and culturally (probably in the 18<sup>th</sup> century) from the rest of the enclave, the village of Hałcnów/Alzen/Alca<sup>5</sup> became the geographically and culturally nearest extreme of the German *Sprachinsel*. The relations with the inhabitants of its “sister” community of Hałcnów became the closest ‘counter-reference’ of Wilmowicean identity until the end of World War II. The two speech communities, two

---

<sup>4</sup>For a detailed outline of socio-cultural and glottopolitical history see:

<http://inne-jazyki.amu.edu.pl/Frontend/Language/Details/10> and

<http://inne-jazyki.amu.edu.pl/Frontend/Language/Details/11> (30 December, 2020).

<sup>5</sup>Polish/German/Halcnian name of the village; nowadays a quarter of the Bielsko-Biała conurbation.

settlements, and two language varieties shared quite a lot of linguistics and extralinguistic characteristics, but it is the German identity of Hałcnowians that became a decisive factor of their doom after 1945.

As described by Mętrak (2019: 10) in his contrastive-comparative study of those two communities:

“(...) modern national conflicts arrived in the region when Poland regained independence in 1918. The whole Bielsko-Biała enclave became a part of the Polish state, but as one of the centers of German minority it posed a problem for the state policies. Both Polish and German activists tried to convert the local population to their cause. In the interwar period, German nationalist-oriented scholars and ideologists treated the linguistic enclave inhabitants as exemplary “Arch-Germans”, who have been struggling for centuries to preserve their identity surrounded by Slavic “barbarians”.

Most of the German population of Bielsko-Biała fell for the Nazi propaganda (as seen in the voting results of the Young-German Party),<sup>6</sup> and the German army arriving to Bielsko-Biała on September 3<sup>rd</sup>, 1939 was greeted as liberators. The inhabitants of Hałcnów, however, traditionally supported the Christian Democratic Party. When the war was lost for Germany, some of the people evacuated with the fleeing Wehrmacht, others awaited the advancing Red Army. When in February 1945 Soviet troops marched into the city, its German population was subjected to harsh persecution: arrests, murders, rapes, and looting. Out of almost 50 thousand Germans of Bielsko-Biała, only a few «indispensable» specialists or pro-Polish activists were allowed to stay as the rest was forcibly resettled to Germany.

The only settlement of the former enclave whose inhabitants were not officially persecuted was Wilamowice, due to their non-German identity. (...) The lack of state-organised repression has not stopped discrimination on the local level, often performed by Poles from neighboring villages, envying the Vilamovicean wealth.”

The very economic wealth of Wilamowicea(ns) has been a decisive factor in the modern history of the development of a separate Sprachinsel-identity and its microlanguage ideology.

---

<sup>6</sup> *Jungdeutsche Partei in Polen* – a National Socialist political party founded in 1931 by members of the German ethnic minority in Poland.

## 2 Classifications and Myths of Origin

Genetic classifications and scientific taxonomies rarely coincide with the folk linguistic theories of language origin and language variation, or with perceptual classification of language varieties (lects). Furthermore, microlanguage communities have been both objects and subjects of quite repetitive discourses on *language-or-dialect* status, which actually followed and resulted from the unending debates and agendas concerning the national declarations of communities without their (kin) nation-states. These issues have actually been most visible and influenced the modern history of the entire Central-Eastern Europe, as well as abundant individual cases within.

Wilamowice, however, thanks to an exceptional tangle of historical, cultural, political, linguistic, economic, and even individual developments, happened to be the only survivor on that Central-European battlefield of languages, nation-states and *Sprachinseln*, macro- and microideologies, into the 21<sup>st</sup> century, one of the most salient factors being the overt and covert, documented and mythical history of the microlanguage.

The actual constellations of languages, patterns of individual and community multilingualism and polyglossia in Wilamowice, as well as in (other parts of) *Bielitz-Bialer Sprachinsel* still require thorough research.<sup>7</sup> There have been, however, some earlier contributions to the historical sociolinguistics of Wymysioeryś that focused on the ethnotheory of origin, shared by and imposed upon the community. To investigate and understand the current position and self-identification of Wilamowiceans, one should sort and analyze the myths and theories that have referred to and tried to categorise Wymysioeryś as a language (variety).

Some of them have already been researched earlier by Morciniec (1984), Ryckebroer (1984), Morciniec (1995), Wicherkiewicz (2003), Ritchie (2012), Wicherkiewicz & Olko (2016) or Wicherkiewicz et al. (2018).

A cutting-edge, comprehensive study of linguistic micro- and macroideologies of and in Wilamowice was compiled by Chromik (2019).<sup>8</sup> Using the approach based on Michael Silverstein's methodology, Chromik set down a monograph of linguistic ideologies understood as a bridge between linguistic and social structures, proposing a twofold typology of linguistic macroideologies (which attempt

<sup>7</sup>Actually, these questions constitute one of the main topics of the research project *Multilingual worlds – neglected histories. Uncovering their emergence, continuity and loss in past and present societies* (MULTILING-HIST), to be started in 2021 by Justyna Olko and her team (including the present author), and financed by a European Research Center grant.

<sup>8</sup>In Polish, although an English-language edition is expected.

to hegemonise the way languages are perceived), and microideologies (as an alternative to the former):

“Due to specific conditions, they are easier to observe in communities using minority languages or nonstandard forms of dominant languages. Unlike macroideologies, they are predominantly transmitted orally within families. As a consequence, they often remain invisible to sociologists, linguists, or historians. They are easier to observe when an ethnographic method, based on direct interaction with people, is used.”<sup>9</sup>

During the long 19<sup>th</sup> and 20<sup>th</sup> centuries, these language microideologies in Wilamowice have frequently been directly interwoven with (mythical) ethnotheories which pertained to the origin of the settlers in Wilamowice. Moreover, during the 20<sup>th</sup> century, they also got tangled with and into the terminological and taxonomical disputes on the linguistic status of Wymysiöeryś (recently also in the never-ending and unproductive debate on *language-vs.-dialect* status).

Therefore, it is language ideology that became an indispensable factor and link to the understanding of the sociolinguistic processes *sensu largo* within the Wilamowicean community, as indeed they significantly influenced the role, prestige, status and attitudes to the native language variety. Even the systemic language variation might be considered not only a consequence but, to some extent, even a product of language ideologies, as most paradigms of historical sociolinguistics do not ignore such variables as language ideologies, language attitudes or social networks as markers or factors of language variation.

### 3 Language ideologies as variables of “Wilamowiceanness”

In the last two centuries, the community of Wilamowice has constantly been subject to constellations of internal and external language ideologies that resulted in the ethnolinguistic choices and declarations, especially during the long 20<sup>th</sup> century. Many linguists (as well as dialectologists and historians) often associated the Bielitz-Bialer Sprachinsel with Silesian German. It must be remembered, however, that the emergence of those varieties is to be ascribed to the medieval settlers who came from various regions of today’s Germany, not to speak about adjacent areas. Silesian, thus, developed an admixture of dialects which consisted of elements from various German varieties. Initially, it probably had been much

---

<sup>9</sup>Quoting the English summary of unpublished Chromik 2019.

more diversified than it was in the 20<sup>th</sup> century. Throughout the centuries of co-existence of various groups, their lects converged. The scholars of language and history did some attempts to reach into the depths of history and determine what particular German dialects influenced the *Bielitz-Bialaer Mundart(en)*. Some of them (e.g., Wagner 1935: 192) referred to the apparent influence of Thuringian and Upper Saxon as the base of Silesian German varieties in the cluster of Bielsko-Biała.

The situation of Wymysiöeryś was different. Actually, every linguist (including philologists, dialectologists and amateur *Heimatkundige*) interested in the (linguistic) history of Wilamowice attempted to explain or determine the origins of Wymysiöeryś. Frequently, those explanations were either based on folk theories or created new myths themselves. One of the proposed hypotheses indicated Schaumburg in north-western Germany (Lower Saxony) as a possible place of origin of the first Wilamowiceans. The theory found a quite a wide distribution at the beginning of the 20<sup>th</sup> century. Józef Latosiński, a teacher and schoolmaster in Wilamowice, in his massively popular *Monografia Miasteczka Wilamowic* (based on authentic sources; 1909), argued that:

“ (...) the first Wilamowiceans did, indeed come from there; firstly, because of the local dialect which had both the native Low German element and Dutch and Anglo-Saxon features as well, and secondly, because to this day the textile industry has been thriving in the Principality of Schaumburg-Lippe (...)”

J. Latosiński's monograph has effectively influenced and to some extent constructed the collective identity of Wilamowice in the face of nationalisms breaching through Central Europe and restructuring the identities of lands and ethnic groups (more in: Wicherkiewicz & Olko 2016: 23ff, and Chromik 2019). The importance of Latosiński's thesis has been crucial for the Wilamowice's group identity within the 20<sup>th</sup> century, emphasised by the fact that the *Monografia* was reprinted in 1990.

What Latosiński's input meant for constructing the Polishness of Wilamowice(ans) was Walter Kuhn's mission-work in respect to their Germanness. Born in Bielitz, W. Kuhn was a prominent and productive ethnohistorian of the German *Ostgebiete* (Eastern territories), which included the areas colonised (also discretely, as scattered *Sprachinseln*) by Germans since the Middle Ages (cf. Wicherkiewicz & Olko 2016: 23ff). His 1934 monograph on *Deutsche Sprachinselforschung: Geschichte, Aufgaben, Verfahren* became an actual foundation of the entire discipline and research paradigm of *Sprachinselforschung*, an interdisciplinary study

of German language exclaves in Central and Eastern Europe, with Bielitz-Biala and vicinities as the actual archetype of a *Sprachinsel*. Himself an involved and convinced Nazi, professor of Universities of Breslau and Hamburg, Kuhn did his best to prepare and later support the ethnic cleansings of local non-German populations and the settlement of German colonists to Germanise Central and Eastern Europe. His texts on Wilamowice endeavored to explicitly display the Germanness of Wilamowice, considering any other theories and ethnic myths as inventions of nationalist Polish propaganda.<sup>10</sup>

Many sources and records have revealed that since the beginning of modern European state nationalism, Wilamowiceans rarely overtly considered themselves Germans. During modern history, i.e., as Habsburg Austrian subjects, they hardly had to declare their ethnicity; instead, the identity of belonging to the Austrian political nation has been expressed for generations: “Usually they said ‘wir sein Esterreichyn’, *we are Austrians*, because we belong to Austria. All inhabitants of the Monarchy were perceived here as Austrians, and it did not depend on the language used by them” (quotation from Filip 2005: 162, after Chromik 2016: 96). The linguistic/dialectal facet of Austrianess was not a stable and independent marker either, as the dialect continuum, referred to as Silesian German (*Schlesisch, Schläsisch/ Schlässch, Deutschschlesisch*), covered the territory administered by various polities in modern times. The *Bielitz-Bialaer Sprachinsel* and the enclave of Wilamowice constituted the easternmost confines of that Silesian-German dialectal continuum, where intense language contact with the Polish language (continuum) left a remarkable imprint. The German(ic)-Polish language contact was much stronger in Wilamowice than, e.g., in Hałcnów/Alzen/Alza, not to mention much less “Polonised” varieties spoken by the much more nationally German-oriented urban population of the nearby Bielitz/Bielsko. The religious factors were also not without significance; worth reckoning is the prevailing Catholic profile of Wilamowice and Hałcnów, while Bielsko and Biała also had considerable shares of Protestant and Jewish population.

Subsequently, Wymysiöeryś has not been considered by its speakers’ community just a language variety of German, as standard German was regarded as a separate language, and as such was hardly present in their everyday lives, with exceptions of the Austrian administration or Nazi occupation times during World War II. Some of the testimonies might have referred to the German substrate of Wymysiöeryś; its “mixed” nature was, however, apparent to external observers,

---

<sup>10</sup> Nevertheless, Kuhn’s articles on Wilamowice may still be considered an inexhaustible source of ethnographic and folkloristic information, as may works by several other Nazi-engaged German fieldworkers (e.g. Alfred Karasek or Hertha Strzygowski) – cf. Wicherkiewicz & Olko (2016) and Chromik (2019).

as, e.g., the Moravian-German minister Karl Franz Augustin, who commented in his 1842 *Jahrbuch oder Zusammenstellung geschichtlicher Thatsachen, welche die Gegend con Oswieczyn und Saybusch angehen*:

“The commune of Wilamowice has a peculiarity. The German language has remained here until the present day. From the immemorial times, despite the fact that all round Polish was spoken in the town, there are remains of the old Vandals and Burgunds, who saved their German language, which naturally is now mixed with Polish” (quoted after Chromik 2016: 96).

What appeared *proto*-Wandal or Burgund to Augustin sounded “like Yiddish or English” to Jan Łepkowski (1853), professor of the University in Cracow, who reported:

“A peculiar vernacular! (...) In spite of centuries-old language contacts between Wilamowice and Hałcnów from one side and the Polish and Silesian Slavdom from the other, the former have totally preserved their Gothic asperity; this is a Germanic dialect, caught and ossified in its medieval form. Some people consider the settlers Dutch, who had arrived here during the earliest religious unrests (...). Apart from the settlers themselves, nobody understands that language. They speak Polish or German to strangers; they pray in Polish; nevertheless, they have extraordinarily preserved their embalmed dialect for ages.”

Another intriguing message concerning the (linguistic) origin of Wilamowiceans was expressed in a collection of folk (or rather folklorised) poems from the region by an amateur ethnographer from Biala, Dr. Jacob Bukowski. In his 1860 volume, the poem *A Welmeßajer ai Berlin* tells the story of a merchant from Wilamowice in Berlin, who tries to sell woven fabrics, shouting:

“Buy (...) beautiful coutils from Wilamowice!”; “he spent a week there” (...) // but he was hardly understood by the foreigners // as he was thought to be from England. // Nothing strange: as, indeed, that is where // the Wilamowiceans came from.”

It is this “embalmed dialect”, “mixed with Polish” and exotic “from-England”, that started functioning as the main factor of the ethnolinguistic identity of Wilamowice in the era of growing nationalism brought to Central Europe in the second half of the 19<sup>th</sup> century.

According to Chromik (2016: 96):

“(...) in the nineteenth century and at the beginning of the twentieth century, Wilamowice drew attention of nationalistically oriented scholars from the nearby big and rich town of Bielsko (Bieltz). They saw the inhabitants of Wilamowice as *Supergermans*, who, although “devoid of any contact with «Germanness», retained their archaic language in the period of German national weakness, [when] the whole German foreground at the edge of the Carpathian Mountains started to yield to Polonisation.” (quotations from Filip 2005) However, if we listen to the voices of Wilamowiceans of the same period, we would see a portrait of a phenomenon that is very seldom nowadays, a prenationalist community. Wilamowiceans, with their distinct language and culture, were fully aware of their separateness.”

The previously mentioned amateur historian and community leader, Józef Latosiński, was probably the most influential and remembered propagator of an “embedded” national character of Wilamowiceans – according to his worldview, they could maintain and cherish their German(ic) language, enclave and ancestry, only if overtly and massively “superstructured” by Polishness at the national level. The overt national plan by Latosiński was preceded by a covert artistic form by the father of literary Wymysiöeryś, Fliöra-Fliöra.<sup>11</sup>

Florian Biesik (1850-1926), the first and most famous Wymysiöeryś-language writer and patron of the modern Wymysiöeryś cultural and linguistic identity, ceaselessly asserted non-German origins of the first Wilamowiceans. The major part of the foreword to his most important work, *Uf jer welt* ('In the other world', 1924), described the odyssey of the predecessors of today's Wilamowiceans. It also depicted the folk linguistic myths promoted by Biesik in order to replace the German plots in the local ethnolinguistic history by new nation-like links and soft language ideologies, connecting Wilamowice(an) with Friesland, Low Countries, or even with the Anglo-Saxons:

“(...) with that language they conquered England and then came back to the mainland, to Friesland and to Wilamowice. Therefore, this language, being the oldest one, is the father of other living Germanic languages of today, i.e., English, Dutch, German, Swedish, and although it served for the transition among them, it has been left isolated (...) The Wilamowicean language

---

<sup>11</sup>Flora-Flora (or Fliöra-Fliöra, according to the current standardised orthography of Wymysiöeryś) the nickname and later penname of Florian Biesik. His literary output and impact have already been documented and analyzed quite thoroughly, by e.g. Wicherkiewicz (2003), Wicherkiewicz & Olko (2016) or Wicherkiewicz (2019). Polish versions (and a selection of English translations) of F. Biesik's texts can be found in Wicherkiewicz (2003) and Wicherkiewicz (2019).

has not had much luck; as the oldest one, it served as the basis for other languages, like earlier Latin, which had served as the basis for the Romance peoples (...)"

Florian Biesik saw himself as a carrier of language ideology, whose mission was both to instruct his compatriots on the "correct" version of their history, and to strengthen the history by means of a Wilamowicean literary microlanguage, following the steps of the *founding fathers* of Europe's major languages:

"(...) Everything in this world depends on chance; and so the speech of the Tuscan peasants accidentally found its great poet Dante, who (...) became the father of the Italian language, or as Luther, with his translation of the Bible (...) became the father of modern German. (...) and so it came to my mind to find all those dear friends (...) in the other world, and at least there, to talk to them in the language of my childhood, i.e. the thousand-year-old Anglo-Frisian-Saxon (or Gothic?) language (...)"

The structure of native bilingualism in the community did and was to exclude German, as Biesik stated when referring to his own family members: "For half a century (...) I have been reading my mother's letters in Polish (around 50, and none in German, as if it had been implied by my brother Dr. Mojmir)".

Consequently, the ethn(ograph)ic and linguistic archaicity of German Wilamowice was to be counterbalanced by their exotic *Occidentalism* (understood as a "West-looking" counterpart of E. Said's *Orientalism*) stressed by the supporters of the Polish national ideology, such as F. Biesik or J. Latosiński. Through the mythical *Occidentalisation*, they would have eagerly seen Wilamowiceans as *Gente Germanici, natione Poloni* ("Germanic by origin, Polish by nationality"), regarding ethnic constructs enforced during the long-lasting cultural and political colonisation of its Eastern territories by the Polish state, where Slavic Ruthenians were encouraged to accept the covertly assimilating status of *Gente Ruthenus, natione Polonus*, while the Balto-Slavic Lithuanians – that of *Gente Lithuanus, natione Polonus*. The embedded ethnic-national identification is still lacking a proper corresponding label as far as (folk) linguistic classification is concerned. Therefore, a mixed Germanic-Polish *variation* would be, and was, the most desirable categorisation from the Polish perspective, while the German linguistics and, primarily, language ideology required the paradigm of *Sprachinsel(forschung)*.

Some linguists, both academic and amateur, however, tried to avoid an ideological context in their studies of Wymysiöeryś. In that respect, one of the most intriguing actors on the stage of Wilamowice's ethnolinguistic history was the

above-mentioned Dr. Mojmir. Hermann Biesik (1874-1919),<sup>12</sup> F. Biesik's much younger brother, a devoted physician and philologist, eventually changed his family name to a more Slavic-sounding *Mojmir*, as a result of a fraternal conflict in the Biesik family, and/or the protest against Austria's Germanisation policy. Hermann Mojmir's most monumental work of enduring value (also in linguistics, ethnolinguistics, and revitalisation of Wymysiöeryś) became his two-volume *Wörterbuch der deutschen Mundart von Wilamowice* (1930–1936). That straightforward classification of the titular *German dialect of Wilamowice* was probably suggested by (either of) the academic supervisors and coeditors: Prof. Adam Kleczkowski (Vol. I) or Heinrich Anders (Vol. II). The former, professor of German linguistics at the University of Poznań and Jagiellonian University in Cracow, published a two-volume grammatical reference handbook of Wilamowicean – with the title (Kleczkowski 1920 and 1921) referring neutrally to the *dialect of Wilamowice in western Galicia*. The former, H. Anders (otherwise a researcher of the medieval German varieties in Poznań/Posen region, and the first editor of selected poem texts by F. Biesik), referred to Mojmir's dictionary as:

“compiled by a native-speaker and completely edited by A. Kleczkowski. (...) Despite the eventful history, the inhabitants of Wilamowice and the surrounding villages have preserved their Silesian dialect with an East-Central German character, very little influenced by the High German written language and by Austrian German, a dialect containing numerous words and word-forms from Old Polish and from Polish.”

Thanks to the works by H. Mojmir, A. Kleczkowski, and H. Anders, Wymysiöeryś became quite famous and one of the best-documented *Sprachinsel*-microlanguages in the world of Germanic linguistics. In 1958, Uriel Weinreich presented his study on “the differential impact of Slavic upon Yiddish and *Colonial German* in Europe”, where Wilamowice and Wymysiöeryś served as one of the main case studies of the latter, and were presented from a perspective of contact linguistics and *bilingual dialectology*. Weinreich tried to draw also some “folkloristic and ethnographic parallels” between the Yiddish-speaking communities and German *Sprachinseln* in Central-Eastern Europe. Nevertheless, Weinreich (1958) also criticised the *Sprachinselforschung* because of its ideological background, too:

“Germanistic research, even at its best, has been preoccupied with the origins and chronology of German eastward migration and with the patterns of dialect mixture and leveling, and it has therefore sought out the archaic,

---

<sup>12</sup>H. (Biesik) Mojmir's most complete biography is available in Polish by Król (2020).

## 9 Variation, myths, and ideology in Wymysiöeryś ethnolinguistic identity

Germanic elements of language and culture, while new acquisitions, seemingly irrelevant to migration history, were deemphasised or overlooked (...)

The physical destruction or dislocation of the societies under discussion [including Wilamowice] rules out the possibility of fresh field work and forces the student to rely to a considerable degree on materials published before the Second World War. For the problem at hand, this turns out to be a major handicap for the ideological framework,”

Weinreich's 1958 work, however, seems to remain unknown to the community of Wilamowice and their researchers, most probably because of the lack of academic communication between the West and East and/or the official *désinteressement* in any German(ic) heritage in post-War Poland (cf. Wicherkiewicz 1993).

For the long years of prevailing communist-nationalist ideologies, between the end of World War II and the political changes of the 1980s, both Wymysiöeryś and its research must have fallen into oblivion by the Polish scholarship and academia. Except for occasional studies (as Weinreich 1958), the only perspective Wilamowice was researched from was that of the German linguists and ethnographers (as Walter Kuhn), who acted mainly as members and representatives of the *Landsmannschaften der Heimatvertriebenen* (associations of expellees). They actually continued research on “the Supergermans” from Wilamowice/Wilmesau, *Bielitz-Bialer Sprachinsel*, and other *altschlesische Sprachinseln*.

The post-War history of Wilamowice community was then studied either from German exile perspective, or subject to the official communist-nationalist propaganda in Poland. Only recently, several projects have addressed the problems of acute language abandonment, community trauma, and the resulting sociolinguistic developments, including the language micro- and macroideologies at stake (e.g. Chromik 2019).

The critical developments that influenced (also mythically and ideologically) the contemporary history of Wymysiöeryś were: signing of the *Volksliste* by Wilamowiceans during the Nazi occupation, massive appropriation of their properties by the neighbors based on an old-established economic grudge, direct ethnic cleansings in the former *Bielitz-Bialer Sprachinsel*, or the official ban issued on Easter 1945:

“(...) from now on, we ban any use of the local dialect – also in family and private situations, the forgoing concerns also wearing the distinct folk costumes. Those who do not comply with the present ban will be brought to

severe punishment; since it is the high time to put stop to any distinctness and its lamentable results" (Filip 2005: 183–184).

A steady and systematic language loss eradicated Wilamowicean from most, and eventually from all, its traditional domains, from family circles to community life. *Wymysiöeryś* lost its communicative and integrative functions, with a few exceptions including individual curricula, as, e.g., grandchildren brought up solely by their grandparents, or displaced families (cf. Wicherkiewicz 1993 and Wicherkiewicz & Olko 2016).

Any return to any noticeable token of Wilamowicean identity became possible in the late 1970s, when a Belgian television station made a documentary on and in Wilamowice *Een dorp van Vlamingen?* ['A village of Flemings?', 1977].<sup>13</sup> It does constitute a priceless record of the Wilamowicean language and culture of the stage; the movie has also become a constitutive factor for the common perception of Wilamowice as a forgotten and remote colony established by settlers from Flanders (or Holland). The fame of a Flemish village attracted also researchers, such as Hugo Ryckeboer (of Ghent) and Norbert Morciniec (of Wrocław), who published two extensively documented articles (Ryckeboer 1984, Morciniec 1984), discussing possible origins of the *Wymysiöeryś* language; the latter evolved his debate as Morciniec (1995).

The turn of the 1980s marked essential changes in Poland's social, political, and economic system. The initially top-down changes were to significantly influence the position, perception and role of minority communities in the new democratic order. The country and society were to restructure their self-view of a monolingual, monoethnic, and homogenous state-nation, and the community of Wilamowice was to start an entirely new period in their history: reclamation of the regional "microethnic" identity and revitalisation of the "microlanguage", not as a (part of any) *Sprachinsel*, but as a newly re-established speech community *in statu renascendi* (cf. Wicherkiewicz et al. 2018).

This article is not discussing any later, recent, and current developments in Wilamowice, as the actual language ideologies that are currently at stake do not focus anymore on language myths or endo- and exo-ethnotheories of origin. The new generations, who have been assuming their roles of leaders and players in the current movements (civic, local, regional, linguistic, language documentation, institutionalisation and legal recognition, etc.), are forming new paradigms. Research of *Wymysiöeryś* as a language system (e.g. Andrason 2014, 2015a,b, Źak 2016) and as a sociolinguistic community (Król et al. 2016, Neels 2016, Wicherkiewicz & Olko 2016) continues on an unprecedented scale.

---

<sup>13</sup> Available: <https://youtu.be/ibUn82Odjpo> (1 February, 2021).

This sketch, thus, may serve as actually a starting point for further research on the correlates between language variation and language ideology from the perspectives of historical sociolinguistics or perceptual dialectology.

## References

- Andrason, Alexander. 2014. The Polish component in the Vilamovicean language. *Glossos* 12. [https://slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/media\\_items\\_files/andrason.original.pdf](https://slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/media_items_files/andrason.original.pdf). (25 June, 2022).
- Andrason, Alexander. 2015a. Slavic-Germanic hybridisation in the Vilamovicean language. In Elżbieta Mańczak-Wohlfeld & Barbara Podolak (eds.), *Words and dictionaries: A festschrift for Professor Stanisław Stachowski on the occasion of his 85th birthday*, 29–40. Kraków: Jagiellonian University Press.
- Andrason, Alexander. 2015b. Vilamovicean: A Germanic-Slavic mixed language? *Studies in Polish Linguistics* 10(2). 57–85.
- Balowska, Grażyna. 1999. Mikrojęzyki literackie. *Socjolingwistyka* 16. 41–49.
- Biesik, Florian. 1924. Uf jer welt. In Tomasz Wicherkiewicz (ed.), *The making of a language: The case of the idiom of Wilamowice, Southern Poland*, 53–265. Berlin: De Gruyter Mouton.
- Bukowski, Jacob. 1860. *Gedichte in der Mundart der deutschen schlesisch-galizischen Gränzbewohner, resp. Bielitz-Biala*. Bielitz: Zamarski.
- Chromik, Bartłomiej. 2016. From Wilamowice to the philosophy of science and back again: Three levels of linguistic ideologies. In Justyna Olko, Tomasz Wicherkiewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 91–109. Warszawa: Uniwersytet Warszawski.
- Chromik, Bartłomiej. 2019. *Mikro-i makroideologie językowe: przykład języka wilamowskiego*. Unpublished. University of Warsaw Faculty of Artes Liberales. (Doctoral dissertation).
- Dulichenko, Aleksandr Dmitrievič. 2006. Sovremennoe slavyanskoe yazykoznanie i slavyanskie literaturnye mikroyazyki. In Aleksandr Dmitrievič Dulichenko & Sven Gustavsson (eds.), *Slavyanskie literaturnye mikroyazyki i yazykovye kontakty: materialy mezhdunarodnoy konferentsii, organizovannoy v ramkakh komissii po jazykovym kontaktam pri mezhdunarodnom komitete slavistov*, vol. 7, 22–46. Tartu: Slavica Tartuensia.
- Filip, Elżbieta Teresa. 2005. Flamandowie z Wilamowic? (Stan badań). *Bielsko-Bialskie Studia Muzealne* 4. 146–198.
- Kleczkowski, Adam. 1920. *Dialekt Wilamowic w zachodniej Galicji: Fonetyka i fлексja*. Kraków: Polska Akademia Umiejętności.

- Kleczkowski, Adam. 1921. *Dialekt Wilamowic w zachodniej Galicji: Składnia i szyk wyrazów*. Poznań: Uniwersytet Poznański.
- Kloss, Heinz. 1967. 'Abstand languages' and 'Ausbau languages'. *Anthropological Linguistics* 9(7). 29–41.
- Król, Tymoteusz et al. 2016. Czym jest dla dzisiejszych wilamowian język wilamowski? Różne funkcje, różne postawy językowe. *Lódzkie Studia Etnograficzne* 55(1). 241–264.
- Król, Tymoteusz. 2020. Hermann Aleksander Mojmir (Biesik). In Katarzyna Ceklarz & Jan Święch (eds.), *Etnografowie i ludoznawcy polscy: sylwetki, szkice biograficzne*, vol. 5, 154–157. Wrocław: Polskie Towarzystwo Ludoznawcze.
- Kuhn, Walter. 1934. *Deutsche Sprachinselkorschung: Geschichte, Aufgaben, Verfahren*. Vol. 2. Plauen im Vogtland: Wolff.
- Latosiński, Józef. 1909. *Monografia Miasteczka Wilamowic: Na podstawie źródeł autentycznych*. Kraków: Drukarnia Literacka.
- Łepkowski, Józef. 1853. Notatki z podróży archeologicznej. *Gazeta Warszawska* 329. 3–4.
- Mattheier, Klaus J. 1994. Theorie der Sprachinsel. Voraussetzungen und Strukturierungen. In Nina Berend & Klaus J. Mattheier (eds.), *Sprachinselkorschung. Eine Gedenkschrift für Hugo Jedig*, 333–348. Frankfurt am Main: Peter Lang.
- Mattheier, Klaus J. 1996. Methoden der Sprachinselkorschung. In Hans Goebel, Peter H. Nelde, Zdeněk Starý & Wolfgang Wölck (eds.), *Kontaktlinguistik/Contact linguistics/Linguistique de contact. Ein internationales Handbuch zeitgenössischer Forschung. An international handbook of contemporary research. Manuel international des recherches contemporaines*, vol. 1, 812–819. Berlin: De Gruyter Mouton.
- Métrak, Maciej. 2019. Wymysorys (Vilamovicean) and Halcnovian: Historical and present-day sociolinguistic situation of microlanguages in a Southern-Polish language island. In Lidija Rezoničnik & Nina Zavašnik (eds.), *Sloveni od preloma 19. in 20. stoletja do danes: Jezikovne, zgodovinsko-politične spremembe in književni doprinos*, 7–19. Ljubljana: Zvesva društva Slavistično društvo Slovenije.
- Mojmir, Herman. 1930–1936. *Wörterbuch der deutschen Mundart von Wilamowice*. Kraków: Polska Akademia Umiejętności.
- Morciniec, Norbert. 1984. Die flämische Ostkolonisation und der Dialekt von Wilamowice in Südpolen. *Slavica Gandensia* 11. 7–18.
- Morciniec, Norbert. 1995. Zur Stellung des deutschen Dialekts von Wilmesau/Wilamowice in Südpolen. In Gundolf Keil & Josef J. Menzel (eds.), *Anfänge und Entwicklung der deutschen Sprache im mittelalterlichen Schlesien*, 71–81. Sigmaringen: Jan Thorbecke.

- Neels, Rinaldo. 2016. Language and identity in Wilamowice: A complex history of language choices and language attitudes. In Justyna Olko, Tomasz Wicherkiewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 111–130. Warszawa: Uniwersytet Warszawski.
- Protze, Helmut. 1995. Zur Erforschung deutscher Sprachinseln in Südost- und Osteuropa. Siedlung, Sprache, Geschichte und Wechselwirkungen. In Gerhard Grimm & Krista Zach (eds.), *Die Deutschen in Ostmittel- und Südosteuropa*, 55–84. Munich: Verlag Südostdeutsches Kulturwerk.
- Riehl, Claudia Maria. 2010. Discontinuous language spaces (*sprachinseln*). In Peter Auer & Jürgen E. Schmidt (eds.), *Language and space: Theories and methods: An international handbook of linguistic variation*, vol. 1, 332–354. Berlin: De Gruyter Mouton.
- Ritchie, Carlo J. W. 2012. *Some considerations on the origins of Wymysorys*. [https://www.academia.edu/8290384/Some\\_Considerations\\_on\\_the\\_Origins\\_of\\_Wymysorys](https://www.academia.edu/8290384/Some_Considerations_on_the_Origins_of_Wymysorys). Sydney: University of Sydney. (BA Thesis). (1 January, 2021).
- Ryckeboer, Hugo. 1984. Die “Flamen” von Wilamowice. Versuch zur Deutung einer bäuerlichen Überlieferung. *Slavica Gandensia* 11. 19–34.
- Stern, Dieter. 2018. Languages without an army: Minority, regional and literary microlanguages. In Dieter Stern, Motoki Nomachi & Bojan Belić (eds.), *Linguistic regionalism in Eastern Europe and beyond: Minority, regional and literary microlanguages*, 14–38. Frakfurt am Main: Peter Lang.
- Wagner, Richard Ernst. 1935. *Der Beeler Psalter. Die Bielitz-Bialaer deutsche mundartliche Dichtung*. Katowice: Verlag der Kattowitzer Buchdruckerei.
- Weinreich, Uriel. 1958. Yiddish and colonial German in Eastern Europe: The differential impact of Slavic. In *American contributions to the Fourth International Congress of Slavists, Moscow, September 1958*, 369–421. The Hague: Mouton.
- Wicherkiewicz, Tomasz. 1993. The impact of politics and social factors on the death of a minority language: The case of Wilamowicean in Poland. In Peter W. Thomas & Jane Mathias (eds.), *Developing minority languages: The proceedings of the Fifth International Conference on Minority Languages (July 1993, Cardiff, Wales)*. Cardiff University Department of Welsh.
- Wicherkiewicz, Tomasz. 2003. *The making of a language: The case of the idiom of Wilamowice, Southern Poland*. Berlin: De Gruyter Mouton.
- Wicherkiewicz, Tomasz. 2019. Mikroliteratura mikrojęzyka mikrospołeczności. Wilamowice, Fliöra-Fliöra i ich literatura. *Litteraria Copernicana* 2(30). 113–124.
- Wicherkiewicz, Tomasz, Tymoteusz Król & Justyna Olko. 2018. Awakening the language and speakers’ community of Wymysiöeryś. In vol. 26, 179–191. Cambridge University Press.

- Wicherkiewicz, Tomasz & Justyna Olko. 2016. Researching, documenting and reviving Wymysioeryś: A historical outline. In Justyna Olko, Tomasz Wicherkiewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 17–53. Warszawa: Uniwersytet Warszawski.
- Wiesinger, Peter. 1983. Deutsche Dialektgebiete außerhalb des deutschen Sprachgebiets: Mittel-, Südost-, und Osteuropa (mit einem Anhang von Heinz Kloss). In Werner Besch, Ulrich Knoop, Wolfgang Putschke & Herbert E. Wiegand (eds.), *Dialektologie: Ein Handbuch zur deutschen und allgemeinen Dialektforschung*, vol. 2, 900–929. Berlin: De Gruyter Mouton.
- Żak, Andrzej. 2016. The influence of the Polish language upon Wymysorys in light of the revitalization process. In Justyna Olko, Tomasz Wicherkiewicz & Robert Borges (eds.), *Integral strategies for language revitalization*, 131–149. Warszawa: Uniwersytet Warszawski.

# Chapter 10

## Evaluating linguistic variation in light of sparse data in the case of Sorbian

Eduard Werner

The severely endangered Sorbian languages (ISO hsb, dsb), endemic to the Eastern part of Germany, are dramatically under-researched. This lack of research extends from basic knowledge about numbers of speakers, competence, and language transmission, but includes also core aspects of linguistics, like phonology, morphology and syntax.

Having experienced centuries of marginalisation, Sorbian texts are (sparsely) attested only from the 16<sup>th</sup> century. This makes evaluation of variation especially difficult, since the variation might be caused by a certain register of the language, for example, a special dialect (our default assumption), but it might also be caused by other factors such as traditions of verbal art, notably in folksongs which unfortunately have not been preserved in their original form either and are therefore hard to evaluate, but which contain very old layers of language.

In this chapter, one of the oldest Sorbian monuments will be compared to folksongs, applying knowledge about neighbouring Germanic and Celtic literatures. From the linguistic side, the results of the comparison lend greater insights into historical sound changes in Sorbian; from the cultural side, we learn about aesthetic concerns of verbal art in this language, which, in turn shed light on a range of linguistic phenomena beyond sound patterns.

### 1 Goal of this paper

The goal of the investigation is to get a more reliable reconstruction for older layers of Sorbian (Upper Sorbian, ISO hsb and Lower Sorbian, ISO dsb) on the one hand and more concrete ideas about Sorbian verbal art on the other hand. As



a starting point, elements of verbal art from other European cultures (restricting ourselves to examples from Old High German and Welsh in this paper in order to demonstrate such elements) will be applied to Sorbian folk song texts and monuments. It will be shown that these elements make a lot of sense especially if applied to a reconstructed text with a phonologically older layer.

## 2 Introduction

The Sorbs, a Slavic people indigenous to the Eastern part of Saxony and Brandenburg (see Figure 1), are one of the four acknowledged autochthonous minorities of Germany, the others being the Frisians, the Sinti/Roma and the Danes.

They are first mentioned in the year 631 of Fredegar's chronicle from the early Middle Ages and they are the last remnants of the Slavic-speaking population which once reached the Baltic sea in the North and Frankfurt/ Main and Hamburg in the West (Stone 2015: 13). During the 8<sup>th</sup> century a stable border between German (the area without red dots) and Slavic (the area dominated by red dots) population emerged as can be seen from the following map (Figure 2), which shows the distribution of place names ending in -itz<sup>1</sup>. However, German slowly expanded, and after the assimilation of the Polabians (in the region of Lüchow-Danneberg at end of the 18<sup>th</sup> century) the Sorbians were the only autochthonic Slavs left.<sup>2</sup>

For Sorbian, the first monuments maintained appear at the beginning at the 16<sup>th</sup> century, roughly 900 years after their first mentioning in the chronicle of Fredegar.<sup>3</sup> Moreover, the first monuments are usually translations of Christian clerical texts for missionary purposes so they do not reflect the normal context of language usage of that time.<sup>4</sup> For most of the originally Slavic territory which has been germanised long ago, no information of the language and culture other than place names has survived.

The Slavic gentry was germanised early on and Christianity extinguished or significantly diminished old domains of Sorbian language and culture like pagan

---

<sup>1</sup>These place names trace back to Slavic patronymic names ending in \*-ici with only rare exceptions like *Urmitz* in Rheinland-Pfalz which come from Latin.

<sup>2</sup>For a comprehensive study see Stone 2015.

<sup>3</sup>An exception is the Kayna stamp presumably from the beginning of the 10<sup>th</sup> century which is an Old Polabian monument (and the only one) found on Old Sorbian territory (Werner 2004). It shows that the Slavic languages were used in official contexts as well. Unfortunately, no other monument from this time seems to have been preserved.

<sup>4</sup>The oldest Sorbian sentence from 1510, however is a declaration of love (Werner 2014).



Figure 1: Area of settlement of the Sorbs in Germany

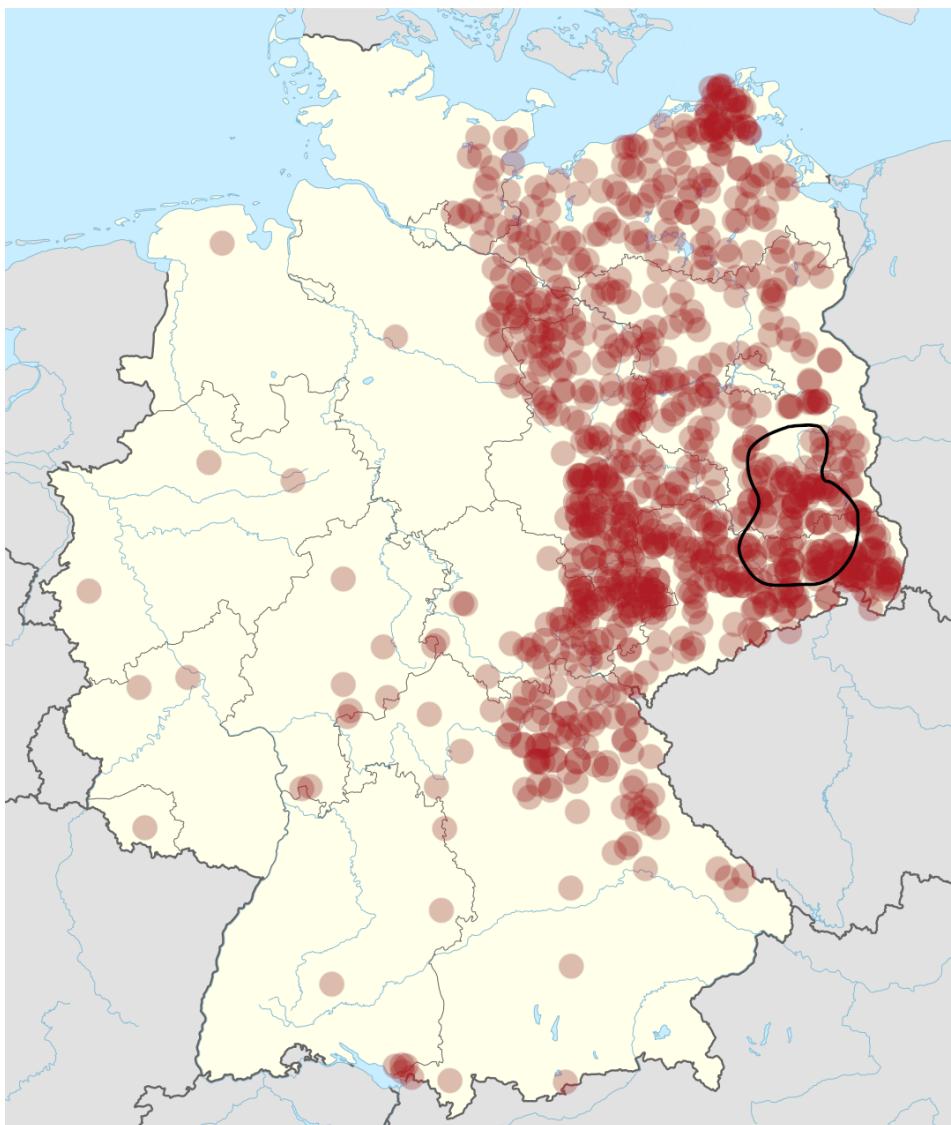


Figure 2: Names ending in -itz. The region outlined in black is roughly the contemporary region of the Sorbian languages. Source: <http://deutschlandkarten.nationalatlas.de/wp-content/namensatlas/>

religion and connected fields such as sorcery and medicine. Apart from the monuments mentioned, only songs have survived. However, they have also changed due to the lack of professional bards; songs got passed on by peasant workers coming from other villages for seasonal work. The oldest Sorbian songs we know have been collected during the 19<sup>th</sup> century. Details regarding they were performed, the context in which they were performed, and even the melodies are sketchy (Nedo 1966: 199) for technological and methodological reasons<sup>5</sup>. Moreover, Smoleń & Haupt (1843) were no folk musicians. They often relied on songs that had been collected by others such as Jordan or Zejler. Altogether we have 1,500 Sorbian song lyrics documented (Nedo 1966: 176).

Nowadays, the Sorbian languages are highly endangered; for Lower Sorbian the number of native speakers is down to a few hundred (Walde 2004, Lewaszkiewicz 2014). For Upper Sorbian, there is still a territory where Sorbian is being passed on as a family language, but their number is declining fast as well and might be no more than 5,000 (*ibid.*). On the one hand, the Sorbian languages display strong German influence in every subsystem of the language, on the other hand we find archaic phenomena like a complete dual<sup>6</sup> or supine<sup>7</sup>.

### 3 Verbal art from Proto-Indo-European to Slavic

We are inclined to view the oral tradition of a culture as an early stage of progress, succeeded by the more stable and permanent stage of recorded language. (Berleant 1973: 340)

PIE seems to have had two distinct types of metrics, a syllable-based one with a quantitative rhythm and a fixed number of syllables as well as a so-called strophic style with relatively short lines and no syllable count. According to Fortson (2010: 35), this style is especially characteristic of archaic liturgical and legal texts.

<sup>5</sup>It was obviously not possible to record performances and intervals and rhythm were written through the filter of musical notation for classical European sheet music.

<sup>6</sup>The dual does not only occur with nouns and adjectives, but pronouns, verbs etc. as well, and it does not need a trigger (like *two* or *both*) to be used, e.g. ‘those (two) children are playing’ would be USo *wonej džěći sej hrajkatej* PRON.NOM.DU N.NOM.DU PRON.DAT V\_2/3duas opposed to ‘those children are playing’ *wone džěći sej hrajkaja*. PRON.NOM.PL N.NOM.PL PRON.DAT V.3PL (cf. Faßke & Michalk 1980: 429ff, fn. 29).

<sup>7</sup>In Lower Sorbian, the supine is a form of the infinitive required when expressing a movement involved in order to act, e.g. *I go to sleep* vs. *I want to sleep*. The first sentence would be expressed with a supine (LSo *du spat* [V.1SG V\_SUP]), the second with an infinitive (*cu spaš* [V.1SG V.INF]). Cf. Janaš (1976: 354)

The *syllabic* type is commonly known from antique epics like the Vedic Rig-veda (presumably from the second millennium BCE<sup>8</sup>), the Sanskrit Mahābhārata (ca. 400 BCE<sup>9</sup>), the Old Greek Illiad and Odyssey (presumably 8<sup>th</sup> century BCE<sup>10</sup>) as well as all the Latin classics. The system is based on vowel length and consonant clusters. Vowels can feature *natural* or *positional* length whereby *positional length* basically means a short vowel and a consonant cluster.

In this type, end rhymes, alliterations and so on can occur. However end-rhymes do not play a central role (Coulson 2017: 17f, 211f).

The strophic type in Hittite, Avestan, Umbrian, Classical Armenian and Old Irish shows that both forms have co-existed with the same geographical scope, since Umbrian coexisted with Latin, and so did Avestian and Sanscrit. This type is characterised by grammatical and phonetic parallelism (Fortson 2010: 35).

As has been argued that the strophic style might be the older one (Fortson 2010: 35), but this discussion is outside the scope of this article. Let it suffice to agree that both types were present at a late PIE period. As it seems, it was perfectly possible to combine both types and end up with a syllable-counting system containing alliterations, as can be seen from the example of a South Picene epitaph given by Fortson (2010: 301), the inscription Sp TE 2, found in Bellante near Teramo:

- (1) postin viam        videtas tetis        tokam        alies  
along road-ACC.SG see-2.PL Titus-GEN.SG toga-ACC.SG Alius-GEN.SG  
esmen vepses vepeten  
this-LOC buried grave-LOC.SG

“Along the road you see the toga of Titus Alius buried in this grave”

While much of the interpretation is still open to guesswork, the artistic part is much clearer: We have three alliterating phrases consisting of seven syllables each which can again be divided into two two-syllable units and a final trisyllable. In each of the phrases we have alliterations: *viam – videtas*, *tetis – tokam*, *vepses – vepeten*.

Repetition of sounds (including alliteration, assonance, and, less frequently, end-rhyme) is characteristic of IE poetry even outside the strophic style. A line like the following, from the Roman comic playwright Plautus (*Miles Gloriosus* 603), is quite typical of the technique:

---

<sup>8</sup>Cf. Fortson (2010: 208).

<sup>9</sup>No exact date can be given, but the compilation must have been undertaken after Pāṇini’s work (Fortson 2010: 208).

<sup>10</sup>Fortson (2010: 249).

- (2) *sī minus cum cūrā aut cautēlā locus loquendī lēctus est*  
 “If your place of conference is chosen with insufficient care or caution  
 ... (trans. P. Nixon)

We have the alliterating *k* sounds (spelled *c*) of *cum cura aut cautela* followed by *l*'s in *locus loquendi lectus*, all of which also have *k* sounds in their interior [...]. In Plautus, the repetition of these sounds is partly for comic effect [...] (Fortson 2010: 37)

Of course, the sound effects can vary and all sound patterns as well as the rhythmical patterns can also serve to provide a mnemonic aid and, therefore, ensure that the text is passed on unchanged. That would be especially important in liturgical and legal texts of all sorts as well as folk medicine, sorcery, etc.

The sound changes which led to the rise of Proto-Slavic<sup>11</sup> (PSI) had a very significant impact on both of these systems. First, the PIE opposition of long and short vowels was abandoned. As a result there was no phonological vowel length (although the PSI yers<sup>12</sup> must have been much shorter than the other vowels, cf. Schaarschmidt 1997: 48ff). Secondly, the principle of open syllables greatly reduced the possibility of having consonant clusters.<sup>13</sup> This would have impacted both the syllabic system and the strophic system as well because many consonants would disappear reducing the possibilities of consonant alliterations. So while it is still possible to count syllables, the aesthetic system of versification must have undergone significant changes in PSI times.

## 4 Verbal art and Sorbian

In Old Sorbian, closed syllables were possible again after the fall of the weak yers.<sup>14</sup> This reduced the number of syllables which would have been the only preserved feature of verbal art from PIE times then. Other changes also occurred affecting consonants or eliminating certain consonant groups. (Due to the lack of written documents it is not possible to date these changes absolutely.)

The following facts are noteworthy:

---

<sup>11</sup>According to Fortson (2010: 419) this occurred before the 5<sup>th</sup> century.

<sup>12</sup>Yers were ultrashort vowels reflecting mostly PIE \*u and \*i; they got lost in many positions later in the individual Slavic languages.

<sup>13</sup>For an extensive introduction of the sound changes of PSI see the introductory chapters of Leskien (1969) and Trunte (1990).

<sup>14</sup>Since the yers were vowels, the loss of them in certain positions (e.g. at the end of the word) led to loss of syllables and to closed syllables preceding the syllable containing the yer, e.g. PSI \*onъ > USo wón, cf. Schaarschmidt (1997: 57f)

- While there is a well-known bardic tradition in both South and East Slavic cultures (as well as Celtic and Germanic), no such traditions are mentioned for Sorbian. This likely owes to the fact that the Sorbian gentry (as well as other Slavic tribes of the Slavia Germanica) was germanised early on. Accordingly, Sorbian bards would not find a society with good conditions to support or sustain Sorbian bardic traditions.<sup>15</sup> This would mean that Sorbian lacked professional bards who could ensure that songs are being passed on properly<sup>16</sup> and that the hidden meanings and metaphors will be taught to the people.
- Traditional South Slavonic and Eastern Slavonic poems are dekasyllabic. Rawp (1957) showed that not even the oldest surviving Sorbian songs are dekasyllabic. So we do not find a cultural connection between Sorbian and East/South Slavonic bardic tradition here.
- Virtually all Sorbian cultural traditions were living and being passed on orally in the so-called *přaza/pšěza*<sup>17</sup> (until well after WWII) by normal village people who were mostly unable to read or write their Sorbian native tongue.
- Songs were passed on by ordinary people coming over from other villages to work. Usually people were keen to be taught new songs and Sorbians travelling to other Sorbian villages could be sure of being asked for all the songs they knew.

Sorbian Studies have so far focused on written texts (e.g. Jenč 1956) depicting Jurij Mjeń (1727–1785) as the founder of profane Sorbian literature<sup>18</sup> with his translation of a part of Klopstock's Messias and his *Ryćerski kěrliš* in clean hexametres (Stone 2012), Handrij Zejler as the father of Sorbian poetry and songs and so on.

---

<sup>15</sup>In the 13th century, there was a known Slavic ministrel Wizlav III from the Isle of Rügen (Cf. Rawp 1978: 12f). While the melodies are said to contain Slavic elements (*ibid.*), the words of all 17 songs which have been preserved are in German, cf. [https://archive.thulb.uni-jena.de/collections/receive/HisBest\\_cbu\\_00008218](https://archive.thulb.uni-jena.de/collections/receive/HisBest_cbu_00008218) (accessed 18-09-2019).

<sup>16</sup>Nedo (1966: 197) remarks that in many cases the lyrics of the songs have been corrupted.

<sup>17</sup>While the translation would simply be *spinning room*, it was more an institution of village life where not only work would be done (like typically spinning and shearing of feathers), but stories were told, songs were sung, people of all generations met and traditions were passed on.

<sup>18</sup>Cf. Jenč (1956: 130), and Čermák & Maiello (2011: 76).

Pawol Nedo (1908–1984) declares Sorbian folksongs<sup>19</sup> as primitive, non-elaborated, imperfect, lacking rhyme, neglecting the fact that in German (which he tacitly keeps in mind) end-rhymes only became predominant in the High MA in the courts of Europe:

Das wichtigste Kennzeichen der gebundenen Volkssprache ist der Rhythmus, mit dem das Volk [...] recht großzügig verfuhr, entweder, weil es ihm keine entscheidende Bedeutung beimaß oder weil es eine klare rhythmische Durcharbeitung nicht bewältigte. (Nedo 1966: 197)<sup>20</sup>

Nedo does notice alliterations, but does not investigate them, and calls them a substitute for the missing or imperfect rhyme<sup>21</sup>:

Die sorbischen Lieder kennen auch keinen bewußt und systematisch angewandten Endreim. Wenn der Endreim erscheint, so hat man oft den Eindruck der Zufälligkeit, und er ähnelt oft mehr einer Assonanz. (Nedo 1966: 197)<sup>22</sup>

Nedo does not give examples for rhymes he considers to be imperfect, so it can only be guessed what he had in mind. He never considers the possibility that traditional Sorbian songs and music might have had its own artistic criteria so that a rhyme he perceives as imperfect or accidental might have been completely fine for the artistically educated Sorb of that time because perception of artistic means is influenced by education to a large degree.<sup>23</sup> For music, something similar has been attested: Christian missionaries described Sorbian singing as cacophonic (Rawp 1978: 11).<sup>24</sup> Thus, it would have been logical and even compelling to consider such a possibility for poetry as well.

---

<sup>19</sup>The songs at issue are the traditional songs found mostly in Smoleń & Haupt (1843), not the romantic songs composed by Kocor and Zejler during the 19<sup>th</sup> century.

<sup>20</sup>“The principal feature of folk poetry is rhythm, which has been treated rather sloppily by the ordinary people, either because it did not seem relevant to them or because they were unable to cope with it properly.” [EW]

<sup>21</sup>“[...] ein Stilmittel [...], das offenbar den fehlenden oder unsauberer Reim ersetzen soll” (Nedo 1966: 199) The term *imperfection* is also being used by Kayser (1954: 98) if an assonance is used at the end of a verse.

<sup>22</sup>“Sorbian songs do not even know a systematical and intentional end-rhyme. Where an end-rhyme occurs, it seems to be accidental and it looks more like an assonance.” [EW]

<sup>23</sup>The Welsh verses we will discuss later in this paper are such an example: Welsh end-rhymes usually consist of a stressed and an unstressed verse while e.g. in German poetry both normally display the same stress pattern (Kayser 1954: 40f). Of course, there are other types of end-rhyme, notably *proest* which requires that the vowel of the rhyming syllables be of different quality (Llwyd 2007: 149ff).

<sup>24</sup>Slaviūnas (1958: 22) attests that in traditional Lithuanian songs parallel seconds are being perceived as aesthetically pleasing.

After the crushing verdict of one of the most outstanding and well-known protagonists of Sorbian culture and Sorbian public life (Nedo had been e.g. head of the Domowina, the chief organisation of the Sorbs, for many years), Sorbian folksongs have not been researched further.<sup>25</sup>

On general grounds, Nedo's view which implies that Sorbian folk songs lack artistic elaboration or structures should be rejected. But his view on his own culture shows that such structures are either very much different from what he expected or hoped to find or veiled by language change (or both). Such different aesthetic systems can be found in Germanic *Stabreimdichtung* or Welsh *cyng-hanedd* (as seen in the following chapter). In both cases, verses are structured by means of alliterations.

## 5 Verbal art in Old German poetry

While Old Germanic poetry is usually dominated by alliteration rhymes (Jan-kuhn & Hoops 2005: 435ff), this is not true for Old German poetry. In German literature these structures have mostly disappeared:

Die Menge überliefelter Stabreimdichtung ist in den einzelnen germanischen Sprachen recht verschieden. Im Ahd. sind es nicht mehr als 200 Zeilen. Hier scheint die Tradition im 9. Jh abgerissen zu sein [...]” (Von See 1967: 1f)<sup>26</sup>

Here is a well-known example from Hildebrand's song (von Eckhart 1729: 864):

- (3) hiltbraht enti haðubrant, untar heriun tuem  
“Hildebrand and Hadubrand, between hosts two”

This is the so-called *long verse* according to Snorri Sturluson (1179–1241) which is supposed to be divided into two parts, the (underlined) *staff* occurs twice in the

<sup>25</sup>A few years before, Jan Rawp (Raupp) had written on Sorbian songs: “Die formale Seite der [...] Liedtexte erweist sich in manchem als eigenartig und reizvoll. Sprachliche Gestaltungsmerkmale wie Epitheta, Alliterationen, Interjektionen u. a. vertiefen Ausdruck und Sinngebung.” [Formally, the lyrics of the songs are in many ways strange and appealing. Linguistic means such as epitheta, alliterations, interjections etc. emphasise expression and meaning.] (Raupp 1966: 10). Rawp was not only a scientist, but also a musician and composer, so he had a different approach than Nedo, but unfortunately he was unable to continue his work due to health reasons.

<sup>26</sup>The amount of traditional alliterational poetry varies according to the language at issue. In Old High German there are not more than 200 verses. Here, the tradition was discontinued during the 9<sup>th</sup> century.

first and once in the second part (where it is supposed to be in the only stressed word). It should be pointed out that this means that there had been at least about 200 years of contact and cultural exchange between Sorbs and Germans at that time.

## 6 Verbal art in Celtic poetry (Welsh)

There are no Celtic monuments documenting verbal art from the region which is now Germany and from the Celtic cultures that must have been in contact with the Slavic and German cultures. Therefore, some Welsh verses will be taken as an example for Celtic since Welsh versification is very well documented (Morris-Jones 1930, Llwyd 2007). The following verses from the 15<sup>th</sup> century by Dafydd ap Edmwnd will illustrate the intricate sound patterns as well as how easy it is to miss them if one is not familiar with them:<sup>27</sup>

- (4) Oeri y bûm ar y barth er cyn cof, a'r ci'n cyfarth.  
“Freezing I was on the ground longer than memory, and the dog was  
barking.”

People mainly familiar with modern Sorbian, Lithuanian or German poetry might only spot the end-rhyme and some individual alliterations (*bûm – barth*, *ci – cyfarth*), but only very few will be aware of the full complexity of the verse if they have never been introduced to the system of *cynghanedd*:<sup>28</sup>

- (5) Oeri y bûm ar y barth  
features a repeating sound pattern *r-b* repeating in the first and the second part of the verse
- (6) er cyn cof, a'r ci'n cyfarth.  
shows an even more intricate pattern *r-c-n-c-f*.

---

<sup>27</sup>The following analysis is by no means complete. It mainly serves as an example of an intricate system of alliteration which is still alive and well documented.

<sup>28</sup>For a full description of *cynghanedd* as well as the *cynghaneddion* given here, see Morris-Jones (1930) and Llwyd (2007).

As mentioned above, there is no surviving poetry from the continental Celtic tribes that were originally in contact with the German and Slavic tribes, but it can be safely assumed that they will also have possessed a system of sound alliterations for verbal art, because, as aforementioned, these alliteration systems can be traced back to PIE times. For the same reason, it can be assumed that the Slavic tribes that were in contact with Germanic and Celtic tribes must have had an at least remotely similar system due to both heritage and ongoing cultural contact.

## 7 Verbal art in Sorbian folk songs

Nedo dedicates only three pages to the language of the Sorbian folk songs (Nedo 1966: 196–199) and describes figures of speech only superficially. He notices the existence of repetitions, epitheta and assonances without investigating them further. Figure 3 provides a sample verse from an Upper Sorbian folk song from Smolef & Haupt (1843: 86), but comprehensive research is necessary.

The theme of the song is the widespread (in folk songs throughout Europe) *knight takes girl* which might have some grounds in Slavic exogamy, but interesting is the lexeme *šelma* since the word originally means ‘carriion, rotting carcass’ and later ‘villain’ and ‘executioner’ (Kluge 2001: 798).

- (7) a. Přišoł je šelma mi šelmowski,  
V\_ł.SG.M COP.3SG N.NOM.SG.M PRON.DAT.1SG A.NOM.SG.M  
Come has villain me villainous  
b. Přišoł je šelma a wzał je ju  
V\_ł-SG-M COP.3SG N.NOM.SG.M CON V\_ł-SG-M COP.3SG PRON.ACC.SG.F  
Come has [a] villain and taken has her  
preč,  
ADV  
away  
c. Hišće mje njeje na kwas prosyl.  
ADV PRON.ACC.1SG COP.3SG.NEG PREP N.SG.ACC.M V\_ł.SG.M  
even me hasn't to wedding asked.  
“Came a villainous villain,  
Came a villain and took her away  
Did not even invite me to the wedding.”

Following Nedo's statements there is a typical repetition (*přišoł je šelma*) and a tautological adjective follwing the noun (*šelma ... šelmowski*) “for greater poet-

**II.**

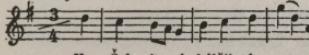
<p><b>Zjěbany luby.</b> (Z Kočiny.)</p> <p><i>Tempo di menuetto.</i></p>  <p>Ha Šol je tón hólčik tón runy pu-éik, ha na-dejšol, na-dejšol runy du-bik, ha na-dej-šol, nadejšol runy du- Der Bursch ist ge-gangen den graden Weg, ge-kommen zum gra-den Ei - chenbaum, ge-kommen zum gra-den Ei - cher</p> <p>Šol je tón hólčik tón runy puéik, Nadejšol, nadejšol runy dubik. .,: Wurjezał je sej tam hałžičku, Wudželal je sebi piščalku. .,: Tak sej wón weselje zapiska, Ze swojej' zrudneje wutroby, Ze swojmaj plačtomaj wočomaj : Ja mam tam lubku haj susodžinku, Ja mam tam lubku mi rjanu holčku. .,: Druzy tam khodža k nej z wesełosću, Ja pak, mój bczo, ze zrudnosću. .,: Přišol je šelma mi šelmowski, Přišol je šelma a wzäl je ju preč, Hišće me neje na kwas prosyl. Džensa je kwas ha za tydžení zas, Néndu ja na tón, da na tamón du, Da na tamón tola zawsjeći du.</p>	<p><b>Der betrogene Liebhaber.</b> (Aus Rotten.)</p> <p>1 Der Bursch ist gegangen den graden Weg, Gekommen zum graden Eichenbaum. .,: 3 Geschnitten hat er sich ein Ästelein, Gemacht hat daraus er ein Pfeifelein. .,: 5 So spielt' er ein fröhliches Lied sich auf, Aus seinem Herzen, das trauerte schwer, Aus seinen Augen, die weinten so sehr: 8 Ich hab' dort ein Liebchen, ein Nachbarskind, Ich hab' dort ein Liebchen, ein Mädchen so schö- 10 Die Andern gehn zu ihr mit Freudigkeit, Ich aber, mein Gott, nur mit Traurigkeit. .,: 12 Ein Schelm ist gekommen, ein schelmischer Set, Ein Schelm ist gekommen und nahm sie mir we- 15 Und heute ist Hochzeit und künftige Woch', Und geh' ich nicht heute, geh' später ich doch, Gewiss und wahrhaftig geh' später ich doch.</p>
--	---

Figure 3: Text fragment from Upper Sorbian Folk Song

ical expressiveness” (Nedo 1966: 198). However, if an older version of the text<sup>29</sup> is assumed, a different picture emerges. Therefore we assume a state after the falling of the weak yers and denasalisation of nasal vowels, but before assibilation of \*í, vowel changes (caused by palatalisations and labialisations) and the establishing of word-initial stress in prefixes:<sup>30</sup>

- (8) Prišel je šelma mi šelmowski,  
Prišel je šelma a wzäl je ju proč,  
Hišće mje njeje na kwas prosyl.

<sup>29</sup>This has so far only been tried for a Sorbian song by Rawp (1957) in order to establish the original number of syllables of the song verses, but not in order to identify other linguistic means of verbal art.

<sup>30</sup>The following (reconstructed) verses are otherwise identical to the ones already annotated. Discussing the individual sound changes in Sorbian is beyond the scope of this article, and we must point the reader to Schaarschmidt (1997).

When looking at the first two verses in example 9, note the complex alliteration rhyme with a staff *šel/zäl*. The *w* belongs phonetically to the preceding syllable:

- (9) Prišel je šelma mi šelmowski,  
Prišel je šelma a wzäl je ju proč,

The first two verses start with the same pattern *pr*, which is repeated in the last stressed syllables of the last two verses creating a frame:

- (10) Prišel je šelma mi šelmowski,  
Prišel je šelma a wzäl je ju proč,  
Hišće mje njeje na kwas prosyl.

The vowel scheme of the first four syllables is identical in all three verses which suggests that the song was originally sung in a canon-like way<sup>31</sup>:

- (11) Prišel je šelma mi šelmowski,  
Prišel je šelma a wzäl je ju proč,  
Hišće mje njeje na kwas prosyl.

Sound structures such as the ones outlined here can be found in many songs. The function of connecting verses can also be taken over by a cyngahanedd-like assonance (they also have the same stress and are accented) which can be found e.g. in the first verse of *Rubieżnicy* (Smoleń & Haupt 1843: 29):

- (12) Jědlenki su rubali a rěbliki su džěłali  
N.ACC-pl COP.3PL V\_ł.PL CONJ N.ACC.PL COP.3PL V\_ł.PL  
pine-trees they-have chopped and ladders they-have wrought

Here a structure *r-b-l – r-b-l* occurs between the verses (*rubali – rěbliki*). Around these, there is twice *ki-su*, adding another element of symmetry as well as an onomatopoëtic feature (the chopping of the axes). Furthermore, the *d-l* of *jědlenki* finds an equivalent in *dž-l* of *džěłali* with palatalisations inverted. Finally, there are also inner rhymes *ki-li-ki-li* (all these syllables are accentuated in the song).

As can be seen, there seems to have been a very rich, dense, and intricate alliterational system which needs further investigation in order to try and reconstruct individual elements of this system.

<sup>31</sup>Slaviūnas states that many of the Lithuanian threefold *sutartinės* were sung in as strict canons (Slaviūnas 1958: 14) and that assonances are a necessary part of them (op. cit.:18).

## 8 Verbal art in Sorbian monuments

In this part, the results of the analysis of the folk song will be applied to one of the oldest Lower Sorbian monuments, i.e. Richter's baptizing agenda from 1543. It sports several unique features, but because of the sparseness of documents from this period, some of them are hard to evaluate (see Figure 4):

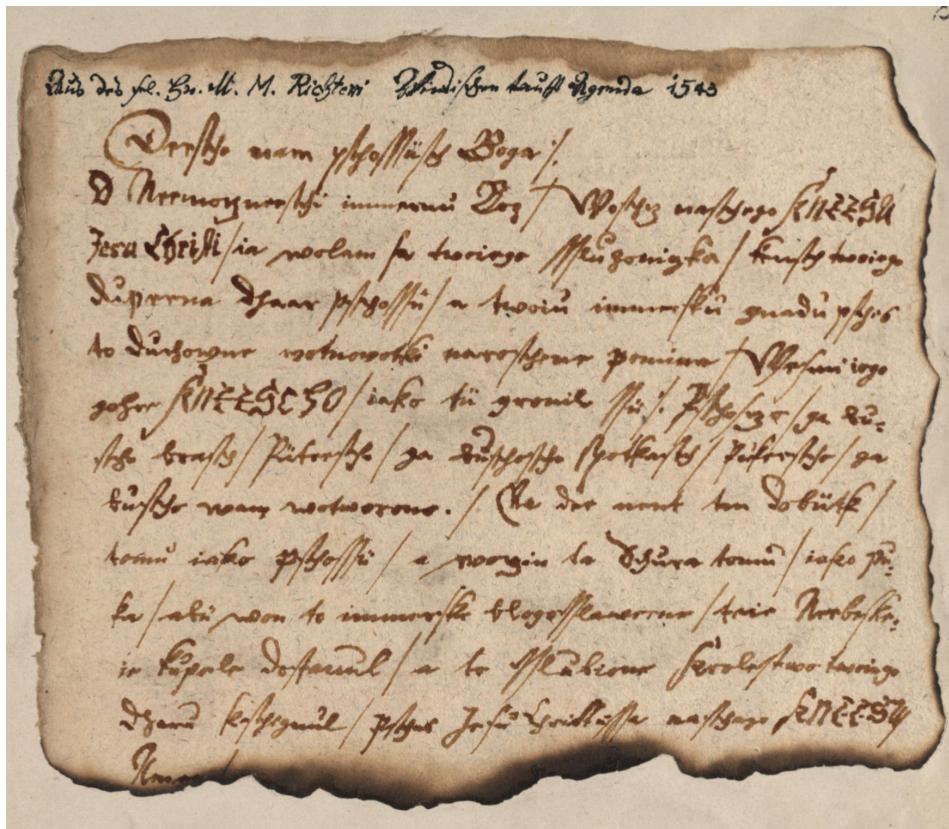


Figure 4: Lower Sorbian baptising agenda from 1543 (Source: <https://sachsen.digital/werkansicht/dlf/172704/3/0/#>)

- The monument shows a change \*aj > ej throughout (\*dajšo 2pl IMP 'to give' > dejšo, \*pytajšo 2pl IMP 'to look for' > pytejšo, \*pukajšo 2pl IMP 'to make burst' > pukejšo<sup>32</sup>), which is phonetically very plausible, but not known from other monuments.

<sup>32</sup>Since orthographical issues are not being discussed here, all examples are given in a modernised orthography.

- There is *kšignuš* (or maybe *kšygnuš*) instead of the expected \**krygnuš* ‘to get’ (loanword from German *kriegen*). Schuster-Šewc (1983: 690) interprets this *kšignuš* as a hypercorrect form, but to all that is known it might as well be a dialectal variation. Probably the sound change at issue had not occurred long ago (people were still aware of it in the late 18<sup>th</sup> century, cf. Schlegel 2019: 31).

Even more striking is the term *blogoslowjenje* ‘blessing’ which is obviously a Church Slavonic (ChSl) term (*blagoslavljenje*) where we would strongly expect something like *žognowanje* (from German *segnen*) because that is the only attested term for ‘blessing’ in any other Sorbian monuments and because there are many other German loanwords in this monument as well. The phonetic adaptation of the word, however, makes it very plausible that it is part of a very old (and at that time long severed) ChSl connection and not merely an ad-hoc-loanword introduced by an educated writer.

The most interesting part, however, is the rendering of Mt 7,7:

Ask, and it shall be given you; seek, and ye shall find; knock, and it shall be opened unto you.

From the phonological system of the monument and other Sorbian (or Polish or Czech) sources, we would expect the following translation:

- (13) Pšosćo, ga bužo wam dano; pytejšo, ga  
V.2PL.IMP CONJ COP.3SG.FUT PRON.D.2PL PART.NOM.SG.N; V.2PL.IMP CONJ  
bužošo namakaš; klapejšo, ga bužo wam  
COP.2PL.FUT V.INF V.2PL.IMP CONJ COP.3SG.FUT PRON.DAT.2PL  
wotworjono.  
PART.N.SG.N

However, the words in the monument read<sup>33</sup>:

- (14) Pšosćo, ga bužo braś;  
V.2PL.IMP CONJ COP.3SG.FUT V.INF  
pytejšo, ga bužošo spotkaš; pukejšo, ga bužo  
V.2PL.IMP CONJ COP.2PL.FUT V.INF V.2PL.IMP CONJ COP.3SG.FUT  
wam wotworjono.  
PRON.DAT.2PL PART.N.SG.N

---

<sup>33</sup>Lines 6-8 of the manuscript, rendered in contemporary orthography for convenience’s sake.

According to any other source of Lower Sorbian, this sentence should be translated as:

Ask, and he will take; seek, and ye shall stumble; make it burst, and it shall be opened for you.

Schuster does not discuss this passage at all (Schuster-Šewc 1967: 293), but simply states in the Sorbian etymological dictionary (Schuster-Šewc 1983) that *spotkaš* also means ‘to find’, and that *pukaš* also means ‘to knock at a door’ (as in Polish *pukać*) although this document is the only source for these meanings. But even then, the passage remains unclear (*ask, and he will take*). In his edition of Sorbian language monuments (1967) Schuster therefore tacitly conjectures *pšosćo, ga bužo braš* to *pšosćo, ga bužošo bras*, which means *ask, and you will take* (Schuster-Šewc 1967: 293). But in spite of the conjecture, the translation is still not an acceptable translation of Mt 7,7.<sup>34</sup> Keeping in mind that the document does not contain any obvious errors, I would be unwilling to accept the conjecture, especially since it fails to provide a full explanation of the deviations from other documents.

One interpretation is that this passage is an old parody<sup>35</sup> which is found in the manuscripts of Thietmar von Merseburg, where the Greek *κύριε ἐλέησον* had been turned into a sentence meaning ‘there is an alder-tree at the bush’ (Stone 2015: 27). The first part (*ask and he shall take*) could then refer to taxes and duties. But it is also possible that the changes were introduced solely in order to produce an aesthetically more pleasing text. Sorbian culture at that time was an oral culture, not a written one, and Christian contents were in any case unintelligible to the Sorbian peasants. Therefore, while a priest could not excel by conveying content, he could still earn the respect of his parish demonstrating oratory skills. Accordingly, the sentence in example 15 will be considered from an artistic standpoint, starting with the expected (reconstructed) wording (differences between this reconstruction and the monument are underlined):<sup>36</sup>

<sup>34</sup>It should be mentioned that there is a translation in the oldest Upper Sorbian catechism of Warichius from 1595 (Schuster-Šewc 2001: 126) which is in line with Schuster’s conjecture (not with the manuscript discussed here) which could be from a different part of the Bible (John 16:24) and which we will not discuss here as it is significantly more recent and from a different region.

<sup>35</sup>I would like to thank Patrick McCafferty from the UL for pointing out that such parodies exist in Irish. In a Slavic context, one could compare the oldest Western Slavonic (presumably Old Sorbian) sentence ukriwolsa.

<sup>36</sup>Newer phonology can be applied here. Only \*r is reintroduced instead of š in *pšosćo* so the reader without knowledge in historical Sorbian phonology can follow more easily. Speakers were still aware of the change \*r > š at the end of the 18<sup>th</sup> century (Schlegel 2019: 31), and the aforementioned example of *kšygnuš* vs. *krygnuš* shows that this sound change was still active for the time of the monument examined here.

- (15) Prosćo, ga bužo Wam dano.  
Pytejšo, ga bužošo namakaš.  
Klapejšo, ga bužo Wam wotworjono.

Observe how all the verses consist of two parts. There is an alliteration p-p in the first two lines, while the second part of all three verses start with *ga bužo*. There is also a sort of climax in the second parts of the verses in so far as there are six syllables in the first verse (*ga bužo Wam dano*), seven in the second (*ga bužošo namakaš*) and eight in the last verse (*ga bužo Wam wotworjono*). This might have been perceived as an aesthetically pleasing starting point, but from an artistic point of view, it could definitely be improved.

Looking at the first verse, *Wam dano* displays no alliteration or assonances and is not linked to anything. However, substituting *Wam dano* with *braš* does two things:

1. It creates a sound chain *pr – b – br*.
2. It changes the meaning from *it will be given to you* to *(he) will take*, which could be a message along the lines of: *Now it sounds correctly and we can understand it properly: so that is what the Christians really mean – they are not giving, but taking*. This might reflect the experience of the Sorbians with the Christian church as they were forced to attend church service (Knauth 1767: 150).

Assuming a change of the text here from *Wam dano* to *braš* would also explain the fact that we have *bužo* 3.SG and not *bužošo* 2.PL and would render Schuster's tacit conjecture unnecessary.

In the second verse, we have the same artistic problem – *namakaš* is not connected. The substitution of *namakaš* with *spotkaš* again has two effects:

- It creates an alliteration *p-t – p-t* with *pytejšo – spotkaš*.
- It creates a parody *seek and you shall – no, not find, but stumble*. Knauth (1767: 120ff) mentions the Sorbians did not like to go into the cold and dark Christian churches.

And finally, when substituting *klapejšo* with *pukejšo* (which is not so far away semantically), there is a much better alliteration, connecting all three verses with *prosćo – pytejšo – pukejšo* (not unlike *veni – vidi – vici*). Furthermore, the second and the third verse are connected with *pytejšo – spotkaš – pukejšo* through the

voiceless stops *p-t – p-tk – p-k*. The substitution again adds to the parody as well, stating that you will *not* find open doors by knocking on them, but that you have to force them open.

So there is a case to be made that the citation of Mt 7,7 found here is an old parody rather than a translation. The fact that this passage is found in a baptising agenda and therefore in a context where a parody should not appear would require that this parody is much older than the monument and that the parody was not perceived as such by the person who incorporated it into the agenda.<sup>37</sup>

The alternative explanation is much less convincing: It requires the dialect of ZosSEN to have evolved lexically rather differently to everything else we know about Lower Sorbian of that time; it would not explain the artistic sound structures (or explain them as coincidences), but even then, it would require a conjecture and leaves the part *pšosćo, ga bužo[šo] braš* partly unexplained.

## 9 Summary

In spite of the assertions made by Nedo (1966), the analyzed samples of Sorbian traditional folk songs feature interesting artistic means which are, however, significantly different from what we would expect from a German perspective. They seem to be based on various types of rhymes including alliteration, not unlike Germanic *Stabreimdichtung* or Welsh *cynghanedd*. Especially alliterating verse may be the oldest surviving attestation of Sorbian verbal art.

As shown, original rhymes were lost to corruption during the oral transmission of texts or obscured by later Sorbian sound changes. We therefore have to assume that it will only be possible to recover parts of the alliteration schemes. However, with *Stabreimdichtung* disappearing in Old High German as early as the 9<sup>th</sup> century, the Sorbian songs would be the oldest (and only) source of poetry of that kind that survived to this day from the former Slavic and now germanised region. As such patterns can also be found in translations of liturgical texts, they for sure go beyond aesthetic purposes. Indeed, apart from their former functions, they help reconstruct texts. Investigating them in detail would require a multidisciplinary philological project since all areas at issue are sparsely documented and have to be further explored. Historical linguistics needs the input

---

<sup>37</sup>Perhaps the parody was at that time so old that it had faded away, or the priest who originally adopted them (who might not have had fluent Sorbian) had been “taught” these words by the community. Cf. again the oldest Western Slavonic sentence cited in the chronicle of Thietmar von Merseburg, where the people claim that their corrupted version is what they had been taught by Boso, the first bishop of Merseburg. (Stone 2015: 27)

from literature, cultural studies, and musical studies as well. For example, on the one hand, historical sound changes help to unearth elements of verbal art; on the other hand, it is possible to date the historical sound changes more exactly because of their effects on alliterations.

## Abbreviations

V	verb	SG	singular
COP	copula	DU	dual
N	noun	PL	plural
A	adjective	M	masculine
ADV	adverb	F	feminine
CON	conjunction	N	neuter
PRON	pronoun	NOM	nominative
PREP	preposition	GEN	genitive
PART	participle	DAT	dative
ł	ł-form used for most compound tenses and moods	ACC	accusative
IMP	imperative	INF	infinitive
		SUP	supine

## References

- Berleant, Arnold. 1973. The verbal presence: An aesthetics of literary performance. *The Journal of Aesthetics and Art Criticism* 31(3). 339–346. DOI: 10.2307/429164.
- Čermák, Radek & Giuseppe Maiello. 2011. *Nástin dějin a literatury Lužických Srbů*. Brno: Středoevropské centrum slovanských studií.
- Coulson, Michael. 2017. *Complete Sanskrit*. revised by Richard Gombrich and James Benson. London: Hodder Education.
- Faßke, Helmut & Siegfried Michalk. 1980. *Grammatik der Obersorbischen Schriftsprache der Gegenwart: Morphologie*. Bautzen: Akademie der Wissenschaften der DDR.
- Fortson, Benjamin W. 2010. *Indo-European language and culture: An introduction* (Blackwell Textbooks in Linguistics). Chichester: Wiley.
- Janaš, Petr. 1976. *Niedersorbische Grammatik: Für den Gebrauch der Sorbischen Erweiterten Oberschule*. Bautzen: Domowina Verlag.
- Jankuhn, Herbert & Johannes Hoops. 2005. *Reallexikon der germanischen Altertumskunde*. 2nd edn. Vol. 29. Berlin/New York: de Gruyter.

- Jenč, Rudolf. 1956. *Stawizny serbskeho pismowstwa*. Vol. 1. Nakl. Domowiny.
- Kayser, Wolfgang. 1954. *Kleine deutsche Versschule*. Lehnen Verlag München.
- Kluge, Friedrich. 2001. *Etymologisches Wörterbuch der deutschen Sprache*. Vol. 24. bearbeitet von Elmar Seibold. Berlin/New York: De Gruyter Mouton.
- Knauthe, Christian. 1767. *Derer Oberlausitzer Sorberwenden umständliche Kirchengeschichte*. Görlitz.
- Leskien, August. 1969. *Handbuch der altbulgarischen Sprache, Grammatik – Texte – Glossar*. Heidelberg.
- Lewaskiewicz, Tadeusz. 2014. Dolnołużycki i górnolużycki - języki zagrożone czy wymierające. *Slavia Occidentalis* 71(1). 37–53.
- Llwyd, Alan. 2007. *Anghenion y gyngahanedd: Fersiwn newydd*. Llandybïe: Cyhoed-diadau Barddas.
- Morris-Jones, John. 1930. *Cerdd dafod sef celfyddyd barddoniaeth Gymraeg*. 2nd edn. Gwasg Prifysgol.
- Nedo, Pawoł. 1966. *Grundriß der sorbischen Volksdichtung: Schriftenreihe des Instituts für sorbische Volksforschung in Bautzen*. Bautzen: Domowina Verlag.
- Raupp, Jan. 1966. *Sorbische Musik*. Bautzen: Domowina Verlag.
- Rawp, Jan. 1957. *Naše golcy z wojny jědu – staroserbski episki spěw*. Vol. C 2. 35–65. Letopis.
- Rawp, Jan. 1978. *Serbska hudźba*. 2nd edn. Budyšin: Ludowe nakładnistwo Domowina.
- Schaarschmidt, Gunter. 1997. *The historical phonology of the Upper and Lower Sorbian languages*. Heidelberg: Carl Winter.
- Schlegel, Matthias. 2019. Leben und Werk Johann Friedrich Fritzes. In *Festschrift zum 200. Todestag Johann Friedrich Fritzes*, 15–48. Kolkwitz: Evangelische Kirchengemeinde Kolkwitz.
- Schuster-Šewc, Heinz. 1967. *Sorbische Sprachdenkmäler, 16.-18. Jahrhundert* (Schriftenreihe des Instituts für Sorbische Volksforschung in Bautzen 31). Bautzen: Domowina Verlag.
- Schuster-Šewc, Heinz. 1983. *Historisch-etymologisches Wörterbuch der ober- und niedersorbischen Sprache*. 1978–1989. Bautzen: Domowina Verlag.
- Schuster-Šewc, Heinz. 2001. Die ältesten Drucke des Obersorbischen. In Wenceslaus Warichius und Gregorius Martini (ed.). Bautzen.
- Slaviūnas, Zenonas. 1958. *Sutartinės: Daugiabalsės lietuviu liaudies dainos. I[-II.] Sudaré ir paruosé Z. Slaviūnas*. Valstybinė grozinės literaturos leidykla. <http://www.sutartines.info> (20 August, 2019).
- Smoleř, Jan Arnošt & Leopold Haupt. 1843. *Volkslieder der Sorben in der Ober- und Niederlausitz: Pěsnički hornich a delních Lužiskich Serbow*. Fotomechanischer

- Neudruck mit einem Vorwort von Jan Raupp, Domowina Verlag, Bautzen 1984.  
Grimma: VEB Domowina-Verlag.
- Stone, Gerald. 2012. Upper Sorbian reczerski in Jurij Mjeń's poem 'Sserfskeje reczje samoženje a kwalbu we recžerskim kyrlischu' (1767): An amendment. In Irina Podterga (ed.), *Schnittpunkt Slavistik, Ost und West im wissenschaftlichen Dialog. Festgabe für Helmut Keipert zum 70. Geburtstag*, chap. Teil 3: Vom Wort zum Text. Bonn: V&R unipress.
- Stone, Gerald. 2015. *Slav outposts in Central European history: The Wends, Sorbs and Kashubs*. digital edition. London/Oxford/New York/New Delhi/Sydney: Bloomsbury Publishing.
- Trunte, Hartmut. 1990. *Ein praktisches Lehrbuch des Kirchenslavischen in 30 Lektionen. Zugleich eine Einführung in die slavische Philologie*. München: Otto Sagner.
- Von See, Klaus. 1967. *Germanische Verskunst* (Sammlung Metzler. Abteilung E: Poetik 67). Stuttgart: J. B. Metzlersche Verlagsbuchhandlung. DOI: 10.1007/978-3-476-99093-8.
- von Eckhart, Johann Georg. 1729. *Commentarii de rebus Franciae orientalis et epis copatus vvirceburgensis...* Vol. 2. sumptibus almae UniversitatisJuliae.
- Walde, Martin. 2004. Demographisch-statistische Betrachtungen im Oberlausitzer Gemeindeverband „am Klosterwasser“. *Lětopis* 51(1). 3–27.
- Werner, Eduard. 2004. Das Petschaft von Kayna. In Uwe Junghanns (ed.), *Lin guistische Beiträge zur slavistik*, 63–72. (= Specimina Philologiae Slavicae 149). München: Sagner.
- Werner, Edward. 2014. Dodawk k serbskej sadže z lěta 1510. *Lětopis* 61(2). 55–57.

# Chapter 11

## Modelling accommodation and dialect convergence formally: Loss of the infinitival prefix *tau* ‘to’ in Brazilian Pomeranian

Gertjan Postma<sup>a</sup>

<sup>a</sup>Meertens Institute Amsterdam

Various pathways with their respective outcomes of multi-dialect interaction have been described in the literature: levelling in the sense of the erasure of linguistic communal differentiation, interdialect formation with compromise forms or fudging, and reallocation of doubles to distinct functions. In this paper I re-evaluate a well-known, but often ignored mechanism and outcome: revert to default settings, the rise of the unmarked, i.e. whenever the result of the change is not a sum or subset of the input forms, but an innovative pattern. Two related models are developed, one for koineisation and one for accommodation, that can serve as an evaluation scheme for a language change. The case study pursued is the loss of the infinitival prefix *tau* ‘to’ in Pomeranian, a West Germanic language, extinct in Europe, but still spoken in isolated communities in Brazil. While the original Pomeranian dialects in Europe had a considerable variation in this particular domain, Pomeranian in Brazil has converged to a remarkably uniform new construction, which was not present in Pomerania in the days of emigration. I show that underlying structures remain constant in all Pomeranian dialects, European as well as Brazilian Pomeranian, but the spellout pattern in Brazil is the cross-linguistic default.

### 1 Introduction

Dialectology and sociolinguistics do not only have a value in themselves, they also offer a window to the formal aspects of language and may function as a



Gertjan Postma. 2022. Modelling accommodation and dialect convergence formally: Loss of the infinitival prefix *tau* ‘to’ in Brazilian Pomeranian. In Matt Coler & Andrew Nevins (eds.), *Contemporary research in minoritized and diaspora languages of Europe*, 303–339. Berlin: Language Science Press. DOI: 10.5281/zenodo.7446983

method to reveal underlying structures of natural language. Language contact is an especially valuable tool for formal research. In language contact, the result transcends the input variants and where the final state is no obvious function (addition, selection, split, superposition, etc.) of the initial state. In this study I report on dialect convergence of a set of mutually intelligible dialects and its outcome. I discuss a grammatical change in a language island in Brazil: the loss of the infinitival prefix *tau* ‘to’ in Pomeranian, a West Germanic language. I will argue that the dialectology and sociolinguistics of this minority language provide evidence for the T-to-C movement in infinitival constructions, as was argued for in Pesetsky & Torrego (2007). First I provide a brief overview of the various mechanisms of convergence that have been discussed in the literature, as well as other mechanisms of language change, especially convergence and accommodation. Then, as a background, I give a description of the nature of the complementiser and the infinitival prefix in Pomeranian. In Section 3 I discuss a possible source of the change: the original Pomeranian dialects in Europe had considerable variation in this particular domain. The pattern of this variation is investigated as well as the underlying syntactic pattern. I list two mechanisms of resolving this variation: convergence of the various dialects to a new koine and accommodation to Portuguese. I then repeat the arguments from my 2016 study, which show that Portuguese is not the likely source of change. The arguments in my previous publication that lead to the conclusion that accommodation to Portuguese is not likely to have given direction and impetus to the change, but rather dialect-internal convergence within the Pomeranian diasystem, still hold. But these must be balanced by new considerations of occurrence frequency.

## **2 Contact-induced language change**

While traditional diachronic linguistics has focused on language change by inherent processes, such as (phonological) erosion and inherent instabilities of linguistic cycles (e.g. Jespersen’s cycle), modern sociolinguistics has made contact-induced language change a major object of investigation. For instance, the arrival of considerable numbers of immigrants usually changes the dynamics of a community thoroughly and its language with it. Colonisation, e.g. the settlement of various dialect speakers in a foreign country, usually gives rise to new social dynamics, a new society, and a language with new properties. Two extreme cases are noteworthy: the circumstance of huge immigration of mutually unintelligible speakers, outside the immediate realm of a roof language, may initiate a creolisation process: the emergence of a completely new structure, albeit

with words of various source languages (Bickerton 2015). The other end of the spectrum is the circumstance of a linguistically inhomogeneous but mutually intelligible group of immigrants, which by some social factor is isolated from the environment. This creates a so called *language island*, where the various source dialects converge to a new *koine* (Frings 1936, Rosenberg 2005). Finally, there is the more moderate circumstance when an (immigrant) group has moderate contact with the dominant group “outside”, the superstrate. In such interactions two processes can be observed: the influence of the minority language on the dominant language, usually by the switch of immigrants to the dominant language (substrate effect, Van Coetsem’s source-language agentivity), and the influence of the dominant language on the minority language (accommodation, prestige, Van Coetsem’s recipient-language agentivity, van Coetsem 1988). It may be clear that an actual situation never realises one of these processes in pure form. Accommodation goes together with convergence, creolisation is not always clearly separable from convergence.

## 2.1 Accommodation

Accommodation is omnipresent in linguistic interactions. When an American hears a speaker who pronounces /o/ in *socks* lower, i.e. identical to *sacks*, he nevertheless perceives it as /o/ if it is embedded in a broader context (Labov 1994: 68–70). The process is automatic and usually unconscious. This is accommodation *in perception*. Accommodation *in production* is a speaker’s adaptation to a hearer in a specific situation. This can be in lexis when one speaks to young children. It can be changes in phonology if one talks with friends in a bar, etc. It is also possible to accommodate in syntactic structures. When accommodation becomes systematic, and conventionalised, it is a source of language change, for instance if it occurs in a linguistic group in interaction with another linguistic group.

Though “accommodation” is used in the literature in various senses, I will reserve it in this paper to the situation where a group of speakers changes its language in order to become acceptable or intelligible to another group, usually the dominant, more prestigious group, i.e. it is *asymmetric*. It is also possible to accommodate the superstrate language to some minority group, i.e. to a substrate. For instance, if Turkish immigrants in the Netherlands use periphrastic constructions to realise the V2 constraint in Dutch more often, it might be seen as an accommodation strategy to retain the basic SOV structure in accommodation to the more rigid SOV order in their Turkish mother tongue (Van de Craats 2009).

## 2.2 Koineisation

While accommodation is conceptually an asymmetric process – one language accommodates to another – *koineisation* is, at least conceptually, a process whereby language variants influence each other. The process is, conceptually at least, symmetric (Gumperz & Wilson 1971). The mechanism involved is *convergence*. Koineisation may give rise to a *Sprachbund*, but it most typically occurs in *Sprachinseln*, language islands: settlements with colonist of various dialect regions. Such (German) language islands were studied “as relics from the past” (Rosenberg 2005: 222) from the 19<sup>th</sup> century onwards, though the explicit mechanisms of the changes only received attention in the 20<sup>th</sup> century. Rosenberg notices that the language islands are not homogeneous, neither linguistically nor socially. “(...) they were often inhabited by settlers of different origins, i.e. by speakers of different dialects” (Rosenberg 2005: 223). Below I mention four mechanisms by which the process of koineisation can come about: levelling, interdialect formation, reallocation, and revert to the default settings. The first and the last mechanism can be considered simplification (L1-L2 language contact), the other two mechanisms are complexification in the sense of Trudgill (2011): they typically occur with 2L1 language contact (bilingualism).

### 2.2.1 Levelling

Most researchers mention levelling as the major process of new dialect formation in closed immigrant groups. It is the process of eliminating prominent stereotypical features of the input dialects (Dillard 1972). Notice that all stereotypical features and locally specific features are typically the first to be eliminated (Thelander 1980, Hinskens 1996). The process is symmetrical, despite the fact that the result is eliminating a certain feature from one of the two dialects in interaction. In many cases, it leads to reduction of inflectional paradigms and morphology in general.

### 2.2.2 Interdialect formation

Interdialect formation is the rise of compromise forms. In the case of two dialects, this can be by simple optionality of two forms, by neutralisation of the feature that defines the two forms, or by superposing the two forms. Chambers & Trudgill (1998) mention the case of [ʌ] and [ʊ] in *strut* in East Anglia, which merge around the isogloss to [y]. A clear example of the superposing process, mentioned in Hinskens (1996: 366) is the emergence of superheavy syllables on the borderline of Limburgian dialects. The eastern dialect has [x] drop in

*nacht* [naxt] ‘night’ under compensatory lengthening of the vowel ([na:t]). The dialect west of the isogloss has [naxt]. On the borderline, new forms such as [na:xt], i.e. with both [x] and the long vowel can be observed. In the latter case, the superposed form is clearly a transitional phenomenon, under the assumption that superheavy syllables are marked. A more complex syntactic example is given in Postma (2014), where on the borderline of two Limburgian dialects with two types of Verb-second (the German type with uniform C-V2, and the Dutch type with C-V2 and T-V2), complex V-AGR-T forms emerge, such as *klöps-de* ‘knock.2SG-ED’. This can be explained if the interdialect complies with both types of V2: the V complex moving to C skipping T, where AGR is the so-called COMP-inflection. This mechanism clearly works on underlying rules, rather than on surface forms. Once again, these superposed forms are marked and often transitional (cf. Cornips 2006).

### 2.2.3 Reallocation

Just like interdialect formation, reallocation gives rise to complexification. Reallocation takes two or more inputs from source dialects and redistributes these over two or three sub-contexts. As an illustration, in Jundiaí, an Italian immigrant city in Brazil, people use both the Portuguese word *pavor* [pa’vor] ‘fear’ and the Italian word *paura* [pa’ura] ‘fear’ in their *Caipira* version of Portuguese, but limit *paura* for the meaning ‘strong fear’. Britain (1997) and Taeldeman (1989) provide more complex phonological cases where two alternates from source dialects distribute in a contact dialect. The distribution is *rule-governed*. These are, of course, the more interesting cases linguistically, because they potentially shed light on underlying linguistic processes.

### 2.2.4 Revert to the default

The final mechanism that I would like to mention, is revert to the default. If two dialects, one with a marked setting, the other with an unmarked setting in some feature, come into contact, the result tends to lean towards the unmarked setting. For instance, if there are two features involved, say,  $F_1$  and  $F_2$ , and if we call + the marked and  $\emptyset$  the unmarked value, contact of a dialect with  $[+F_1, \emptyset F_2]$  and a dialect with  $[\emptyset F_1, +F_2]$  might give rise to the new variant  $[\emptyset F_1, \emptyset F_2]$ . Dependent on the nature and abstractness of  $F_1$  and  $F_2$ , the contact variant might have a rather different appearance without obvious connection to the properties of the source dialects. In Postma (2004, 2012), I give a case of two variants of late Middle Dutch (MD), that lack a reflexive pronoun, i.e. these dialects circumvent the

Binding Theory, albeit for different (marked) mechanisms. Cross-linguistically, the reflexivity of pronouns is dependent of feature underspecification, typically number, but also person, or case (Reuland & Reinhart 1995), while referential pronouns are (fully) specified. By a marked parameter setting, however, the referential pronoun, MD *hem* ‘him’ had number underspecification in the Southern Dutch dialects (meaning either ‘him’ or ‘them’) and could be used as a reflexive, while it had ACC/OBLIQUE underspecification in the Northern dialects (cf. Hoekstra 1994 for modern Frisian). Both settings are marked settings, cf. Table 1.<sup>1</sup>

Table 1: Feature analysis of a koineisation process in Dutch reflexive constructions.

Variety	Pattern	Feature Setting		Markedness	
		F <sub>1</sub>	F <sub>2</sub>	F1	F2
a. Southern MD	NP <sub>i</sub> ... hem <sub>i</sub>	yes	no	+	0
b. Northern MD	NP <sub>i</sub> ... hem <sub>i</sub>	no	yes	0	+
c. <i>koine</i>	*NP <sub>i</sub> ... hem <sub>i</sub> / NP <sub>i</sub> ... sick <sub>i</sub>	no	no	0	0

F<sub>1</sub> = Number neutralisation in pronouns; F<sub>2</sub> = ACC/OBL neutralisation in pronouns.

What we observe then is that both marked strategies are lost in the contact-induced variant. The contact dialect then comes in need of a special, underspecified, reflexive pronoun. It then *actively* borrows it from neighboring German dialects, first *sick*, later *sich*. It was in need of the borrowed form, rather than accommodating to it. The result with a reflexive is a result of contact between two variants without reflexive. It may be clear that the grammatical system is a creative force which transcends the dialectal input. We might call this tendency towards the default variant in contact “micro-creolisation”. This shows that convergence to the default can not only be the result in cases of a set of unrelated source languages without mutually intelligibility, but also in closely related mutually intelligible dialects.

In the next sections, I present a case of contact of many minimally distinct Pomeranian dialects, which merge in a language island in Brazil. I will investigate if revert to the default is active in this case.

<sup>1</sup>It is slightly more complicated. In the case of oblique, it is the feature inventory that is marked, not the setting.

### 3 European Pomeranian (EP)

#### 3.1 Background

Pomeranian is the dialect (or set of dialects) of Coastal Germanic roughly between the Oder river and the Vistula river, an area which is called Hinterpommern. Until 1945 it was first part of Prussia, later Germany, but lies in present-day Poland. The dialect of Mecklenburg-Vorpommern in present-day Germany is rather different (henceforth Mecklenburgian) and should be discussed separately from Hinterpommersch, henceforth simply Pomeranian. The map in Figure 1 below, slightly adapted from Brockhaus (2012: 128), gives an impression of the Pomeranian area, indicated with “Ostpommersch”.

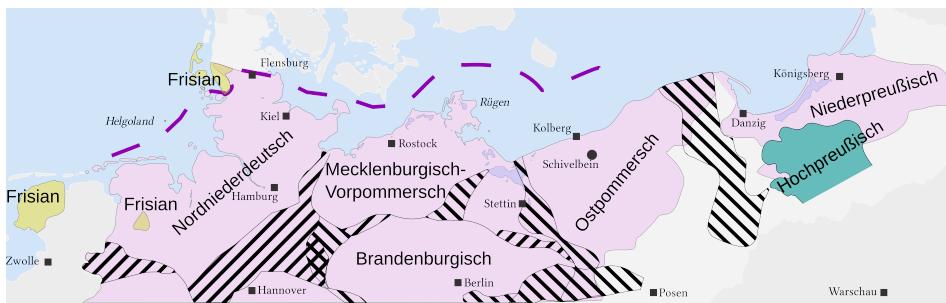


Figure 1: Coastal Germanic dialectal areas in the first decades of the 20<sup>th</sup> century (after Brockhaus 2012).

Pomerania was Germanised in a geographically scattered way during the so-called *Ostsiedlung*, the “going east” of settlers, land developers, and merchants coming from Flanders, Holland, and Frisia and later from the core Saxon areas. The newly emerged variant of Low Saxon, Pomeranian, has been in close contact to High-German and Slavonic, especially Slovincian/Kashubian.<sup>2</sup> The origin from the North Sea area might explain the consistent Ingwaeonisms in the language, characteristics of the North Sea Germanic area, such as loss of /n/ before spirants, development of a -s plural in nouns. The linguistic roof of High German through religion and education explains the many German loans, e.g. in the ordinals (*fünft* instead of the expected *fi:wd* ‘fifth’), in kinship terms (*grosmuter*

<sup>2</sup>Slavonic influence on Pomeranian can be ignored from the 13<sup>th</sup> century onward, except for Slovincian. In the 20<sup>th</sup> century, Slovincians were, like the Pomeranians, predominantly Lutheran, and were expelled with them from the new Polish areas in 1945.

instead of the expected *groutmuder* ‘grandmother’, etc.). Virtually all Pomeranians in Europe were Lutherans.<sup>3</sup>

A distinguishing feature of the Pomeranian vis à vis Mecklenburgian in the west and Low Prusian in the east is the existence of two infinitival forms: an infinitive in *-a* ([ə] or [ɐ]), and one in *-en* ([ən] or [ŋ], Wrede 1895: 295).<sup>4</sup> Two types of infinitives are further encountered in Frisian and North Frisian (Hoekstra 1997: 4–5).<sup>5</sup> In Pomeranian, the infinitive in *-a*, which we call infinitive-1 (INF1), is used in clauses under modals, under causatives (*läta* ‘let’, *daua* ‘do’), verbs of motion (*gåa* ‘go’), and control predicates, as exemplified by the Wenker-sentence<sup>6</sup> 16b in example (1). The example is taken from location 20, the village of Schloenwitz (present-day Słonowice) in the municipality Schivelbein (see map). This schwa-infinitive (INF1)<sup>7</sup> is used without complementiser and without infinitival prefix.

- (1) European Pomeranian (19<sup>th</sup> century (Schloenwitz))  
Du must eista no 'a inn wass-a  
you must first still a bit grow.INF1  
'you must first still grow a bit'

The infinitive in *-en*, which we will call infinitive-2 (INF2), is used in embedded infinitivals with a leading complementiser, as exemplified in the Wenker-sentence 16a in (2), again taken from the village of Schloenwitz.

- (2) European Pomeranian (19<sup>th</sup> century (Schloenwitz))  
Du bust nog ni groot naug um 'n Flasch Wiin ut-tau-drink-en  
you are yet not big enough COMP a bottle wine PRT-to-drink.INF2  
'you are not big enough to drink out a bottle of wine'

In this paper I study the changes in infinitival syntax of such rationale clauses.

<sup>3</sup>Data for the entire Pommern Province in the year 1932: Lutherans (90.7 %), other Protestants (1.3%), Catholics (6.7%), Jews (0.5%). For the region of emigration (see the map in Figure 2), the ratio of Lutherans ranges from 97–98.9%. Cf. GLFP (1932).

<sup>4</sup>Neither Vor-Pommersch (to the west) nor Low Prusian (to the east) participates in this characteristic feature.

<sup>5</sup>Alemannic dialects also have two infinitival forms, one in *-a/e* and one in *-i(n)t* (Bayer & Brandner 2004). The syntactic distribution is rather different from the *-a/e* vs *-en* infinitive in Coastal Germanic. See also Höhle (2006).

<sup>6</sup>The Wenker-sentences are a set of 40 sentences that Georg Wenker used in a questionnaire for dialect research in 1880 in 40,000 locations in Germany. The sentences have also been elicited in The Netherlands, Belgium, Luxemburg, Austria, and Switzerland.

<sup>7</sup>Please see the Abbreviations section.

## 4 Variation in the Infinitival Syntax of European Pomeranian

Rationale clauses in European Pomeranian can be studied using the Wenker sentences,<sup>8</sup> that were elicited around 1880.<sup>9</sup> Using the online database, I checked more than 300 locations in coastal Pomerania i.e. in municipalities Schivelbein, Regenwalde, Belgard, Colberg-Cörlin, Cöslin, Greifenberg, and Schlawe, as the emigration into Espírito Santo was mainly fed from this coastal area (cf. Granzow 2009: 167). The various municipalities are indicated in Figure 2.



Figure 2: Municipalities (Kreise) covered in the search on infinitival constructions.

It turns out that there is some variation in the realisation of this construction in European Pomeranian with respect to the infinitival prefix *tau* ‘to’. Apart from (3a) where, as in Standard German, Dutch and Frisian, both *um* and *tau* are realised, (e.g. *um* and *zu* in German, *om* and *te* in Dutch/Frisian), we observe two alternative patterns in Pomeranian. In one of these, the ‘*to*’-prefix *tau* remains unrealised (3b), and in another variant, *um*, the ‘*for*’ complementiser, remains unrealised (3c).<sup>10</sup>

<sup>8</sup>Cf. Demske (2011). The Margburg digitalisation project, led by Jürg Fleischer, made Wenker sentence 16 available through a grid of 1250 datapoints (of the 40,000 data points).

<sup>9</sup>The Wenker sentences are not available in digital format, but scans of the questionnaires can be inspected at [www.regionalsprache.de](http://www.regionalsprache.de).

<sup>10</sup>These are not necessarily different dialects, as optionality might be involved.

- (3) European Pomeranian (1880, Schloenwitz, Lankow, and Schlenzig, resp.)
- du bust nog nich grot naug üm an Flasch Wiin ut-tau-drinken
  - du bust nog nich grot naug üm an Flasch Wiin ut-∅-drinken
  - du büst no ni groot naug ∅ ain Flasch Winn  
you are yet not big enough COMP a bottle wine  
ut-tau-drinken  
PRT-to-drink.INF2  
'you are not big enough to drink out a bottle of wine'

The fourth conceivable option with both *üm* 'for' and *tau* 'to' unrealised, is not found. I summarise the patterns in Table 2 for the entire coastal area. From now on I will gloss *üm* as 'for' and *tau* as 'to'.

Table 2: Occurrences of infinitive constructions in European Pomeranian

	Pattern	Occurrence	Frequency	N
a.	FOR ... TO	general	83%	258
b.	∅ ... TO	rare	11%	34
c.	FOR ... ∅	rare	6%	20
d.	*∅ ... ∅	absent	0%	0

The complementiser *üm* 'for' can remain empty only if the verbal prefix *tau* 'to' is not empty; conversely, the verbal prefix *tau* can be empty only if the complementiser *um* is not. This is cast in a cross table in Table 3 on the basis of the Wenker sentences of 312 locations in Pomerania.<sup>11</sup>

Table 3: Cross table of occurrences of infinitival constructions in European Pomeranian

	+FOR	-FOR
+TO	258	34
-TO	20	0

This shows a structural absence of the [∅ ... ∅] pattern with *p*-value of 0.09 in Fisher's test. To be more precise: The hypothesis  $H_0$  that the absence of the

<sup>11</sup>The six places where the Wenker sentence 16 has been translated by a finite embedded clause (*du bist noch nicht groß genug daß du eine Flasche Wein austrinken kannst*) were ignored. They occur scattered over the area and it does not seem a structural effect.

$[\emptyset \dots \emptyset]$  pattern is a mere result of the (low) probability ( $\text{FOR}=\emptyset$ ) and the (low) probability ( $\text{TO}=\emptyset$ ), is rejected with a p-value of 0.09.

I therefore conclude that both positions T and C must “see” each other at some level of representation (Bennis & Hoekstra 1984: 55). This suggests that the *tau*-marker in Pomeranian, at least in these rationale clauses, concerns the syntactic type of the infinitival marker as described in Brandner (2006). Following standard assumptions on these markers, I assume that **FOR** (*um, om, üm* ...) sits in C (Koster & May 1982: 133, Vanden Wyngaerd 1987: 108) while **TO** (*zu, te, to, tau*, ...) sits in T (Evers 1990, Sabel 1996).<sup>12,13</sup> Since we are dealing with constructions that have a lexicalised complementiser in continental Germanic, I assume that there is T-to-C movement at some level of representation and that the complementiser C must be lexical at that level. I, therefore, make the assumptions in (4), taken from Hoekstra (1997: 106, 116) developed for (Fering) Frisian. The lexicalisation requirement of C already holds in West Germanic for main and embedded *finite* clauses and (4b) is a natural extension to non-finite sentential constructions.

- (4)    a.  $C \dots T \Rightarrow C^\circ[C+T] \dots T$
- b.  $[C^\circ]$  is overt in all types of clauses in Pomeranian<sup>14</sup>

Notice that T-to-C movement in infinitival constructions is independently motivated from a theoretical perspective, cf. Pesetsky & Torrego (2007), who derive the T-movement chain from basic syntactic principles. In the next section I will provide evidence that these assumptions also hold in Brazilian Pomeranian.

## 5 Brazilian Pomeranian (BP)

### 5.1 Background

While Pomeranian is not used anymore in cohesive communities in Europe since 1945, it is still in full use in various parts of Brazil, with many children not learning Portuguese at all until schooling at age six or so. These communities derive from immigration as early as 1850, and have been rather isolated until recently.

---

<sup>12</sup>Bennis (1987) argues that so-called prepositional adjunct clauses have P in the C position.

<sup>13</sup>Arguments have been raised against treating ZU in German as a functional head (I or T), see e.g. Haider (2010: 273–274). Brandner (2006) argues that one should distinguish morphological ZU from syntactic ZU. If this is correct, dialects with only syntactic ZU cannot be excluded. There is no evidence in Pomeranian that a morphological *tau* should be distinguished. On the contrary, most of the evidence forwarded in Postma (2014) only follows under the assumption of an exclusively syntactic *tau* in Pomeranian.

<sup>14</sup>It would be attractive to extend this to West Germanic infinitivals without *um* in general, as in Bayer (1984). I only defend the claim for Pomeranian here. Kayne (1999) argues that all Romance complementisers are complex: a W head that have attracted the infinitival prefix *di*.

In this article I will use the variant spoken in the state of Espírito Santo, in the municipality of Santa Maria de Jetibá and surroundings.<sup>15</sup> I simply call it Brazilian Pomeranian, though there might be differences with the variants in the South (in the states of Santa Catarina and Rio Grande do Sul) or in the Amazone region (Rondônia), which left the Northern parts of ES in the 1970s. This community is rather big<sup>16</sup>. Virtually all Brazilian Pomeranians are Lutherans (Droogers 2008). Although Pomeranian was never used in the liturgy until quite recently (first in High German, since 1942 in Portuguese), the religion is an important factor of social cohesion that safeguards the language in Brazil (Schaffel Bremenkamp 2014). Within the various groups of Germanic immigrants, the Pomeranians have become the dominant group, both economically, religiously, and sociologically. For instance, virtually all Dutch immigrants that arrived at the same time and who were Calvinists, have converted to Lutheranism and speak Pomeranian now.

Recently, a collection of Brazilian Pomeranian tales was published under the title *Upm Land* (Tressmann 2006b, henceforth *UmL*), as well as a dictionary of Brazilian Pomeranian (Tressmann 2006a). The data used in this paper are mainly from this corpus of tales, provided by a variety of authors and registered by Anivaldo Kuhn and Ismael Tressmann. The orthography that is used is the one developed in Tressmann (2006a). Apart from this corpus<sup>17</sup> I completed my data with two interviews in March 2013 (Elizana Schaffel) and September 2013 (Elizana Schaffel and Tereza Gröner).

## 5.2 The infinitival syntax of Brazilian Pomeranian

As said above, the distinction between the two infinitives has been fully retained in ES.<sup>18</sup> The complementiser in infinitive-2 constructions, however, is never realised as *üm*, but as *taum* ([ta<sup>u</sup>m]/[tam]). Interestingly, the verbal prefix is always null, indicated with Ø. So while the *tau* prefix position is systematically zero, the complementiser position has changed from *üm* to *taum*. I give some examples in (5), rationale clauses, taken from UmL (78, 114, 115).

### (5) Brazilian Pomeranian

- a. Dai lüür sin arm un häwa kair gild [taum sich air huus Ø  
The people are poor and have no money FOR.TO REFL a house Ø

<sup>15</sup>Santa Maria de Jetibá, Caramuru, Garrafão, Melgaço, and Domingo Martins.

<sup>16</sup>Tressmann (1998) estimates the population to be 300,000.

<sup>17</sup>Cf. Postma (2014) for the details.

<sup>18</sup>Under influence of High German (in older speakers) or Hunsrückisch (in some areas), deviations from the Pomeranian pattern occur: overgeneralised *n*-forms and overgeneralised *e*-forms, respectively. These are not present in the corpus used in this study.

buugen].

build.INF2

'The people are poor and have no money to build themselves a house'

- b. Dai blaumasuuger is air seir hübsch tijr. Hai hät aina langa  
The flowersucker is a very elegant animal. He has a long  
snåbel [taum dai saft uuta blauma Ø suugen].  
beak FOR.TO the juice out-the flowers Ø suck.INF2  
'The hummingbird is a very elegant animal. It has a long beak to suck  
the juice out of the flowers'
- c. Dai ima maga seir geirn dai maluulabüsch eer blauma [taum  
the bees like.PL very much the malula bushes their flowers FOR.TO  
sich eera hoinig Ø måken].  
REFL their honey Ø make.INF2  
'The bees like the malula-bushes's flowers very much to make honey'

In contrast to the situation in European Pomeranian, there is virtually no variation left in Brazilian Pomeranian. There is no variation in the complementiser position, which is always *taum*. Only in 3 of the 127 cases (2%) in the corpus does the original *tau* show up, but it is not adjacent to the verb, i.e. we may assume that it has always moved up to the C-position.<sup>19</sup> There is no variability in the

---

<sup>19</sup>For further reference, I give these three cases. Only in one case (i) is *tau* a true complementiser. In the other cases (ii–iii), *tau* assigns a deviant dative case to the embedded object under surface adjacency, similar to English *For me to go....* Apparently, the intervening subject PRO does not block case assignment to the object in BP.

- (i) [Tau-Ø dai köirn afstampen] gewt dat aina stampküül.  
to the.ACC grains crush.INF2 is there a pounder  
'There is a pounder to crush the grains'
- (ii) Suurdaig dörwt ni feigla [tau dem daig anmåken ] < taum de daig  
sourdough may not fail to the.DAT dough produce.INF2 < TO the.ACC dough  
anmåken  
produce  
'Sourdough may not be absent upon making dough'
- (iii) [Tau dem rijs weglegen] mud man em mita slusa forwåra < taum de  
to the.DAT rice store.INF2 must one him with the peel stock.INF1 < TO the.ACC  
rijs weglegen  
rice store.INF2  
'In order to store the rice, one needs to store it with the chaff'.

These are not performance errors, as informants accept both variants. I leave these sentences for further research.

lexicalisation of the lower infinitival prefix position, either: it is without exception without spellout. Hence we may conclude that Brazilian adjunct infinitivals have obligatorily lexicalisation of C and no low spellout of T in these infinitival constructions.

One way of understanding this innovation is to postulate that Brazilian Pomeranian has reanalyzed *taum*, which was originally a P+CASE complex [tau+m], into [tau+um], i.e. as a C+T-complex of *tau* and *um*. It is then an overt realisation of the rule in (4) that I inferred from European Pomeranian dialect set. So, the surface *variability* of lexicalizing C and T in European Pomeranian has been replaced by a surface *rigidity* in Brazilian Pomeranian. The underlying formal rigidity of spelling out the C-T chain in European Pomeranian has been retained and recaptured by an overt marking of the head of the C-T chain.

- (6) a. in  $[C_i \dots T_i \dots]$ , the *chain* must be lexicalised in EP  
b. in  $[C+T]_i \dots \emptyset_i$ , the C+T complex must be lexicalised in BP

The scheme in (6a) shows that the variability in spellout in EP has been replaced by one spellout form, under retention of the more abstract underlying syntax.

The *taum+INF2* construction had a precursor in European Pomeranian, illustrated in (7). It is the nominalised use of the *-en* form, illustrated by Wenker sentence 20, given for Schloenwitz.

- (7) European Pomeranian (19<sup>th</sup> century (Schloenwitz))  
Hai deer so, as hann-e in taum dörsch-en bistellt  
He did so, as if had he him for-the.DAT threshing invited  
'...as if he had invited him for the threshing'

In this construction, *tau* is a preposition enriched with a dative marker (*taum* < *tau+(de)m*). This construction allows modification but it must be done adjectivally, by PPs, or under incorporation: no direct object arguments between *taum* and the nominalised verb are possible, because the deverbal noun cannot assign case.<sup>20</sup> The infinitival construction in *-en* has been reinterpreted in Brazilian Pomeranian as a verbal construction<sup>21</sup> in which the verb in the *-en* infinitive does assign Accusative case, e.g. *air huus* 'a house' in (5a), *dai saft* 'the juice' in

<sup>20</sup>Incorporated objects are possible even when no accusative is available. Incorporated objects do not need Accusative Case cross-linguistically (Baker 1988). The dimension of case assignment is often ignored in the literature (cf. for instance Demske 2011).

<sup>21</sup>Cf. Haspelmath (1989) for the grammaticalisation pathway of infinitives.

(5b) and *eera hoinig* ‘their honey’ in (5c). Since syntactic categories that receive case cannot assign case, cf. the Case Resistance Principle (Stowell 1981) or the Unlike Category Constraint (Hoekstra 1984), the case assigning preposition *tau(m)* was obligatorily reanalyzed into a non case assigning tense head.

The question is now what caused this change, which is minimal with respect to the surface string but with considerable structural consequences. Why does only the complementiser position receive lexicalisation in Brazilian Pomeranian? Is it an accident that the superstrate language Portuguese does not have an infinitival prefix and systematically lexicalises C in this context (*para* ‘for’)?

## 6 Other contact varieties

In the previous section, I showed that the BP verbal *taum* construction is a Brazilian innovation. It does not occur in the Wenker material of the Pomeranian areas in Europe. But I also showed that the C-T link also had deep structural parallels in the dialect continuum of European Pomeranian. Therefore, it does not come as a surprise that we encounter similar constructions in other West Germanic dialects. In this section I review some of these.

### 6.1 Middle English

The oldest West Germanic counterpart to the *taum* construction of Brazilian Pomeranian is found in Middle English. We can compare this construction with the Middle English split infinitive (where the verbal prefix *to* has undergone T-to-C in forming a complex *for-to* complementiser (8), taken from Visser (1963: par. 982); see also Mustanoja (1960)).

- (8) Middle English (Pecock, Repr. 219)
- a. A nurish or a modir is not bounde forto alwey and for euere Ø fede her children.  
‘a nurse or a mother is not bound to always and for ever feed her children.’
  - b. He eoden (...) forto fully that folk and godes lawe techen  
he went (...) FOR.TO fuly that people and god’s law Ø teach.INF  
‘he went in order to fully teach God’s law to that people’

- c. if it schulde plese god forto bi miracle make a fier and a watir  
if it should please god FOR.TO by miracle Ø make a fire and a water  
tgidere  
together  
'if it would please God to combine fire and water'

If we identify Eng. *for* with Pom. *um* and Eng. *to* with *tau*, the parallel is striking. Admittedly, it is not certain that *for* actually resides in C. It might sit in a lower position (van Gelderen 1998). Nevertheless, the processes share the raising of the infinitival prefix away from the verb and clustering with a higher functional morpheme. The question is: "What triggered this change? English went through a process of dramatic changes in the Middle English period with respect to word order and morphology. But it is also tempting to tie it to external influence. Did these changes emerge under French influence from the south? Was it accommodation to a dominant language like French without infinitival prefix?

## 6.2 Pella (Wisconsin)

The *taum*-construction is also found in a Low-German recording from Pella (Wisconsin), available from the *Databank für Gesprochenes Deutsch*. Though without metadata documentation, the recording seems Pomeranian to my ear, and my transcription of the same Wenker sentence 16 in (3) and (5) in this variety is presented in (9).<sup>22</sup>

- (9) Pella Pomeranian (DGD-IDS, MV-E138)  
Du büst no nit groot nauch to 'n bottel ut-Ø-drinken.  
you are yet not big enough to a bottle PRT-Ø-drink.INF2

Notice that C is lexicalised with simple *tau* rather than *taum*. This is evidence for the movement of T to C. These data might feed the idea that the split infinitive originates from Europe. However, as T-to-C is, by hypothesis, a formal option

---

<sup>22</sup>IDS database, <http://dgd.ids-mannheim.de>.

- (i) File: MV--\_E\_00138\_SE\_01\_A\_01\_DF\_01.WAV, time 00:02:01.0.  
(ii) du büs no ni groot nauch to ... ain flash wiin ut-Ø-drinke:  
File: MV--\_E\_00136\_SE\_01\_A\_01\_DF\_01.WAV, time 00:01: 54.0.

Louden (2009: 175) reports the more traditional [um ... Ø]-pattern in Hamburg (Marathon county (Wisconsin)):

- (iii) Du bist noh nit groot genaug, um et Glas Wien ut-Ø-drinken.

of UG that can arise at various moments, we should not exclude the possibility that the split *tau* + V-en construction has arisen as a consequence of language contact between Germanic with a prefix (European Pomeranian) and a language without such prefix (Modern English).<sup>23</sup>

### 6.3 Altschlage

In the Pomeranian area that was checked (the regions Schivelbein, Regenwald, Belgard, Colberg-Cörlin, Cöslin, Greifenberg, Schlawe), I found one case with a raising of the *tau* prefix, though without deletion of the lower copy, given in (10).

- (10) Altschlage (W 00148)

Du büst no nie grot nouch to ne Flasch Wiin ut tau drinken.  
you are yet not big enough FOR a bottle wine empty TO drink.INF2

We might see this as a precursor of a high spellout of *tau* in the chain. Altschlage is present-day Ślawa (Świdwin). It used to be a Wendish settlement. Slavic languages lack an infinitival prefix, and it lexicalises the complementiser. In this case, accommodation to a language with infinitival prefix is not very plausible as the prefix is retained. Only the chain as such is lexicalised, which is a universal structure. It maximally shows a kind of agreement between the C position and the T position. It might be used as evidence for the abstract movement of T to C, but not for accommodation.

### 6.4 Alemannic

The *taum* construction also occurs in the Wenker material in the Alemannic dialects of Switzerland and Austria (Vorarlberg) (Seiler 2005), as illustrated in (11a) and (11b), respectively.

- (11) Alemannic (Fläsch (Graubünden) and Krumbach, resp)

- a. du bisch noh z Klii zum a Fläsch Wi us-∅-trinka
- b. du binscht no nit gros gnug, zum a flöscha wing us-∅-trinken

---

<sup>23</sup>Modern English lost *to* as a prefix, as *to* can be separated from the verb by adverbs (“split infinitives”).

(i) My mother asked me **to** quickly ∅-go to the market.

It is unclear to which functional projection it has raised. It has not raised as far as C.

Direct influence of Alemannic on Brazilian Pomeranian is improbable. Though there is a Swiss community in the Pomeranian area in Espírito Santo, its earliest immigration to the Santa Leopoldina area consisted of 30 Catholic families (Franceschetto 2014: 155). The Pomeranian community and the *Suiça* community were segregated by religion: Lutheran versus Catholic.<sup>24</sup> As to the origin in Europe, it must be noticed that Alemannic is in close contact with Rhaeto-Romance and vice versa. For instance the V2 properties in Rhaeto-Romance are probably due to language contact with Germanic. If so, the *taum* construction could be a sign of language contact in reverse direction. Notice that this contact has happened before the split in religion during the reformation. There is evidence that there is T-to-C as early as in Middle Alemannic of around 1470.<sup>25</sup> Such contacts with Romance in the Vorarlberg are also reported, as it has been germanised from the 9<sup>th</sup>–16<sup>th</sup> century (Klausmann & Krefeld 1995: 4).<sup>26</sup>

## 6.5 Schwabian

A similar construction is reported in Schwabian (cf. Hoekstra 1997: 23), who analyzes the floating ‘to’ as head movement to C or to Asp. I give three examples in (12).

(12) Schwabian (Müller 1996)

- a. dass'r extra hoimkomma isch [zom schnell des Päckle  
that he expressly home come is FOR.TO quickly the parcel  
auf-∅-macha]  
open make  
'that he came specially home to open the parcel quickly'

---

<sup>24</sup>In 1860, there was a big Catholic church in the center of the area, and a small Protestant chapel at the edge, which were in conflict to a point that the governor of the state had to intervene (de Tschudi 1860: 139).

<sup>25</sup>Examples from MHG bible of ~1470 in an Alemannic/Schwabian dialect:

- (i) vnd er gabe in [zewerden offen]  
'and he gave him to become open'
- (ii) Du gibst nit deinen heiligen [ze sehen die zerbrochenkeit]  
'you give not your saints to see the broken-ness'

The infinitival prefix *ze* ‘to’, being a bound morpheme, pied-piped the verb, creating VO contexts.

<sup>26</sup>I thank one of the reviewers for drawing my attention to this.

- b. dass'r mir a weng Zoet brauchat [zom des neie Haus en dr  
that-PRT we a little time needed [FOR.TO the new house in the  
Garteschtross en Tiebenga zom/z' baua]  
Gartenstraße in Tübingen FOR.TO/TO build  
'that we needed little time to build a new house in Garden Street in  
Tübingen'
- c. dass'r [kurz den Zättl zom aus zom/z' fülla] ogfanga hat.  
that he quickly the form FOR.TO out FOR.TO/TO fill begun has  
'that he began to quickly fill out the form.'

Schwabian is not in direct contact with Romance, though it, of course, participates in the wider Alemannic linguistic space which is in contact with Italian and French. The lack of direct contact, however, makes the accommodation scenario improbable.

## 6.6 Tyrolese

The taum-construction can be found in the Wenker sentence 16 in some villages in Tyrol, as given in (13).

- (13) Tyrolese (Reith b. Brixlegg)  
 Du bischt no nit grousz gnuag zum a flosch win aus-∅-drink'n  
 'you are not big enough yet to drink a bottle of wine'

Direct influence of Tyrolese on Brazilian Pomeranian is improbable. There is a community *Tirol* in Espírito Santo not far from the Pomeranian area, but the inhabitants are separated by religion (Catholic versus the Lutheran Pomeranians). Segregation on the basis of religion has always been strong (Schabus 2009), even until the present day. As to the origin of the construction in Europe, the construction might have emerged in Tyrol in Europe by contact with Romance, in this case Rhaeto-Romance.

## 6.7 Twentieth century European Pomeranian

There is also the possibility that the *taum* construction is native from Pomerania. In Stritzel (1974: 69), a Pomeranian grammar from the 1930's, a similar construction for the village of Grossendorf<sup>27</sup> is reported, given in (14a). Furthermore, there is at least one example (an idiomatic expression) in a Pomeranian dictionary (cf. 14b, taken from Laude & Schnibben 1995).

---

<sup>27</sup>Present-day Wielka Wieś (Pomeranian Voivodeship).

(14) 20<sup>th</sup> c European Pomeranian (Großendorf and Kowalk, resp.)

- a. dan is də še<sup>i</sup>nstə tīd [tum draxən stījən lō<sup>u</sup>tən]  
then is the nicest time FOR.TO drake rise Ø let.INF2  
'then it is the best time to let climb the dragon/kite'
- b. dat is já tam up d' boim kleppre  
that is PRT FOR.TO upto the tree Ø climb.INF  
'that is to become desperate.'<sup>28,29</sup>

This might be a sign of an older presence of the *taum* construction, but it may also be a later, parallel development. It is certainly not evidence that the construction was already in Pomerania in the days of the Pomeranian emigration to Brazil.

## 6.8 Flemish in Brazil

An extremely interesting case is the mixed speech of the Dutch immigrants from the Flemish part of the province of Zeeland (Zeeuws-Flemish). These settlers were Calvinist but converted to Lutheranism and are now part of the Pomeranian community. In multilingual speakers (Flemish, Pomeranian, Portuguese), one can observe constructions, like the ones in (15), where the infinitival prefix *te* has been attached to the complementiser *om*. These are obvious constructions under a strong Pomeranian influence (calques), as might be derived from the typical *do*-support, and from the participle without prefix *ge-* in *kommen* 'come' instead of the expected form *gekommen*. The fact that Flemish *te* has overly moved to *om* in C is a welcome confirmation of my analysis of Pomeranian *taum* as *um+tau*. What makes this construction special is that the order of the lexical ingredients in *om-te* is reversed with respect to the *taum* construction, where the prefix is initial.

<sup>28</sup>Notice the form in *-e*, where one would expect INF2. Kowalk (Kowalki, no Wenker location) patterns with the villages Zeblin (Cybulino, W00453), Groß Leistikow (Lestkowo, W50506), Barfussdorf (Zolwia Bloc, W51121), Köpik (Kopice, W50482), Drammin (Dramino, W50731), Liepnitz (Lipnica, W00374) in two Pomeranian properties: they have no INF2 form and display strong adjectival endings. Kowalk's neighboring village Groß-Tychow (Tychowo, W00346) displays *n*-infinitive and weak endings.

<sup>29</sup>A reviewer draws attention to the fact that this expression also exists in Standard German: "Das ist ja zum auf die Bäume klettern". The Pomeranian example may be a translation of the Standard German saying.

## (15) (Zeeuws) Flemish in Brazil

- a. dat es dan vier dagen om-te naar Santa Leopoldina    kommen  
that is then four days FOR.TO near Santa Leopoldina Ø come.INF  
en dan vier dagen weer om-te terug    kommen  
and then four days again FOR.TO back Ø come.INF  
'It is then four days to go to SL and four days to come back'
- b. as jinne krank worden deed, dan was gien auto om-te die    weg  
if one ill    get    did, then was no car FOR.TO those away  
            Ø bring  
'if somebody got ill, there was no car to bring them away'
- c. om-te dan goeid    bikieken waar ons folk kommen es  
FOR.TO then well Ø look.INF where our folks come    is  
'to see well in what situation our people has arrived'

These structures are, therefore, more similar to the Middle English constructions discussed in Section 6.1. The more conservative linear order *om-te* is also what I expect, as Flemish lacks a precursor like Pomeranian *taum + N*, illustrated in (7) above. The Standard Dutch counterpart [*ten + N*] is a high-register structure, and absent in Dutch dialects. Without doubt, language contact with Pomeranian is responsible for the emergence of this overt T-to-C movement. Important to note is that the overt movement of T-to-C can be observed in the Garrafão area (with a high density Pomeranian speakers), not in the Holandinha area (with a low number of Pomeranians), cf. (16).

## (16) Dutch and Pomeranian Varieties in ES

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>a. om frunne te maken</li> <li>b. om-te frenne Ø maken</li> <li>c. taum farijn Ø måken</li> </ul> | Flemish in Holandinha<br>Flemish in Garrafão<br>Brazilian Pomeranian<br>'to make manioc flour' |
|--|--|

Apparently, the presence of Portuguese is not a sufficient trigger for the change I am discussing in this paper, while the presence of Pomeranian did cause such a change in this variant of Flemish.<sup>30</sup> Hence, accommodation is not a sufficiently

---

<sup>30</sup>Their Flemish is a rather uncertain heritage Flemish, while their Pomeranian is robust, just as the Pomeranian of the Pomeranians. These people are Pomeranians with an additional heritage Flemish.

explanatory factor. We rather must think in terms of variation: while the internal variation of these Flemish variants was not rich enough to cause a change to overt T-to-C, the Flemish-Pomeranian melting pot was sufficiently rich for dialect convergence towards both the *taum*-construction and the *om-te* construction.

## 7 Dialect convergence or contact-induced accommodation?

In the previous sections, I discussed a range of West Germanic varieties that have lost the infinitival prefix and realised it, or rather its functional head, higher up in the syntactic hierarchy. We are now in the position to evaluate the various scenarios that might have led to the innovation, shared by Middle English and modern Alemannic. These dialects behave parallel to Pomeranian in Brazil in that they lexicalise the T-chain high. The null hypothesis is that all these parallel cases receive a parallel explanation. There is the accommodation scenario, which hypothesises that the *taum* construction emerged in Brazilian Pomeranian in contact with Portuguese, which lacks the infinitival prefix, just like French, Slavic, and modern English. Alternatively, we have the koineisation scenario in a newly created melting pot community. This scenario fundamentally reduces the number of variants furnished by the source dialects. This explanation has the variability in the source dialects as a fundamental ingredient. It is obviously an advantage of the latter scenario that it puts the variability discussed in Section 3 on a fundamental footing. Long-term, structural accommodation is only possible upon intensive contact. If we now see to what extent there has been actual contact in all these cases, the balance is not completely positive, as can be seen in Table 4.

Let us discuss the table briefly. Language contact between Middle English and Anglo-Norman is uncontroversial in both directions (Mustanoja 1960, Dalton-Puffer 1996, Ingham 2012, Rothwell 2001, Steiner 2010).<sup>31</sup> In the case of Altschlage, there is no positive evidence of the contact with Slavic, but it cannot be excluded, as it was a Wendic settlement. This might have triggered a C+T complex, as the high *to* in (10) indicates. However, the lower copy *tau* is not silent. If language contact was involved, it apparently did not occur on surface level. We leave this

<sup>31</sup>The influence of French on Middle English in the domain of the lexicon is better studied than for syntax and morphology. For some curious reason, the influence of (Anglo)French on Middle English is not as well studied as the influence of Middle-English on (Anglo)French. It is often downplayed as in Thomason & Kaufman (1988: 306ff), but see Ingham (2009) for noteworthy remarks on this issue.

Table 4: Various Germanic contact varieties with complex FOR-TO complementisers.

Language	Loss of prefixal TO	Contact with prefix-less language	Sufficient contact
Middle English	+	Anglo-Norman	yes
Pella(Wisconsin)	+	English	yes
Altschlage	-	Slavic	doubtful
Alemannic	+	Rhaeto-Romance/ Franco-Provençal/ French	yes
Schwabian	+/-	no	no
Tyrolese	+	Ladin	yes (Western part)
20 <sup>th</sup> c EPom	+/-	Slavic	unknown
BPom	+	Portuguese	yes
Flemish in Hollandinha	-	Portuguese	yes
Flemish in Garrafão	+	BrPomeranian	yes

case open. For Pella (Wisconsin), language contact may have been present beyond doubt, but it is not clear if there has been a Pomeranian cohesive community. There are too few speakers to evaluate this single fact<sup>32</sup>, but contact with English has been strong. For Schwabian, direct contact with a prefix-less language is absent, though it can have happened indirectly through Swiss sister dialects. For Brazilian Pomeranian, contact with Portuguese is present in modern times, as has been shown by Schaffel Bremenkamp (2014: 177, graph 4), though 50% of the older present-day speakers are still monolingual. If accommodation were the causing factor, we would expect that the *taum*-construction would be less used by older speakers. There is no evidence of this kind.<sup>33</sup> Taking all these doubts into account, I conclude that there is too little evidence to either support or to reject the accommodation hypothesis.

This brings us to evaluating the koineisation scenario with its four mechanisms as discussed in Section 2. The options in Sections 2.2.1–2.2.3 only fit with

<sup>32</sup>The Wenker-sentences given in Louden (2009: 175) make a distinction between two infinitives 1 and 2, as in EP and BP. The infinitival *tau* is silent and the complementiser is *um*, as in Lankow (4b) above.

(i)Du bist noh nit groot genaug, **um** et Glas Wien ut-∅-drinken

<sup>33</sup>In 4 interviews by Anivaldo Kuhn in 2003 of a ~75 years old Pomeranian, the *taum*-construction already occurs abundantly: 30 times (on ~4000 words) of which 15 with an actual lexical split [*taum* xxx V-en] (of which 5 bare nouns/adjetives might have been incorporated into the verb). The interviews are in Seibel (2010: 507–556).

some artificiality on the facts under scrutiny. One could argue that instead of lexicalizing a chain optionally in a scattered way, as the European Pomeranian dialects do, Brazilian Pomeranian opts for lexicalizing T higher up jointly with C (*taum=tau+um*). This strategy can be seen as a very particular variant of levelling (i.e. loss of most source variants): in this case loss of *all* variants. However, Brazilian Pomeranian did not just lose the three input variants of the scheme in (5), it also created a new one (6b) on the basis of the underlying syntactic skeleton. So, levelling is an insufficient mechanism to capture what happened. One could also argue that it is a very particular kind of interdialect formation: the emergence of new forms that are intermediate of the input dialects. To what extent lexicalizing two positions higher up in the syntactic hierarchy instead of scattered lexicalisation of a coindexed chain is a case of “intermediate”, is of course open to debate. Finally, one could argue that it must be interpreted as a very particular version of *fudging*: the combination or superposition of two ingredients taken from distinct dialects: lexicalisation of the higher member of the C-T chain (dialects with *um*) and silence of the lower member of the C-T chain (dialects with *tau-drop*) is re-analyzed as *movement*: the lower copy is spelled out high as C+T: *taum* emerges. This is what comes closer to what has happened. But probably the most apt interpretation of the facts is that it should be explained as *revert to the default setting*. Most of the world’s languages lack an infinitival prefix comparable to *tau/to/zu*. Absence of it seems to be the default.<sup>34</sup> And Brazilian Pomeranian complies with it. Moreover, the majority of the world’s languages do lexicalise complementisers in purpose infinitivals, and Brazilian Pomeranian patterns with it as well.<sup>35</sup> Finally, as Pesetsky & Torrego (2007) have argued on formal grounds, there is always an overt or covert T-to-C movement in infinitivals. And this is precisely what *taum* is, the lexicalisation of T+C. So, on all points does Brazilian Pomeranian pattern with the default setting, while this default setting was not present in the source variants. So, theoretically, the dialect convergence scenario seems to have strong cards. Is there then any empirical evidence that can be decisive?

---

<sup>34</sup>The claim that the infinitive is without prefix does not only hold for rationale clauses, but for infinitival clauses in general. In the perspective of *revert to the default*, this does not come as a surprise, cf. (i):

- (i) ik fersuik ais [aira nå hus gåa]  
I try PRT early to house go.INF1  
'I finally try to go home early'

In most of the cases, the German/Dutch construction corresponds to a bare infinitive1 or a finite clause in BP.

<sup>35</sup>It is often difficult to separate prepositions and complementisers in this context. For a discussion and tests, cf. Bennis (1987).

In the next section I make a feature analysis of the constructions and design two models on its basis.

## 8 Modelling accommodation and dialect convergence formally

In this section I make a formal implementation of the two scenarios by which Brazilian Pomeranian infinitival construction [taum ... Ø] can be explained: accommodation to Portuguese and/or dialect convergence to the default settings. I will take the mechanism of *revert to the default*, discussed in Section 2.2.4, as starting point.

### 8.1 Modelling dialect convergence

As we have seen in Table 2, European Pomeranian shows at least 3 variants of this infinitival construction, while one is structurally absent. These variants were the input for the newly created *lingua franca* in Brazil. In the first columns of Table 5, I characterise these 3 + 1 variants in terms of their spellout patterns of functional morphemes in the secound column. Logically, there are  $2^3=8$  possible patterns in total. For completeness, I have added the remaining possibilities below the separator.<sup>36</sup>

Table 5: Chain analysis of infinitival constructions

	Surface Pattern	Underlying Pattern	Frequency (%)	Variant
a.	um ... tau	um-tau ... tau	83	EP
b.	um ... Ø	um-tau ... tau	6	EP
c.	Ø ... tau	um-tau ... tau	11	EP
d.	Ø ... Ø	um-tau ... tau	0	(EP)
e.	taum ... Ø	um-tau ... tau	98	BP
f.	tau ... tau	um-tau ... tau	—	Alt-Slage
g.	tau ... Ø	um-tau ... tau	2	BP/Pella
h.	taum ... tau	um-tau ... tau	—	Schwab

<sup>36</sup>The extra patterns include those of Altschlage (cf.(10)), the BP pattern (i) in note 18 and Pella Pomeranian (cf. (9)), and the Schwabian variant mentioned by Müller (1996) in (12b)

The classification in terms of its surface appearance does not display the underlying grammatical factors, however. There are three grammatical features involved, which all concern the spellout of *chains*. First, there is a  $\pm$ -lexicalisation of the FOR-chain, which is a singleton chain with or without spellout. Secondly, there is a  $\pm$ -lexicalisation of the TO-chain, which is a binary chain (“movement”). It does or does not have a chain spellout. Thirdly, the TO-chain, which is a movement chain, can have a high spellout (overt movement) or a low spellout (covert movement). This is ruled by the DELETE-process of chain reduction, as described in Nunes (1995). This is captured by  $\pm$ low-DELETE. In columns 3-5 of Table 6, I describe the parameter settings of these input variants with values yes/no. Finally, these features must be projected in a consistent way on markedness of the settings: marked (+) or default (0). Let us assume that lexicalizing a chain is the default (applied to FOR and TO-chain equally). Let us furthermore assume that overt movement is the default i.e. DELETE of the lower copy is the default. I indicate the corresponding markedness values of the input in gray-shade. These are the EP input varieties upon entering Brazil. The BP parameter output is in the fifth row (dashed) in (row e).

Table 6: Convergence Model - Feature analysis and markedness

Infinitival construction		Parameter Settings			Markedness			
Pattern	Variety	P1=FOR chain	P2=TO chain	P3=low DELETE	P1	P2	P3	
a.	um-tau... tau	EP	yes	yes	no	0	0	+
b.	um-tau... tau	EP	yes	no	yes	0	+	0
c.	um-tau... tau	EP	no	yes	no	+ 0	+ 0	+
d.	um-tau ... tau	-	no	no	yes	+ +	+ 0	0
e.	um-tau ... tau	BP	yes	yes	yes	0 0	0 0	0 0
f.	um-tau ... tau	Alt-Sl	no	yes	no	+ 0	0 +	
g.	um-tau ... tau	BP/Pella	no	yes	yes	+ 0	0 0	
h.	um-tau ... tau	Schwab	yes	yes	no	0 0	0 +	

Let me now show the convergence mechanism in progress. As to the FOR-chain lexicalisation (shaded P1 column), the input dialect set contains two dialect types with a default setting (row a and b) and one dialect type with a marked setting (row c). Upon interaction, the outcome in (row e) is the default value. As to the lexicalisation of the TO-chain (shaded P2 column), the input set contains two dialect types with default setting (row a and c) and one dialect type with marked

setting (row b). The outcome in (row c) opts for the default setting. Finally, as to the chain spellout (low or high) in the last column, we observe interaction of one dialect type with default setting (row b) and two dialect types with marked settings (row a and c). Once again, the outcome is the default setting. In sum, for the three relevant parameters, revert to the default setting describes the dominant outcome in Brazil adequately. This default setting of the three features as well as the marked settings were already present in one of the input dialects. Therefore, the process can be described as a purely Pomeranian-internal effect: dialect mixing can produce Brazilian Pomeranian under revert to the default if present in the linguistic input. This might be taken as evidence that the Pomeranian *lingua franca* in Brazil has resorted to the default setting in all three relevant parameters upon language contact with conflicting input in the three parameters.

One might wonder why the majority choice of European Pomeranian [FOR .... TO] did not impose itself in Brazil. Under the assumption that the figure of 83% in Table 5 is valid for the immigrants as well, it might come as a surprise that the emigrants followed a completely different path, especially considering the fact that the European [FOR ... TO]-variant is identical to the Standard German variant, a prestige variety that was taught in the parochial schools to some of the community members. There are three points to consider here. In the first place, the interaction (convergence) of two closely related dialects takes place on parameter level, not on surface level. This is precisely the point I want to make: the default setting approach can produce something new, which cannot be explained by considerations of dialect dominance. So the outcome in BP converging to the new [FOR-TO .... Ø] is a strong argument in favor of the parameter approach. Secondly, neither the dominant EP variety nor HG with [FOR ... TO] realise the default setting according to the analysis in Table 5. Hence, even these varieties might decline if they were sufficiently shuffled upon social changes. Third, in the case of, say, two or three caretakers with slightly different dialects, we have the situation of 2L1 or 3L1 and the interaction takes place according to the scheme in Table 5, not on the level of societal statistics. That being said, I do think that societal statistics are relevant: They play a role in the case of accommodation, as we will see in the next section.

## 8.2 Modelling accommodation

By a simple modification of the model of Section 8.1, I can turn it into a model of accommodation, as we will see in an instance. Since we use universal claims of what is default and what is marked, the only locus for a different implementation is the parameter describing covert and overt movement. I captured this dimension in Section 8.1 by checking if the lower link of the chain was deleted or not,

which was default or not default, respectively. However, it can also be checked if the upper link is deleted or not, of course with the reverse markedness assignments. Let us, therefore, consider a parameter P4 that describes the lexicalisation of the higher copy. For this P4, upper link deletion is marked (instead of lower-link deletion being the default). The core cases of overt and covert movement then still project on the same markedness as they did in the convergence model of the previous section. Only in the case of double spellout or non-spellout does the new parameter give distinct results. The two models are compared in Table 7.

Table 7: Chain reduction: low/high DELETE as ruling parameters and their respective markedness.

	Convergente Model using P3		Accommodation Model using P4	
Pattern	low DELETE	markedness	high DELETE	markedness
a. tau ...tau	no	+	no	0
b. tau ... <del>tau</del>	yes	0	no	0
c. <del>tau</del> ...tau	no	+	yes	+
d. <del>tau</del> ... <del>tau</del>	yes	0	yes	+

P3 is low DELETE; P4 is high DELETE.

With the Model-2 implementation, I arrive at the evaluation Table 8. To see how it works, let us inspect Row-a and Row-b in Table 8 with respect to P4 (the features P1 and P2 remain unchanged). Row-a has [um ... tau], which is, as to the *tau*-string: [~~tau~~ ... tau], which is the case of Table 6c with markedness value +. The next case in Row-b is [um ... Ø], which is, as to the *tau*-string, [~~tau~~ ... ~~tau~~], which is the case of Table 6d with markedness +, etc. Only the P4 column differs from the Convergence Model of Table 5. Once again, the Brazilian Pomeranian [*taum* ... Ø]-pattern realises the default setting (000), which BP now shares with the Schwabian [*taum* ... *tau*]-pattern. This model has two absolute default settings: the BP [*taum* ... Ø] in Table 8e and the Schwabian [*taum* ... *tau*] in Table 8h.

The most important consequence is that the four source dialects, Table 8a-d, are homogenous in P4 (with a marked setting), while all high-contact varieties in Brazil, Pella(Wisconsin), and Alt-Schlawe, are homogenous with an unmarked setting. In this model, the flip in the P4-value cannot be produced by internal dialect convergence (there is no variation in the P4 parameter) and must be due to an external trigger of the P4-flip. If one can prove that the Portuguese pattern [*para* ... Ø] does not realise the case of Table 8b, but either Table 8e or Table 8g, then the flip in P4 might have been caused by language contact and accommodation to Portuguese. Let us assume that there is such evidence.

Table 8: Accommodation Model - Feature analysis and markedness

Infinitival construction		Parameter Settings			Markedness		
Pattern	Variety	P1=FOR chain	P2=TO chain	P4=high DELETE	P1	P2	P4
a.	um-tau... tau	EP	yes	yes	yes	0	0
b.	um-tau... tau	EP	yes	no	yes	0	+
c.	um-tau... tau	EP	no	yes	yes	+	0
d.	um-tau ... tau	-	no	no	yes	+	+
e.	um-tau ... tau	BP	yes	yes	no	0	0
f.	um-tau ... tau	Alt-Sl	no	yes	no	+	0
g.	um-tau ... tau	BP/Pella	no	yes	no	+	0
h.	um-tau ... tau	Schwab	yes	yes	no	0	0

The question is then if we can find independent evidence to choose between the two models in Table 6 and Table 8, i.e. we must choose between the features P3 and P4. I will now show that frequency values of the dialects provides us with such independent evidence. To see how, one should realise that it is plausible that a higher level of markedness corresponds to a lower occurrence of the variant and vice versa. So let us define the total markedness,  $\mu$ , of a language variant as the sum of its markedness values. In Table 9 I have represented the Dialect Convergence Model (Model 1) with P1, P2, P3 and the Accommodation Model (Model 2) with the features P1, P2, P4. In the columns headed by  $\mu$ , I added the respective sums of the marked settings.

In order to evaluate the two models with more ease, I displayed the values of the total markedness  $\mu$  and the occurrence rates of the varieties into the *markedness graphs* under Figure 3 and Figure 4. These graphs have the total markedness  $\mu$  on the vertical axis. The horizontal axis is the time axis with before and after the language contact, convergence in Figure 3 and accommodation in Figure 4. In both graphs we observe a local minimum before and after the interaction. Moreover, the local minimum before the interaction is higher than the local minimum after the interaction. So, what happens in both models is a decrease in markedness. However, the models differ in what feature(s) cause(s) this decrease. In the model in Figure 3, all three features are involved and choose the value of the lowest markedness. Hence, this can be interpreted as a convergence model. However, if I put the occurrence rates in the graph (as a % subscript), we must conclude that

Table 9: Comparison of the Convergence Model and the Accommodation Model

Surface Pattern	Freq. in %	Infinitival				Convergence				Accommodation			
		P1	P2	P3	$\mu$	P1	P2	P4	$\mu$	P1	P2	P4	$\mu$
a. um ... tau	83	0	0	+	1	0	0	+	1				
b. um ... Ø	6	0	+	0	1	0	+	+	2				
c. Ø ... tau	11	+	0	+	2	+	0	+	2				
d. Ø ... Ø	0	+	+	0	2	+	+	+	3				
e. taum ... Ø	100	0	0	0	0	0	0	0	0				
f. tau ... tau	0.3	+	0	+	2	+	0	0	1				
g. tau ... Ø	-	+	0	0	1	+	0	0	1				
h. taum ... tau	-	0	0	+	1	0	0	0	0				

P1–4 are the features involved (see the text);  $\mu$  is the total markedness

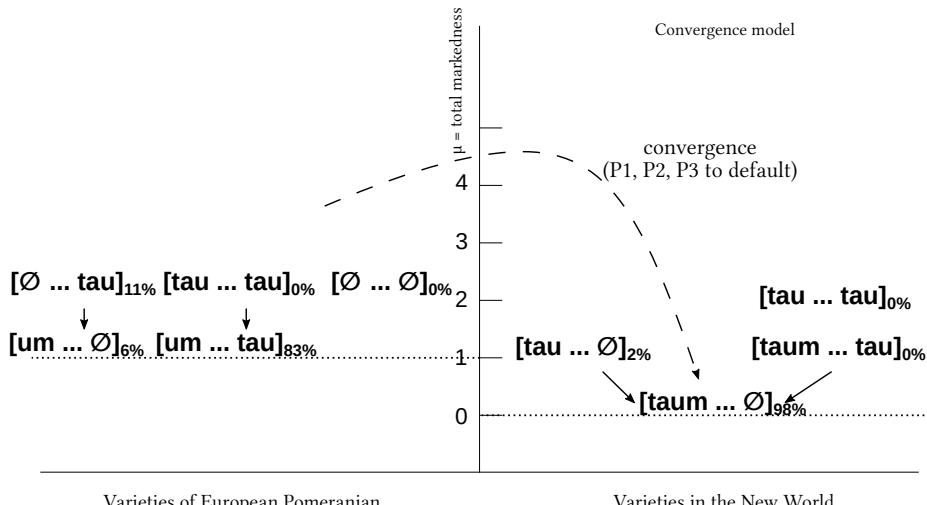


Figure 3: Markedness graphs belonging to the Convergence Model with occurrence rates.

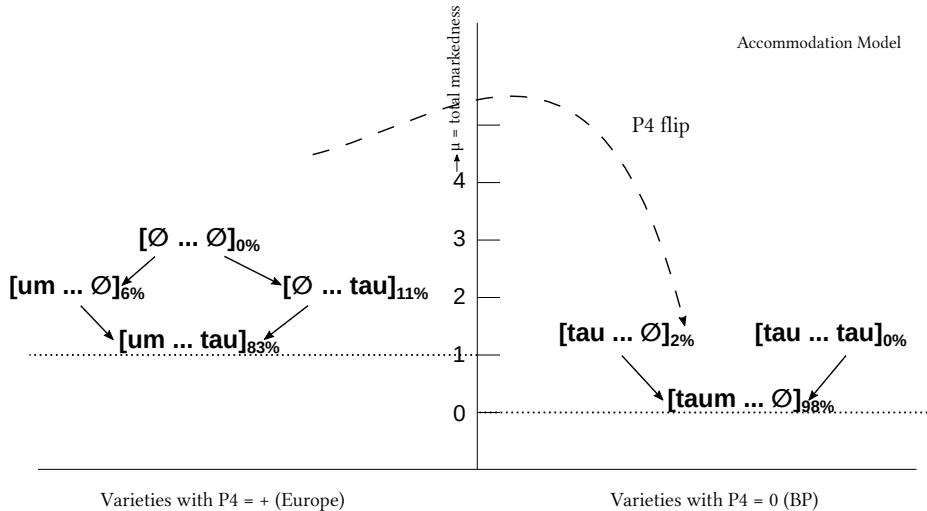


Figure 4: Markedness graph of the Accommodation Model with occurrence rates.

the occurrence rates do not correlate in any way with the level of markedness.

In the markedness graph in Figure 4, on the other hand, we observe two sets of dialects with respect to their value of P4. The flip in P4 coincides with their identification as low and high contact varieties. Interestingly, the value of their markedness neatly correlates with their relative frequencies. The absent  $[\emptyset \dots \emptyset]$ -pattern has the highest markedness of  $\mu = 3$ . The most general [FOR ... TO] pattern is a local minimum of 1. The general [FOR.TO ...  $\emptyset$ ] in BP has markedness 0.

We may, therefore, use the occurrence rates of the varieties and their relation to markedness as independent evidence that the P4-feature provides a better model of the change that Pomeranian underwent upon its settlement in Brazil, than the convergence model with the P3 parameter. It might also be taken as evidence that P4 is a better measure of the difference in markedness of covert-overt movement in general.

If we take the occurrence rates into account, I come to a different conclusion than my 2016 study: what has happened in the emergence of BP, is not dialect convergence within Pomeranian itself, triggered by the high level of *variation* present in the input dialects, but accommodation to an external language, Portuguese.

## 9 Conclusions

The sociolinguistic observations on Pomeranian, with language variation in Europe and convergence to a uniform construction in Brazil, provides evidence for an underlying syntactic C-T chain in natural languages, as was argued for in Pesetsky & Torrego (2007) on formal grounds. While European Pomeranian shows *variation* in the lexicalisation of this [FOR ... TO] chain with a three-fold optionality, Brazilian Pomeranian displays obligatory lexicalisation of the higher link of the chain and obligatory silence of the lower link. This configuration is reanalyzed as an overt movement relation of T to C, which is the default option in natural language. There are language-internal arguments that the new construction is a result of dialect-convergence to the default setting of the parameters involved. However, when we take the external occurrence rates into account, the data indicate that the similarity in this respect between Brazilian Pomeranian and (Brazilian) Portuguese might be analyzed as accommodation of Brazilian Pomeranian to the dominant language Portuguese.

## Abbreviations

ACC	accusative	OBL	oblique
COMP	complementiser	PRT	particle
DAT	dative	PL	plural
INF1	infinitive in -e	REFL	reflexive
INF2	infinitive in -en	SG	singular

## Acknowledgements

This is an extended version of Postma (2016), written in German, that models dialect convergence. This English version models both accommodation and dialect convergence. Moreover, it adds a graphical tool to render the level of markedness of the various variants. These insightful diagrams turned out so powerful that they partly changed the conclusions with respect to Postma (2016). I thank my colleagues at the Meertens Institute, the audiences of the Workshop “German Abroad”, Vienna 2014, the audience of a talk at FFLCH at USP, Nov 11, 2013, of the Colóquios de Sintaxe, Aquisição e Mudança, University of Campinas, Nov 12th, 2013, of the workshop on Heritage Languages, Amsterdam, August 18, 2014, and of ICLaVe10, 26 June 2019, Leeuwarden for their comments and suggestions. I am grateful to two anonymous reviewers their helpful comments. A word of

gratitude to my informants Elizana Schaffel, Teresa Gröner, and Hilda Braun for their help and patience. A special thanks to Andrew Nevins with whom I did part of the fieldwork. All errors are mine.

## References

- Baker, Mark C. 1988. Incorporation: A theory of grammatical function changing.
- Bayer, Josef. 1984. COMP in Bavarian syntax. *The Linguistic Review* 3. 209–274.
- Bayer, Josef & Ellen Brandner. 2004. Klitisertes zu im Bairischen und Allemannischen. In F. Patocka & P. Wiesinger (eds.), *Morphologie und Syntax deutscher Dialekte und Historische Dialektologie des Deutschen*, 160–188. Wien: Edition Präsens.
- Bennis, Hans. 1987. *Gaps and dummies* (Linguistic Models 9). Dordrecht: Foris Publications.
- Bennis, Hans & Teun Hoekstra. 1984. Gaps and parasitic gaps. *The Linguistic Review* 4. 29–87.
- Bickerton, Derek. 2015. *Roots of language*. Martin Haspelmath & Stefan Müller (eds.) (Classics in Linguistics 3). Berlin: Language science press. DOI: 10.17169/langsci.b91.109.
- Brandner, Ellen. 2006. Bare infinitives in Alemannic and the categorial status of infinitival complements. *Linguistic Variation Yearbook* 6(1). 203–268.
- Britain, David. 1997. Dialect contact, focusing and phonological rule complexity: The koineisation of Fenland English. *University of Pennsylvania Working Papers in Linguistics* 4. 141–169.
- Brockhaus. 2012. *Brockhaus Enzyklopädie in 30 Bänden*. Annette Zwahr (ed.). F.A. Brockhaus.
- Chambers, Jack K & Peter Trudgill. 1998. *Dialectology*. Cambridge: Cambridge University Press.
- Cornips, Leonie. 2006. Intermediate syntactic variants in a dialect-standard speech repertoire and relative acceptability. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradience in grammar: Generative perspectives*, 85–105. Oxford: Oxford University Press.
- Dalton-Puffer, Christiane. 1996. *The French influence on Middle English morphology: A corpus-based study on derivation*. Berlin: De Gruyter Mouton.
- Demske, Ulrike. 2011. Finale Marker in Infinitivkonstruktionen: Wandel und Variation. In Helen Christen, Franz Patocka & Evelyn Ziegler (eds.), *Strukturen, Gebrauch, und Wahrnehmung von Dialekt*, 29–46. Präsens Verlag.

- de Tschudi, Johann Jakob. 1860. *l'Envoyé extraordinaire au Brésil, Mr de Tschudi, sur les colonies de Santa Isabel, Santa Leopoldina et Rio Novo, etc. (Du 20 Décembre 1860)*. Vol. 13. Berne: Feuille Fédérale Suisse.
- Dillard, Joey Lee. 1972. *Black English: Its history and usage in the united states*. New York: Vintage Books.
- Droogers, André. 2008. Religião, identidade e segurança entre imigrantes luteranos da Pomerânia, no Espírito Santo (1880–2005). *Religião & Sociedade* 28(1). 13–41.
- Evers, Arnold. 1990. The infinitival prefix “zu” as INFL. In Günther Grewendorf & Wolfgang Sternefeld (eds.), *Scrambling and barriers* (Linguistik Aktuell 5). Amsterdam: John Benjamins Publishing.
- Franceschetto, Cilmor. 2014. *Imigrantes – Espírito Santo; Base de dados da imigração estrangeira no Espírito Santo nos séculos XIX e XX*. Vitória: Arquivo Público do Estado de Espírito Santo.
- Frings, Theodor. 1936. *Die Grundlagen des Meissnischen Deutsch: ein Beitrag zur Entstehungsgeschichte der deutschen Hochsprache*. Tübingen: Max Niemeyer Verlag.
- GLFP. 1932. *Gemeindelexikon für den Freistaat Preussen: Provinz Pommern*. Berlin: Preußisches Statistisches Landesamt.
- Granzow, Klaus. 2009. *Pomeranos sob o Cruzeiro do Sul: Colonos alemães no Brasil*. Vitória: Arquivo Público do Estado do Espírito Santo.
- Gumperz, John J. & Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In Dell Hymes (ed.), *Pidginization and creolization of languages*, 151–167. Cambridge: Cambridge University Press.
- Haider, H. 2010. *The syntax of German* (Cambridge Syntax Guides). Cambridge: Cambridge University Press.
- Haspelmath, Martin. 1989. From purposive to infinitive: A universal path of grammaticalization. *Folia Linguistica Historica* 10. 287–310.
- Hinskens, Frans. 1996. *Dialect levelling in Limburg: Structural and sociolinguistic aspects*. Tübingen: Niemeyer.
- Hoekstra, Jarich. 1994. Pronouns and case: On the distribution of Frisian *harren* and *se* ‘them’. *Leuvense Bijdragen VL* 83. 47–65.
- Hoekstra, Jarich. 1997. *Syntax of infinitives in Frisian*. Groningen University. (Doctoral dissertation).
- Hoekstra, Teun. 1984. *Transitivity: Grammatical relations in government-binding theory*. Dordrecht: Foris.

- Höhle, Tilman N. 2006. Observing non-finite verbs: Some 3V phenomena in German-Dutch. In Patrick Brandt & Eric Fuß (eds.), *Form, structure, and grammar: A Festschrift presented to Günther Grewendorf on occasion of his 60th birthday*, 55–77. Berlin: Akademie Verlag.
- Ingham, Richard. 2009. *Middle English and Anglo-Norman in contact*. Talk presented at Conference of the English Literary Society of Japan, Tokyo University.
- Ingham, Richard. 2012. The transmission of Anglo-Norman. In Richard Ingham (ed.), *Language history and language acquisition*. Amsterdam: Benjamins.
- Kayne, Richard S. 1999. Prepositional complementizers as attractors. *Probus* 11. 39–73.
- Klausmann, Hubert & Thomas Krefeld. 1995. Zwischen Isolation und Sprachkontakt: Der romanische Wortschatz der Vorarlberger Walser. *Zeitschrift für Dialektologie und Linguistik* 62. 1–28.
- Koster, Jan & Robert May. 1982. On the constituency of infinitives. *Language* 58(1). 116–143.
- Labov, William. 1994. *Principles of linguistic change: Cognitive and cultural factors*. Vol. 3. Oxford: Basil Blackwell.
- Laude, Robert & Frank Schnibben. 1995. *Hinterpommersches Wörterbuch des Per-santegebietes*. Vol. 12. Köln, Weimar: Böhlau Verlag.
- Louden, Mark. 2009. Das Wisconsin-Pommersche im Spiegel deutsch-amerikanische Sprachvarietäten. *Niederdeutsches Jahrbuch* 132. 165–176.
- Müller, R. 1996. *Historisch vergleichende Studien zum Infinitiv im Friesischen*. Tübingen: Eberhard-Karls University. (MA-Thesis).
- Mustanoja, Tauno F. 1960. *Middle English syntax*. Helsinki: Société Néophilologique.
- Nunes, Jairo. 1995. *The copy theory of movement and linearization of chains in the Minimalist Program*. Maryland University. (Doctoral dissertation).
- Pesetsky, David & Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In Simin Karimi, Vida Samiian & Wendy K. Wilkins (eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation*, 262–294. Amsterdam: John Benjamins Publishing.
- Postma, Gertjan. 2004. Structurele tendensen in de opkomst van het reflexief pronomene *zich* in het 15de-eeuwse Drenthe en de Theorie van Reflexiviteit. *Nederlandse Taalkunde* 9. 144–169.
- Postma, Gertjan. 2012. Language contact and linguistic complexity: The rise of the reflexive pronoun *zich* in a fifteenth-century Netherlands border dialect. In Dianne Jonas, John Whitman & Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcomes*, 139–159. Oxford: Oxford University Press.

- Postma, Gertjan. 2014. Complex complementizers and the structural relation with weak T: New (morpho)syntactic data from a Pomeranian language island in Brazil. *STUF – Language Typology and Universals* 67. 627–663. DOI: 10.1515/stuf-2014-0032.
- Postma, Gertjan. 2016. Der Verlust des Infinitivpräfixes *tau* ‘zu’ im Brasilianischen Pommersch: Akkommodation an das Portugiesische oder Dialektkonvergenz. In Alexandra Lenz & Nikolaus Ritt (eds.), *German abroad: Proceedings of the workshop “German Abroad”, 4–5 july 2014* (Wiener Arbeiten zur Linguistik (WAL)), 177–211. Vienna.
- Reuland, Eric & Tanya Reinhart. 1995. Pronouns, anaphors and case. In Hubert Haider, Susan Olsen & Sten Vikner (eds.), *Studies in comparative germanic syntax*, 241–269. Dordrecht: Kluwer.
- Rosenberg, Peter. 2005. Dialect convergence in the German language islands (Sprachinseln). In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect change: Convergence and divergence in European languages*, 221–235. Cambridge: Cambridge University Press.
- Rothwell, William. 2001. English and French in England after 1362. *English Studies* 6. 539–559.
- Sabel, Joachim. 1996. *Restrukturierung und Lokalität: Universelle Beschränkungen für Wortstellungsvarianten*. Berlin: De Gruyter.
- Schabus, Wilfried. 2009. *Tirol no Brasil*. Wattens: Berenkamp Verlag.
- Schaffel Bremenkamp, Elizana. 2014. *Análise sociolinguística da manutenção da língua Pomerana em Santa Maria de Jetibá, Espírito Santo*. UFES. (MA-thesis).
- Seibel, Ivan. 2010. *Imigrante no século do isolamento 1870–1970*. Escola Superior em Teologia, São Leopoldo (RS) Brasil. (Doctoral dissertation).
- Seiler, Guido. 2005. Wie verlaufen syntaktische Isoglossen, und welche Konsequenzen sind daraus zu ziehen? In Eckhard Eggars, Jürgen Erich Schmidt & Dieter Stellmacher (eds.), *Zeitschrift für Dialektologie und Linguistik. Beihefte*, vol. 130, 313–341. Stuttgart: Franz Steiner Verlag.
- Steiner, B. Devan. 2010. Middle English influence on late Anglo-Norman syntax: The effect of imperfect L2 acquisition. *IULC Working Papers* 10(3). 99–115.
- Stowell, Tim. 1981. *Origins of phrase structure*. MIT. (Doctoral dissertation).
- Stritzel, Herbert. 1974. Die Gliederung der Mundarten um Lauenburg in Pommern. In *Deutsche Dialektgeographie XXXIII*. Tübingen: Sändig.
- Taeldeeman, Johan. 1989. A typology of dialect transitions in Flanders. In *New methods in dialectology: Proceedings of a workshop held at the Free University, Amsterdam, December 7–10, 1987*, 155–164.
- Thelander, Mats. 1980. *De-dialectisation in Sweden* (FUMS rapport 86). Uppsala.

- Thomason, Sarah & Terence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tressmann, Ismael. 1998. Bilingüismo no Brasil: O caso da comunidade pomerana de Laranja da Terra.
- Tressmann, Ismael. 2006a. *Pomerisch-Portugiesisch Wörbauk – Dicionário Encyclopédico Pomerano-Português*. Santa Maria de Jetibá: Secretaria de Educação.
- Tressmann, Ismael. 2006b. *Upm Land – up pomerisch språk*. Santa Maria de Jetibá: Secretaria de Educação.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Van de Craats, Ineke. 2009. The role of IS in the acquisition of finiteness by adult Turkish learners of Dutch. *Studies in Second Language Acquisition* 31(9). 59–92.  
DOI: 10.1017/S0272263109090032.
- van Coetsem, Frans. 1988. *Loan phonology and the two transfer types in language contact*. Berlin: De Gruyter Mouton.
- Vanden Wyngaerd, Guido. 1987. *Control Theory*. Antwerp University. (Doctoral dissertation).
- van Gelderen, Elly. 1998. *For to in the history of English*. *American Journal of Germanic Language and Literature* 10(1). 45–72.
- Visser, F. Th. 1963. *Historical syntax of the English language*. Leiden: Brill.
- Wrede, Fred. 1895. Berichte über Wenkers Sprachatlas des deutschen Reichs. XII. *Anzeiger für deutsches Altertum und Deutsche Literatur* xxi. 260–296.



# Chapter 12

## Using data of Zeelandic Flemish in Espírito Santo, Brazil for historical reconstruction

Kathy Rys<sup>a,b</sup> & Elizana Schaffel Bremenkamp<sup>c</sup>

<sup>a</sup>Dutch Language Institute (INT), Leiden   <sup>b</sup>Meertens Institute, Amsterdam

<sup>c</sup>Stricto Sensu Post-Graduate Linguistics Program of the Federal University of Espírito Santo, Brazil

This paper focuses on the case of Zeelandic Flemish in Espírito Santo, an obsolescent language variety spoken by about twenty descendants of Dutch immigrants to Brazil in the nineteenth century. The speech of rusty speakers can be used to reconstruct the original immigrant language. We perform a historical reconstruction of the old Zeelandic Flemish dialect as spoken in the days of emigration, with respect to three linguistic cases: (1) deletion of /l/ in codas and coda clusters, (2) subject doubling in inversion contexts and (3) the inflected polarity markers *yes* and *no*. Our findings demonstrate the historical value of transplanted dialects or speech island varieties (Rosenberg 2005). However, a comparison of our findings with historical data demonstrates that reliance on rusty speaker data alone may sometimes lead to incorrect conclusions and that the data should always be considered from the perspective of language contact as well.

### 1 Introduction

In this paper, we present data from the speech of the last speakers of Zeelandic Flemish in Espírito Santo, Brazil. These speakers are descendants of Dutch immigrants, who left Zeeland in 1858–1862, but faced deprivation and difficulties in adaptation and integration into Brazilian society, with their language threatened



by the majority language Brazilian Portuguese and by another heritage language, viz. that of the Pomeranian immigrants who arrived in Espírito Santo at the same time.

Among the last speakers of Brazilian Zeelandic Flemish, there are true semi-speakers, who only acquired the language incompletely and only by listening on an irregular basis to older speakers, and rusty speakers, who came a long way in learning to speak their mother language perfectly, but who stopped using their language on a regular basis and therefore forgot how to use some of its more complex features. In this paper, we discuss three linguistic features that occur in the speech of four rusty speakers of Brazilian Zeelandic Flemish. We discuss indications that the immigrants have been more conservative and less innovative than their counterpart speakers in the Netherlands. Our findings support the view that heritage language research can help in the historical reconstruction of “protolanguages”.

In this paper, we start out from a reduced data set, i.e. confined to the modern varieties in Brazil and the Netherlands only. This implies that in first instance we ignore the available historical data. In this way we want to investigate the extent to which rusty speaker data alone can contribute to the reconstruction of the old language spoken by the ancestors in the time of emigration, which – under the assumption of lacking historical data – can be called the “protolanguage”. However, we confront our findings with the available historical data in the end, which forces us to reconsider some of our conclusions.

## 2 Zeelandic Flemish in Brazil: a case of a transplanted dialect

### 2.1 Historical background

In the 19<sup>th</sup> century an association named *Associação Central de Colonização* was established by the Brazilian imperial government. The goal of this association was twofold: on the one hand, they wanted to recruit European immigrants in order to have more manpower for the cultivation of agricultural land after the abolition of slavery, and on the other hand they wanted to attract more “civilised whites” to the country (Roos & Eshuis 2008: 11). To this purpose, leaflets with promises of a better life were distributed in port cities of Europe. Thousands of fortune seekers from different European countries were persuaded in this way to

migrate to Brazil.<sup>1</sup> Among those immigrants, there were also 323 Dutchmen who settled in the state of Espírito Santo. Within Espírito Santo, there were two destinations: the colony of Rio Novo in the South and the colony of Santa Leopoldina in the interior of Vitória, the latter of which is of interest to this paper. The first Dutch immigrants arrived in the Santa Leopoldina colony in 1858, but new immigrants were arriving each year, up until 1862 (Roos & Eshuis 2008: 50, 121). In total, 243 Dutch immigrants settled in Santa Leopoldina.

Upon arrival, the Dutch immigrants were immediately confronted with many difficulties and deprivations: they had to survive in dense forest, unused to the heat, on infertile land, without the equipment or money to rebuild their lives, short of food and without any assistance. Furthermore, the little that the immigrants were able to cultivate was in the possession of a colonel – also a Zeelandic Flemish immigrant – who controlled the planting and the harvesting. The immigrants sold their products at the colonel's sale house (the *venda*) for a meager price, but had to buy what they needed (e.g. salt) at the same *venda* for exorbitant prices. Because of a negative report about the colony in 1862, the Brazilian imperial government ceased to offer any support (Von Tschudi 2004). This resulted in total isolation and abandonment of the Dutch immigrants in Santa Leopoldina. Because of this situation, the Dutch immigrants did not integrate with other groups, which contributed to the maintenance of customs, such as the preservation of their dialect and religion (i.e. Calvinism) (Buysse 1984, Roos & Eshuis 2008). Another factor that contributed to the preservation of the Zeelandic Flemish language throughout the 19<sup>th</sup> and 20<sup>th</sup> century was the fact that only few members of the community had attended any school (Schaffel 2010: 69).

As time went by, the descendants of the Zeelandic Flemish immigrants got intermingled more and more with another group of immigrants, that is, the Pomeranian immigrants, mainly because of two reasons: because the Zeelandic Flemish community never attained their own Calvinist churches in Brazil – apart from a small chapel in Holandinha – they were mainly forced to go to church with the Lutheran Pomeranians, and because the total number of Zeelandic Flemish immigrants was rather small, they sometimes had to marry members from outside the community, which resulted in a growing number of Zeelandic Flemish-Pomeranian marriages.

---

<sup>1</sup>Von Tschudi (2004) gives the following numbers for Santa Leopoldina in 1860: total number of colonists: 1,003 (232 heads of family), of which: Swiss (104), Hannover (4), Luxembourg (70), Prussia (384), Bavaria (10), Baden (27), Hessen (61), Tirol (82), Nassau (13), Holstein (13), Mecklenburg (5), Saxonia (76), Belgium (8), Holland (126), France (1), England (1), and some Brazilians. The Prussian immigrants probably consisted mainly of Pomeranians. Initially, this latter group was only twice as large as the Dutch and Belgian immigrants together.

While the transplanted Zeelandic Flemish dialect in Brazil has been subject to internal changes in the last decades due to the multilingual setting and regular processes of language obsolescence (see Sections 2.2 and 3, see Schaffel Bremenkamp et al. 2017), the Zeelandic Flemish dialects as spoken in the motherland were modified by processes of dialect loss and convergence to the Northern Dutch standard language. Consequently, these two varieties increasingly diverged from one other.

## **2.2 Sociolinguistic situation**

Sociolinguistic research into the Zeelandic Flemish dialect spoken in Espírito Santo (Schaffel 2010) revealed that the language is currently spoken by just 13 people.<sup>2</sup> These people, who are all descendants of the Zeelandic Flemish immigrants of the nineteenth century, speak the language with varying levels of proficiency and are of different ages, though the majority is more than 60 years old.<sup>3</sup> Among the younger descendants, there are just a few who can understand the Zeelandic Flemish dialect or who speak a few words.<sup>4</sup>

Most of these 13 speakers state that they do not use the Zeelandic Flemish language on a regular basis. This is very likely induced by the fact that they live geographically dispersed in the old colony of Santa Leopoldina and have not much contact with each other. Next to this, the fact that the group of immigrants was small,<sup>5</sup> that these immigrants were mainly forced to go to church with the Lutheran Pomeranians, that they were abandoned by their motherland rather soon after migration and finally, the highly frequent occurrence of exogamous marriages (especially with Pomeranians), are considered to be the most important factors in the disappearance of Zeelandic Flemish in Espírito Santo (Schaffel 2010: 83–85). Since the Pomeranian immigrants were more numerous,<sup>6</sup> and since they could practise their own Lutheran religion, their cultural values and their language were much better preserved (see also Postma 2014, 2019).<sup>7</sup> As a con-

---

<sup>2</sup>In 2015 five more speakers were identified.

<sup>3</sup>To be precise, only 1 speaker is between 20 and 39 years old, 5 speakers are between 40 and 60 years old, and 7 speakers are older than 60.

<sup>4</sup>See Schaffel Bremenkamp et al. (2017) for a more elaborate discussion on Schaffel's (2010) sociolinguistic findings.

<sup>5</sup>According to Roos & Eshuis (2008: 50, 121) 243 people migrated from West Zeelandic Flanders to Espírito Santo between 1859 and 1862.

<sup>6</sup>Of the group of 3933 German immigrants in Espírito Santo, about 2000 were Pomeranian.

<sup>7</sup>This is, among other things, reflected in the fact that there is a Pomeranian language radio program, as well as a dictionary of Brazilian Pomeranian (Tressmann 2006a) and a collection of tales (Tressmann 2006b).

sequence, the most prevalent home language in exogamous families is Pomeranian, not the Zeelandic Flemish dialect.<sup>8</sup> This implies that Zeelandic Flemish is no longer transmitted to the next generation. Cessation in the intergenerational transmission of a language inevitably leads to the gradual loss of that language (Sasse 1992).

Almost none of the contemporary descendants of the Zeelandic Flemish immigrants living in Espírito Santo have Zeelandic Flemish as their mother tongue. Schaffel (2010: 77) found that only eight informants classified it as their only mother tongue, seven of whom were older than 60 years old. A further two informants specified Zeelandic Flemish and Portuguese as their mother tongues, and eight informants mentioned Zeelandic Flemish and Pomeranian as their mother tongues.

The linguistic situation among the Zeelandic Flemish descendants in Espírito Santo is trilingual. Portuguese is the national language that is used in official organisations and in education. Pomeranian and Zeelandic Flemish are both transplanted languages that were taken to Brazil by immigrants in the 19<sup>th</sup> century. As stated above, the Pomeranian language was almost always preferred as the home language in the numerous exogamous families. This situation has led to a gradual shift of the Zeelandic Flemish community to the dominant languages Portuguese and Pomeranian.<sup>9</sup> The inevitable outcome of this situation for the Zeelandic Flemish language in Brazil is language death.

### **3 Gradual language death and different types of speakers**

The last speakers of Zeelandic Flemish in Espírito Santo have not passed on the language to their children, so they can be considered as so-called “terminal speakers” (Sasse 1992) of a moribund language. As argued by Dressler (1996: 195), bilingual or multilingual speech communities are the ideal breeding ground for a situation of gradual language death, in which the minority language community shifts to the dominant language(s) of the majority. In our case, Zeelandic Flemish speakers have gradually shifted to Portuguese and/or Pomeranian. The outcome of this “language shift” – a notion focusing on the speech community rather than on the language (Rottet 1995: 5) – for the receding language is “language

---

<sup>8</sup>The Zeelandic Flemish language has never been recorded in publications of any kind.

<sup>9</sup>Whether the Zeelandic Flemish descendants have shifted to Portuguese or Pomeranian depends on the region they live in: a large majority shifted to Portuguese in (what is today called) Santa Leopoldina, whereas Pomeranian is the dominant language in Santa Maria de Jetibá and Itarana (see Schaffel 2010, Schaffel Bremenkamp et al. 2017).

loss” (Appel & Muysken 1987, Fase et al. 1992) or “language decay” (Field 1985, Sasse 1991) – notions that focus on the internal linguistic changes that the moribund language often undergoes. The final outcome of these internal changes and of the shift to the dominant language is called “language death” (Dressler 1972, Dorian 1977).

The internal changes affecting a declining language are not always distinguishable from changes affecting languages that are involved in “normal” contact situations. However, it is often a combination of various processes that contributes to language decay (Campbell & Muntzel 1989: 188). Some characteristic structural changes in dying languages are for example borrowing (resulting in loanwords and loan constructions or calques), reduction in syntagmatic redundancy and inflectional morphology, replacement of synthetic with analytic grammatical structures, reduction of stylistic variation, extreme phonological variation, extensive code switching, and so forth. In Schaffel Bremenkamp et al. (2017: 453–465) we showed that most of these linguistic characteristics actually occur in the language of the Zeelandic Flemish descendants in Espírito Santo. It is typical of language death situations that these internal changes progress rapidly.<sup>10</sup>

In this contribution, however, the focus is not on the internal changes that have affected the Zeelandic Flemish language of the last speakers, but rather on the archaic features that have remained unchanged in the language of so-called “rusty speakers”, a term coined by Sasse (1992). Sasse makes the distinction between two types of imperfect speakers of a dying language, that is, “rusty speakers” versus “semi-speakers”. He defines rusty speakers as “former fluent speakers who were on their way to becoming full speakers, but never reached that degree of competence due to the lack of regular communication in the language” (Sasse 1992: 62). He considers rusty speakers as a special type of L1 learners, who have “a comparably good proficiency in the grammatical system of the language and a perfect passive knowledge”, but who “suffered from severe memory gaps, especially in vocabulary, but also in more complicated areas of the grammatical system” (Sasse 1992: 61). The imperfect language of a rusty speaker is the result of “later loss”. On the other hand, there are also true semi-speakers, whose command of the dying language is, as argued by Sasse (1992: 61), from the beginning “imperfect to a pathological degree”. Because of the interruption in the transmission of the language, the semi-speaker acquires the language incompletely, and not “by way of normal acquisition processes (i.e. parent-to-child, by means

---

<sup>10</sup>This effect, in which a transplanted language changes more rapidly than the motherland language, due to fading linguistic norms and the influence of other dominant languages among others, is discussed by Rosenberg (2005) in his work on German language islands in Brazil and Russia.

of conscious language transmission strategies on the part of the parents), but rather “by chance”, by interacting more than usual with elderly members of the community” (Rottet 1995: 36–37). Semi-speakers should therefore be considered as L2 speakers. In practice, however, it is often difficult to categorise strictly between either rusty speakers or semi-speakers, because, as Sasse points out, there is a language proficiency continuum between individuals who learned the dying language by chance and those who learned it in more regular ways.

Regardless of the question of how speakers acquired the moribund language, it is characteristic of situations of language death that linguistic norms are broken down, due to the fact that none of the speakers are regarded by the language community as “local authorities on language questions” (Rottet 1995: 39). As a result, there is a “relaxation of internal monitoring” (Dorian 1981: 154): language community members come to accept and tolerate a gamut of uses of linguistic features that would have been regarded as mistakes in earlier times. This huge amount of variation in the use of a declining language by different speakers is one of the reasons why researchers might wonder “whether data from a dying language can reliably be used by linguists for other purposes, e.g. by historical linguists for purposes of reconstruction of protolanguages” (Rottet 1995: 3). In this paper we argue that linguistic data from rusty speakers of Zeelandic Flemish in Espírito Santo can indeed be used to reconstruct the “protolanguage” of the original immigrants.<sup>11</sup> Since rusty speakers do no longer have many opportunities to speak their mother language, it does not get mixed up so strongly with the dominant language(s) and remains relatively “authentic”, i.e. close to the original variety of the ancestors. This variety is not similar to the Zeelandic Flemish dialect as it is spoken in the motherland nowadays, since the latter variety has been subject to processes of dialect loss under the influence of northern Standard Dutch.<sup>12</sup> We discuss three linguistic phenomena that occur in the language of these rusty speakers: (1) deletion of /l/ in codas and coda clusters with concomitant compensatory lengthening of the preceding vowel, (2) subject doubling and (3) inflected polarity markers *yes* and *no*. All three phenomena still occur in

---

<sup>11</sup>The term “protolanguage”, borrowed from comparative linguistics, may not be completely appropriate for the situation at hand, because there is ample historical data available for the linguistic situation in the days of emigration. However, we use this term in the context of the methodological approach of this paper (see Introduction), in which we start out from a reduced data set, i.e. confined to the modern varieties in Brazil and the Netherlands only (and thus ignoring any historical data), in order to investigate the extent to which rusty speaker data alone can help in reconstructing the old language as spoken in the days of emigration.

<sup>12</sup>Northern Standard Dutch refers to Standard Dutch as it is spoken in the Netherlands, as opposed to southern Standard Dutch, which is the standard language spoken in the Dutch-speaking part of Belgium, i.e. Flanders.

some Belgian Flemish dialects, but have disappeared from the contemporary variety as spoken in the Dutch province of Zeelandic Flanders, which has strongly converged to northern Standard Dutch. Our findings demonstrate the potential historical value of transplanted dialects or speech island varieties (Rosenberg 2005), in that such a variety may – as in the case of the rusty speakers of Brazilian Zeelandic Flemish – resemble the original immigrant (proto-)language more closely than is the case for the homeland variety.

## 4 Methodology

The linguistic data that will be discussed in the next section are based on spoken language material that was collected in 2012 and 2013, by Gertjan Postma (Meertens Institute), Andrew Nevins (Federal University of Rio de Janeiro and University College London) and both authors of this paper. Interviews were taken from nine (five male and four female) speakers of Zeelandic Flemish, living in Holandinha (part of Holanda), Garrafão, Alto Jatibocas, Caramuru and Alto Jetibá (see Figure 1).

All of these speakers were older than 40. Recordings were made at people's homes, where two or more speakers were present and talked with each other and with the interviewers. Three of the four interviewers spoke Portuguese with the informants, one also spoke northern Standard Dutch, one of them was a native speaker of Pomeranian, and another interviewer was a native speaker of a Belgian Flemish dialect which closely resembles the Zeelandic Flemish dialects of the places the immigrants originally came from. The interviews were a mixture of these four languages with a lot of code switching in the informants' speech, but with the general aim of eliciting as much Zeelandic Flemish dialect as possible. Topics that were talked about during the interviews were the informants' past, their language, their religion, the work on the land, and so forth. The interviews were audio- and video-recorded and cover more than four hours of speech, only part of which is in Zeelandic Flemish. The recordings are preserved at the Meertens Institute (Amsterdam). Next to this, one interview made by Arjan Van Westen and Monique Schoutsen for their movie *Braziliaanse Koorts* ("Brazilian Fever") was also used for the data discussed in this paper. Large parts of the recordings have been transcribed, glossed and annotated, with segmentation in Praat by Gertjan Postma, Kathy Rys and Lea Busweiler. The linguistic data discussed in the next section were selected from these transcripts. These data come from four rusty speakers: they are all speakers who used to speak more Zeelandic Flemish in the past, but who lost the opportunity to speak it on a daily



Figure 1: Locations of the speakers of Zeelandic Flemish in Espírito Santo, Brazil  
 (c) OpenStreetMaps contributors

basis from the 1990s onwards (Schaffel 2010). The four speakers concerned have the following profiles:

- speaker 1, named A., is an 84-year-old man from Holandinha, whose mother tongue is Zeelandic Flemish and who also speaks Portuguese, which he acquired at school, but no sooner than at the age of six; his ancestors came from the West-Zeelandic village of Nieuwvliet;
- speaker 2, named J., a male informant living in Holandinha, whose mother tongue is Zeelandic Flemish and who also speaks Portuguese, which he acquired at school; his ancestors came from the West-Zeelandic village of Retranchement;
- speaker 3, named R., is a 65-year-old man from Alto Jatibocas, whose mother tongue is Zeelandic Flemish, who is married to a Pomeranian woman, and who speaks Zeelandic Flemish, Pomeranian, Hochdeutsch and Portuguese, the latter of which he started to acquire at school at the age of six;

- speaker 4, named B., is R.'s sister, age unknown, a woman from Alto Jatibocas, whose mother tongue is also Zeelandic Flemish,<sup>13</sup> who is married to a Pomeranian man who is also able to speak the Zeelandic Flemish language because his mother was Zeelandic Flemish; like R., B. speaks four languages: Zeelandic Flemish, Pomeranian, Hochdeutsch and Portuguese;
- all four speakers have in common that they did not go to school for more than four years.

The places where these speakers lived belong to different language areas. Hollandinha belongs to the Portuguese area, i.e. the area where a lot of descendants of the former slaves live. In this area Portuguese is the only dominant language and there is not much interference of Portuguese with Zeelandic Flemish (Schaffel Bremenkamp et al. 2017). Alto Jatibocas belongs to the Pomeranian area, where the Zeelandic Flemish people have often married Pomeranian partners and where the Pomeranian language is the main language used in families, which has resulted in a lot of interference between Pomeranian and Zeelandic Flemish.

## 5 Results

In this section we argue that the speech of rusty speakers contains certain features that deviate from the current motherland variety, but that cannot be attributed to language contact. Instead, these features are remnants of the protolanguage, that was once also spoken in the motherland, and that was transplanted to Brazil by the immigrants of the 19<sup>th</sup> century. These protolanguage features allow us to reconstruct the original language of the immigrants and of the motherland. In this way, the speech of so-called “terminal” speakers may be of historical value. We give evidence supporting this claim by discussing three linguistic phenomena: (1) the deletion of /l/ in codas and coda clusters with concomitant compensatory lengthening of the preceding vowel, (2) subject doubling, and (3) the inflected polarity markers *yes* and *no*.

---

<sup>13</sup>R. and B.'s mother only spoke Zeelandic Flemish with her children when they were little.

### 5.1 Deletion of /l/ with concomitant compensatory lengthening of the preceding vowel

The speech of all four rusty speakers, contained a phonological feature that is illustrated in the underlined parts of the example sentences (1) to (6):<sup>14</sup>

- (1) speaker 1 (A.)

Dat is al twee of drie keer vermaakt hé  
[ɑ:]

'that has already been repaired two or three times you know'

- (2) speaker 1 (A.)

Den eersten a' j' ier stoeng, das al al kapot  
[ɑ:]

'the first one that was standing here, that's already completely broken'

- (3) speaker 3 (R.)

Dan is't zo, ielk brieng wat ja  
[i:k]

'then it is like this, everybody brings something you know'

- (4) speaker 3 (R.)

Koeien zo kalvers  
['ka:vərs]

'cows, you know, calves'

- (5) speaker 4 (B.)

Op 't land gewèèrkt, ja alles...koffie, koeien, koeien melken  
['mæ:?:?ə]

'worked on the land, yes everything...coffee, cows, milking cows'

- (6) speaker 4 (B.)

Toe 's navonds eh... achte... half negene  
[a:f]

'until in the evening eh... eight o'clock...half past eight'

The underlined words are realised with a deleted /l/ and with concomitant compensatory lengthening of the preceding vowel, resulting in an overlong vowel.

---

<sup>14</sup>The phonetic realisation of the underlined words is given with each sentence, as well as the initials and informant number of the speakers involved.

The /l/ that is deleted is always a so-called “dark /l/” (phonetically transcribed as [ɫ]). In Dutch, /l/ is realised differently according to its position in the syllable: clear [l], realised as a lateral approximant, in the onset of a syllable and dark [ɫ], realised as a pharyngealised approximant, in the coda before a pause or another consonant. Vocalisation of dark /l/ occurs in some varieties and some speakers of Dutch. A word like *geel* ‘yellow’, with the underlying form /yel/, is then realised as [yew] (Van Reenen 1986, Botma & van der Torre 2000: 17). The complete deletion of prepausal and preconsonantal /l/, however, is not a common feature in varieties of Dutch.<sup>15</sup> Nevertheless, there is one variety in the Dutch-speaking area which is characterised by the categorical deletion of prepausal and preconsonantal /l/ and by a very strong compensatory lengthening of the preceding vowel. It is a Flemish dialect spoken in the village of Maldegem, which belongs to the Dutch-speaking part of Belgium (represented by the blue dot in Figure 2).

The phonological process of /l/-deletion in the Maldegem dialect occurs in codas and coda clusters of stressed syllables<sup>16</sup> and can be represented as in (7) (see Rys 2007: 182–186):

$$(7) \quad \left[ \begin{array}{c} +\text{STRESSED} \\ +\text{LATERAL} \end{array} \right] \rightarrow \emptyset / -\left\{ \begin{array}{c} C \\ \#C \end{array} \right\}$$

The representation in (7) indicates that /l/ is deleted before a consonant or a pause, but only if that pause is followed by a consonant. This is illustrated in (7).

- (7)    a. de bal\_ pakken [dəm'ba:.pa?̩]
- ‘take the ball’
- b. de bal\_ is...[dəm'bal.ɛs]
- ‘the ball is...’

The examples (1)–(6) of our Brazilian informants are instances of exactly the process described in (7), that is, of preconsonantal or prepausal /l/-deletion with compensatory lengthening of the preceding vowel, resulting in an overlong vowel. The occurrence of a phonological feature in the speech of our rusty speakers that is also found in a current Flemish dialect, suggests that this feature might have been brought along with the original immigrants and thus might be a feature that also occurs in the dialects spoken in the villages of the motherland where

<sup>15</sup>Cross-linguistically, /l/-vocalisation and /or/-deletion occurs in certain dialects of English, Old French, Korean, and Swiss German. In Hungarian, we find complete deletion of prepausal and preconsonantal /l/ and compensatory lengthening of the preceding vowel, comparable to the examples (1)–(6) (Feyér et al. 2012).

<sup>16</sup>In a word like *appel* ‘apple’, which has stress on the first syllable, /l/ is not deleted.



Figure 2: Zeelandic Flemish villages (in the Netherlands) where the immigrants came from and the East Flemish village of Maldegem (in Belgium). Source: Schaffel Bremenkamp et al. (2017: 438)

these immigrants came from (see 2). However, there are only scant indications of the occurrence of /l/-deletion in the current dialects of West Zeelandic Flanders.

Van den Broecke-de Man (1978) mentions /l/ as one of the omitted consonants in her monograph on the dialects of West Zeelandic Flanders. She argues that /l/ is deleted at the end of a word and illustrates this with the examples “à, wè, zà, wìje?”, which is her own representation of regular Dutch orthography *al* ‘already’, *wel* ‘surely’, *zal* ‘shall’ and *wil je?* ‘do you want?’ (lit. want-you) (van den Broecke-de Man 1978: 9).<sup>17</sup> It strikes us that the examples that are given are either frequent

<sup>17</sup>This author uses the ‘-symbol (e.g. à, è, i) to represent vowels that she calls “iets gerekt” (‘slightly prolonged’) (van den Broecke-de Man 1978: 9). It is unclear whether her representation indicates the overlong vowel quality that is typical of the Maldegem dialect.

function words like *al* ‘already’ and *wel* ‘surely’, or frequent verbs like *zal* ‘shall’ and *wil* ‘want’. van den Broecke-de Man (1978: 20–25) describes some grammatical aspects of the dialect of the village of Eede<sup>18</sup> in a separate chapter because of the exceptional character of this dialect, viz. the many grammatical similarities with the Maldegem dialect, which is spoken right across the border. Unlike the other West Zeelandic Flemish dialects, the Eede dialect is characterised not only by prepausal /l/-deletion (i.e. in the coda), but also by preconsonantal /l/-deletion (i.e. in coda clusters, e.g. *melk* → [mæ:k]) (see Rys 1999, 2000: 352). Nevertheless, among the 504 Zeelandic Flemish people who migrated to Espírito Santo between 1858 and 1862 there were no inhabitants of Eede (Roos & Eshuis 2008: 12).

According to Taeldeman (1979: 163), however, /l/-deletion with compensatory vowel lengthening only occurs in the dialect of Eede, in preconsonantal as well as prepausal contexts,<sup>19</sup> whereas the other Zeelandic Flemish dialects display slight lengthening of vowels preceding /l/ in coda clusters (especially with alveolar consonants), but no deletion of /l/.

In order to check whether /l/-deletion occurs in the dialects of the villages the immigrants departed from, we consulted sound recordings of spoken dialogues that were made in the 1960s and 1970s by researchers of the university of Ghent and the Meertens Institute (Amsterdam). These recordings and the broad transcriptions of it are both part of a database called *Stemmen uit het verleden* ('Voices from the past')<sup>20</sup> as well as a database called *Nederlandse Dialectenbank* ('Dutch Dialect Database').<sup>21</sup> We studied the transcriptions of recordings from several West Zeelandic Flemish places.<sup>22</sup> In these transcriptions, we only came across instances of prepausal /l/-deletion in the following words: *veel* ‘much’ (also *zoveel* ‘this much’, *hoeveel* ‘how much’), *wel* ‘PART’, *zal* ‘shall’, *al* ‘already’, and *nogal* ‘quite’. The process of /l/-deletion is thus restricted to a very small set of words, which implies that it is a lexically determined process.<sup>23</sup> This is in contrast to the process of /l/-deletion in the dialects of Maldegem and Eede, where it is applied (nearly)<sup>24</sup> categorically (Versieck 1989, Rys 1999). In addition, the transcriptions

<sup>18</sup>Eede does not appear on Figure 2. It should be located north of the village of Maldegem, just above the border between the Netherlands (which Zeelandic Flanders is part of) and Belgium.

<sup>19</sup>This is confirmed in Rys (1999; 2000).

<sup>20</sup><https://www.variaties.be/portfolio-item/stemmen-uit-het-verleden/>

<sup>21</sup><https://www.meertens.knaw.nl/ndb/>

<sup>22</sup>These places were Aardenburg, Biervliet, Breskens, Cadzand, Groede, Hoofdplaat, IJzendijke, Nieuwvliet, Retranchement, Schoondijke, Sint Kruis, Sluis, Waterlandkerkje, and Zuidzande. For all of these places, one transcribed recording was available. Together, these recordings covered more than nine hours of speech.

<sup>23</sup>The transcriptions do not allow us to conclude whether these cases of /l/-deletion are accompanied by a strong lengthening of the preceding vowel. There are indications that the vowels are only slightly lengthened, in agreement with van den Broecke-de Man (1978).

<sup>24</sup>In Eede it does not apply in all cases (see below).

display a lot of realisations of these words without /l/-deletion.<sup>25</sup> Moreover, all of these instances are examples of prepausal /l/-deletion. Examples of preconsonantal /l/-deletion, as are produced by the Brazilian rusty speakers in the examples (3)-(6), do not occur in these recordings of West Zeelandic Flemish dialects, except for one case: one informant from Hoofdplaat is talking about the North Sea flood of 1953, when he produces the following example of preconsonantal /l/-deletion:

- (8) en me zien de golven zò mà over de diek nar ons toe komm'n<sup>26</sup>  
 [ho:vən]

'and we see the waves suddenly come to us over the dike'

Altogether, we can say that the West Zeelandic Flemish dialects of the 1960s-1970s are not characterised by a categorical process of /l/-deletion. Rather, it seems that /l/-deletion is a lexically determined process that is restricted to a small set of words and that is only applied in prepausal contexts. The occurrence of one instance of preconsonantal /l/-deletion might be a remnant of a process that was more common and widespread in the past.

A more recent source of information on the phonological characteristics of the current West Zeelandic Flemish dialects is the *Fonologische Atlas van de Nederlandse Dialecten*, which is abbreviated as *FAND*, 'Phonological Atlas of Dutch Dialects' (De Wulf et al. 2005). This publication is based on a questionnaire that was conducted in 578 places in the Netherlands and the Dutch-speaking part of Belgium. The *FAND* comprises 496 maps showing the most common pronunciations of particular words in different places. Only three Zeelandic Flemish places relevant to our study (i.e. places of origin of the Zeelandic Flemish immigrants) are represented on these maps, viz. Breskens, Zuidzande and IJzendijke<sup>27</sup> *FAND*-research, but it is not represented in Figures 3 and 4. (encircled with red in Figures 3 and 4).

Figure 3 shows the phonetic realisations of the word *schuld* 'guilt' in different places in the Dutch-speaking areas (viz. the Netherlands and Flanders, Belgium). Thus, this map illustrates the process of preconsonantal /l/-deletion. Prepausal /l/-deletion is illustrated in Figure 4, which represents the phonetic realisations of the word *vol* 'full'. The hollow circles in Figures 3 and 4 represent complete

<sup>25</sup>In the transcription of the Nieuwvliet dialect, for example, there are 13 cases of /l/-deletion (in the words *wel*, *veel*, and *nogal*), but also 13 cases without /l/-deletion (in the words *al*, *wel*).

<sup>26</sup>The transcription as an overlong vowel is our own interpretation, since the original (broad) transcription is not explicit about the vowel length.

<sup>27</sup>Eede is another West Zeelandic Flemish place that was included in the

deletion of /l/ in front of a consonant (Figure 3) or a pause (Figure 4), rendering the phonetic realisations [sxɛ:t] and [vɛ:],<sup>28</sup> respectively. This only occurs (almost) categorically in two East Flemish (Belgian) places, viz. Kleit and Middelburg (encircled with blue in Figures 3 and 4), which are both submunicipalities of Maldegem.<sup>29,30</sup> The word *schuld* in the dialect of Nukerke, which is in the southwest of the province of East-Flanders, as well as in some dialects in the deep south of the Dutch province of Limburg; and Figure 4 shows complete deletion of /l/ in the word *vol* in the dialect of Nukerke and the neighbouring dialect of Ronse. Some other maps in the *FAND* show a more widespread deletion of /l/: the map of *wolf* ‘wolf’, for example, shows deletion of /l/ in some West- and East Flemish places (e.g. in Nieuwpoort, Wervik, Nevele, Gent). However, Kleit is the only place where /l/-deletion applies categorically (in the dialect of Middelburg it applies nearly categorically). This was found by investigating 39 nouns and adjectives containing /l/ in codas or coda clusters that were included in the GTRP-database (<https://www.meertens.knaw.nl/projecten/mand/GTRPdataperitem.html>) on which the *Morfologische Atlas van de Nederlandse Dialecten* (abbreviated: *MAND*, ‘Morphological Atlas of Dutch Dialects’; De Schutter 2005) was based. In addition, the other West- or East Flemish places where /l/-deletion occasionally occurs do not constitute a concatenated area with the West Zeelandic Flemish places the immigrants originated from, unlike Kleit and Middelburg, which do make up a connected area with these places. None of the Zeelandic Flemish places where the immigrants came from<sup>31</sup> display deletion of /l/. The hollow square in Breskens (Figures 3 and 4), represents a velar approximant that tends to vocalisation, but without complete deletion. So, in recent sources like the *FAND* (2005) there are no indications of preconsonantal or prepausal /l/-deletion in the current West Zeelandic Flemish dialects.

Next to the data provided in the *FAND*, the *GTRP*-database<sup>32</sup>, which constitutes the source material of the *MAND* (De Schutter 2005), also provides information on the pronunciation of lexemes in the dialects of the Dutch-speaking area. In order to investigate whether or not /l/-deletion applies categorically, we looked up the pronunciation of 39 lexemes containing /l/ in a coda or coda clus-

<sup>28</sup>The dialects of Kleit and Middelburg (both belonging to the municipality of Maldegem) are characterised by unrounding of the rounded vowels /y/ (in *schuld*) and /ɔ/ (in *vol*).

<sup>29</sup>The dialect of the main municipality, which we could call “Maldegem-centre”, was not included in the data compiled for the *FAND*.

<sup>30</sup>Figure 3 also shows complete deletion of /l/ (symbolised by a hollow circle) in the word.

<sup>31</sup>Encircled with red in Figures 3 and 4.

<sup>32</sup><https://www.meertens.knaw.nl/projecten/mand/GTRPdataperitem.html>

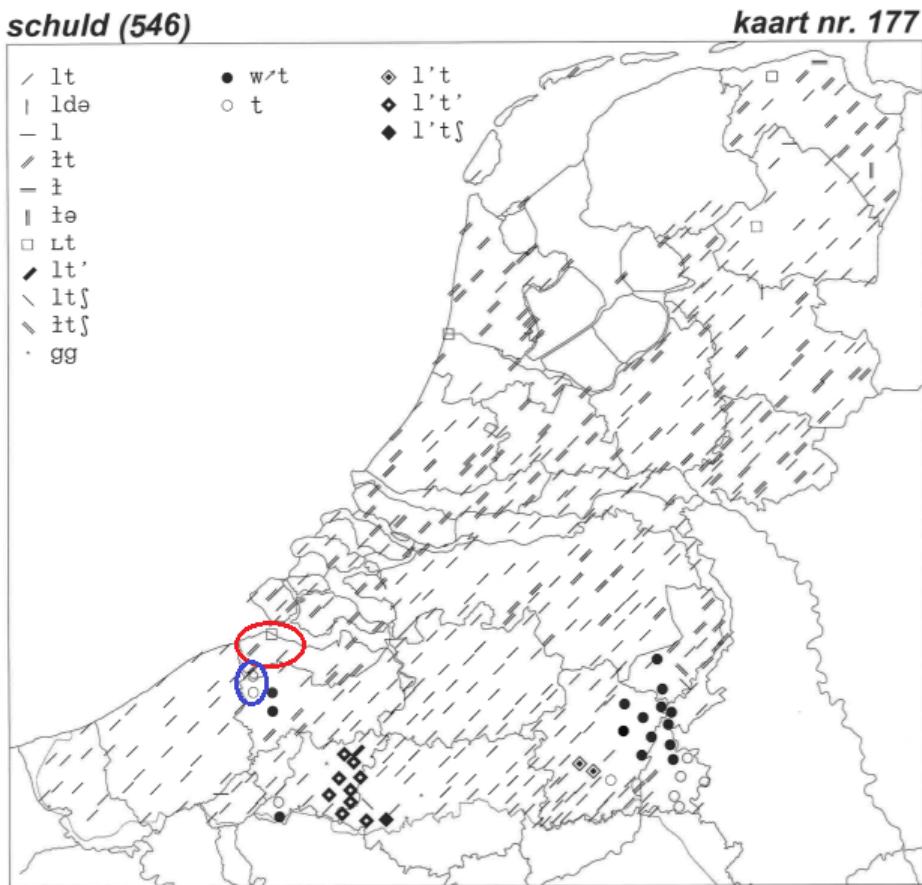


Figure 3: Phonetic realisations of the word *schuld* ‘guilt’ in dialects of the Dutch-speaking area. (Source: De Wulf et al. 2005)

ter<sup>33</sup> in the various dialects of the Dutch-speaking area using this database. As in the *FAND*, Kleit and Middelburg belong to the places that were included in this research, whereas Maldegem-centre does not. Kleit is the only place where complete deletion of /l/ and compensatory lengthening of the preceding vowel

<sup>33</sup>The 39 lexemes under examination were: *balk* ‘beam’, *beeld* ‘statue’, *bril* ‘spectacles’, *deel* ‘part’, *dweil* ‘floor-cloth’, *geld* ‘money’, *heft* ‘half (noun)’, *hol* ‘hole’, *kalf* ‘calf’, *kelder* ‘cellar’, *melk* ‘milk’, *naald* ‘needle’, *pols* ‘wrist’, *schelp* ‘shell’, *schuld* ‘debt, guilt’, *spel* ‘game’, *stal* ‘stable’, *steel* ‘stalk, handle’, *stoel* ‘chair’, *uil* ‘owl’, *volk* ‘people’, *wolf* ‘wolf’, *wolk* ‘cloud’, *zalf* ‘ointment’, *zeil* ‘sail’, *dol* ‘silly’, *zolder* ‘attic’, *fel* ‘fierce’, *geel* ‘yellow’, *half* ‘half (adj.)’, *heel* ‘whole’, *kalm* ‘calm’, *scheel* ‘cross-eyed’, *smal* ‘narrow’, *vals* ‘false’, *vol* ‘full’, *vuil* ‘dirty’, *wild* ‘wild’, *zilveren* ‘silver’.

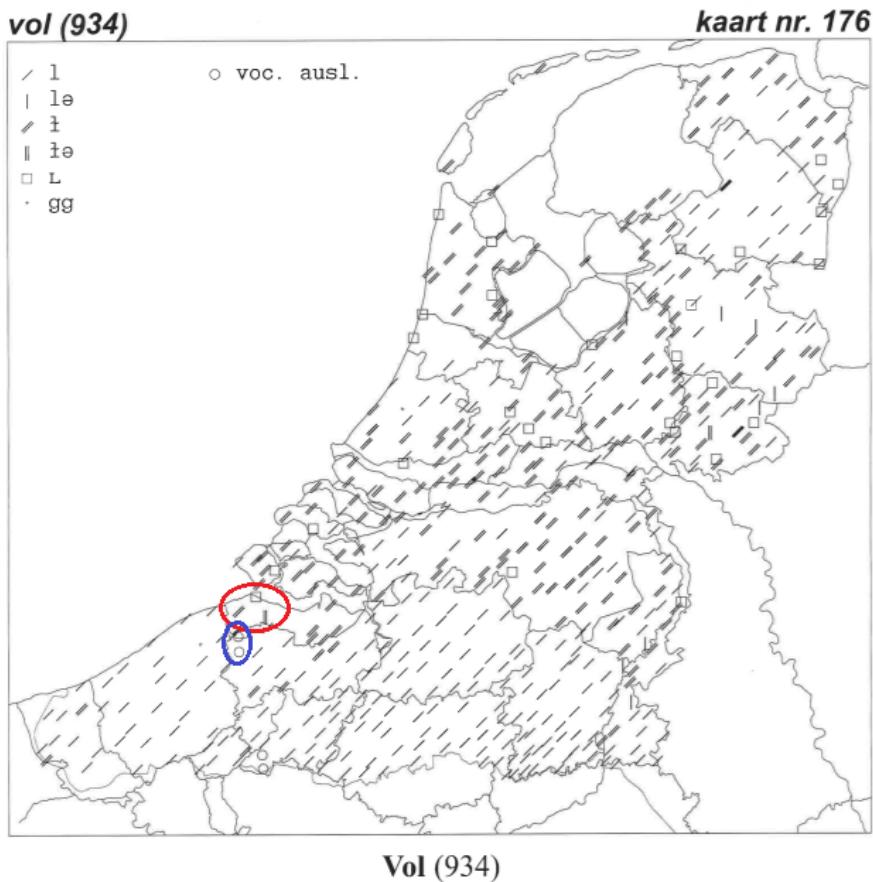


Figure 4: Phonetic realisations of the word *vol* ‘full’ in dialects of the Dutch-speaking area. (Source: De Wulf et al. 2005)

takes place in all 39 lexemes. The dialect of Middelburg does not have categorical deletion: for some lexemes it joins the neighbouring West Flemish dialects, in which case /l/ is not deleted (e.g. in *balk* ‘beam’ > [balk]). In other cases /l/ is deleted but there is no compensatory lengthening of the preceding vowel (e.g. in *heft* ‘half’ > [æft]), and in several cases /l/ is vocalised (e.g. *kelder* ‘attic’ > [kæjdərə]). In addition, the *GTRP*-database offered the opportunity to investigate whether deletion of /l/ occurs in the West Zeelandic Flemish places which the emigrants once departed from.<sup>34</sup> We found one attestation of (preconsonantal) /l/-deletion for the dialect of IJzendijke, more particularly in the word *beeld*

<sup>34</sup>These are the same places as in the *FAND*, i.e. Breskens, Zuidzande, IJzendijke and Eede.

‘statue’ (pronounced as [be:t]). The dialect of Eede, which – as was mentioned above – has some exceptional grammatical features with respect to the other West Zeelandic Flemish dialects (van den Broecke-de Man 1978: 20–25, Rys 1999) displays complete deletion of /l/ in 22 out of 39 lexemes,<sup>35</sup> in prepausal as well as preconsonantal contexts. To conclude, the *GTRP*-database indicates more or less the same results as the sources discussed above (i.e. van den Broecke-de Man 1978, Taeldeman 1979, De Wulf et al. 2005, *Nederlandse Dialectenbank*): except for Eede and one instance in IJzendijke, /l/-deletion does not seem to occur in contemporary dialects of West Zeelandic Flanders.

As the examples (1)–(5) show, however, we find relatively many instances of prepausal as well as preconsonantal /l/-deletion in the speech of rusty speakers of Zeelandic Flemish living in Espírito Santo. In one interview with speaker 1 (A.), /l/ deletes in 8 out of 14 cases, that is, in 57% of possible cases. In one recording with speaker 3 (R.), we find /l/-deletion in 4 out of 11 cases (36%). In speaker 4 (B.)’s speech, there are at least 8 clear cases of /l/-deletion.<sup>36</sup>

To summarise, rusty speakers of Brazilian Zeelandic Flemish have prepausal as well as preconsonantal /l/-deletion in their speech, whereas this phenomenon is only infrequently observed in Zeelandic Flemish dialects as spoken in the 1960s–1970s (van den Broecke-de Man 1978, *Nederlandse Dialectenbank*) and – except from Eede – hardly observed at all in contemporary Zeelandic Flemish dialects (*FAND*, *GTRP*). These findings suggest that /l/-deletion has disappeared from contemporary varieties, but was a common and more widespread feature of Zeelandic Flemish dialects in the past, and thus, of the protolanguage that was spoken by the original Zeelandic Flemish migrants to Brazil.

## 5.2 Subject doubling

The rusty speakers’ speech also exhibits a syntactic construction that nowadays still occurs in many (Belgian) Flemish dialects (Haegeman 1991, Van Craenenbroeck & van Koppen 2002, De Vogelaer & Devos 2008, De Vogelaer 2008), but which seems to have disappeared from West Zeelandic Flemish dialects altogether. Again, there are indications that this construction was a feature of the protolanguage of the Zeelandic Flemish immigrants in Espírito Santo. It concerns the construction referred to as pronominal subject doubling, which implies the use of a combination of the full and reduced form of the subject pronoun. As argued by De Vogelaer & Devos (2008: 249), the distribution of this phenomenon in

---

<sup>35</sup>In 14 out of 22 lexemes there is deletion of /l/ without compensatory lengthening of the preceding vowel.

<sup>36</sup>Some fragments of the interview were unclear and therefore difficult to transcribe.

the Flemish dialects depends on a number of parameters, including word order, clause type (main clause vs. subclause), type of subject in the clause (pronoun or not), the number of pronouns, etc. This is illustrated in the examples in (10)–(12) (examples are based on De Vogelaer & Devos 2008: 243–244):

- (9) Regular word order: clitic preceding the verb and strong pronoun after the verb
- ‘**k** Zal **ik** dat wel krijgen  
 I<sub>clitic</sub> shall I<sub>strong</sub> that ADV get  
 ‘I will get that’

In example (10) a regular order is attested in which the inflected verb is preceded by the subject. This example contains the 1SG-clitic ‘*k*’ in subject position and it is doubled by an optional strong pronoun *ik* ‘I’.

- (10) Inverted word order: clitic and strong pronoun following verbs
- Mag=**ek=ik** dat wel weten?  
 may=I<sub>clitic</sub>=I<sub>strong</sub> that ADV know  
 ‘Am I allowed to know that?’

The example in (10) is an “inverted” word order in which the inflected verb precedes the subject.<sup>37</sup>

- (11) Subclause: clitic and strong pronoun following complementiser
- ...da=**k=ik** dat mag weten  
 that=I<sub>clitic</sub>=I<sub>strong</sub> that may know  
 ‘...that I am allowed to know that’

In example (11), the clitic ‘*k*’ and the strong pronoun *ik* ‘I’ follow the complementiser *dat* ‘that’.<sup>38</sup>

- (12) Clitic and strong pronoun following a particle that introduces comparison<sup>39</sup>
- Hij is groter dan=**ek=ik**  
 he is taller than=I<sub>clitic</sub>=I<sub>strong</sub>

---

<sup>37</sup>It is likely that the inverted clitic pronoun duplicated by a strong pronoun is not in the subject position. The inversion can mark discursive effects such as “focus”, for example.

<sup>38</sup>As observed in example (11), the clitic ‘*k*’ and the strong pronoun *ik* ‘I’ are morphophonologically linked to the complementiser that is on the left periphery of the sentence (and not in subject position).

<sup>39</sup>The combination of reduced and full pronominal form cannot be called “doubling subject” in (12). The subject in (12) is *Hij* ‘he’ (and it is not doubled).

The construction of subject doubling with inverted word order (following verbs and following complementisers) occurs in the speech of the rusty speakers in Espírito Santo, as is illustrated in the examples (13)-(18). Mostly, these speakers use it in the case of the first person singular, but it also occurs with first person plural (as in example (14)).

- (13) Speaker 1 (A.)

Dan gaon=k=ik      mien tuug      uut      de (?) dan kom ik (?)  
 then go=I<sub>clitic</sub>=I<sub>strong</sub> my material out-of the (?) then come I (?)  
 'Then I will (get?) my material from the (?), then I come(?)'

- (14) Speaker 1 (A.)

Ja die      kon goed Hollands praten gelijk a=me=wudder  
 yes that-one could well Hollandic talk like COMP=we<sub>clitic</sub>=we<sub>strong</sub>  
 hier praten  
 here talk  
 'yes he could speak Dutch well like we talk here'

- (15) Speaker 3 (R.)

Dan doe=k=ik      dat in zakken, dan leg=ek=ik      die weg  
 then do=I<sub>clitic</sub>=I<sub>strong</sub> that in bags then put=I<sub>clitic</sub>=I<sub>strong</sub> that away  
 'then I put it in bags, then I put it away'

- (16) Speaker 3 (R.)

En dan doe=k=ik      daor weer frisse      onderbrengen  
 and then do=I<sub>clitic</sub>=I<sub>strong</sub> there again fresh-one bring-under  
 'and then I add some fresh one again'

- (17) Speaker 4 (B.)

An=k=ik      mee mien zusters topekommen      ja  
 COMP=I<sub>clitic</sub>=I<sub>strong</sub> with my sisters together-come yes  
 'when I come together with my sisters, you know'

- (18) Speaker 4 (B.)

An=k=ik      ilder eentwa      zeggen dan verstaon      zulder da  
 COMP=I<sub>clitic</sub>=I<sub>strong</sub> them something say<sub>1SG</sub> then understand they that  
 'when I say something to them then they understand it'

The extent to which subject doubling occurs, depends on the speaker. Speaker 1, for instance, uses the construction in the first person singular following a verb

(as in (13)) in 1 out of 4 possible cases, whereas speaker 3 uses subject doubling in the first person singular following a verb in 10 out of 13 possible cases.<sup>40</sup> Although the subject doubling construction occurs in the Zeelandic Flemish language of the speakers in Espírito Santo, it seems to be absent from contemporary Zeelandic Flemish dialects as spoken in the present-day motherland. According to De Vogelaer (2008: 243) and Barbiers et al. (2006) subject doubling in inverted word order for 1<sup>st</sup> person singular is only observed in one Zeelandic Flemish place (Hulst), but this place is in the eastern part of Zeelandic Flanders and falls outside the region the immigrants came from. Further, the construction is restricted to Belgian (Flemish) dialects, as is illustrated in Figure 5.

Likewise, only scant evidence of the occurrence of subject doubling in Zeelandic Flemish dialects is found by Will (2004), who studied the spontaneous speech of respondents from 39 Zeelandic Flemish villages as recorded in the 1960s-1970s by researchers of Ghent University and the Meertens Institute.<sup>41</sup> He observes a kind of remainders of subject doubling in the Zeelandic Flemish region, but only in 3% of possible cases and mostly in dialects from places that fall outside the region of the immigrants. Will concludes that already in the 1960s, subject doubling was a marginal feature of the Zeelandic Flemish dialects. However, van den Broecke-de Man (1978: 12) does mention subject doubling as a characteristic of West Zeelandic Flemish dialects and gives examples of the construction in normal and inverted word order.

As in the case of /l/-deletion, the occurrence of a linguistic feature in the speech of the rusty speakers which still occurs in the broader Flemish region, but seems to have disappeared almost entirely from the contemporary Zeelandic Flemish dialects, suggests that this feature is probably a protolanguage feature, which was brought along with the 19<sup>th</sup> century immigrants.

### 5.3 Inflected polarity markers *yes* and *no*

Two of our rusty speakers, more specifically speakers 1 and 2, both living in Holandinha, use a construction known as inflected *yes-* / *no-*particles, which im-

<sup>40</sup>This concerns possible contexts for subject doubling, in which doubling may or may not be applied, such as in the utterance in (i) in which the subject (*ik* 'I') is only expressed once, which means that there is no subject doubling in this case.

(i) da weet ik zo niet goed  
that know I PART not well  
'I'm not so sure about that' (speaker 1, A.)

<sup>41</sup><https://www.variaties.be/portfolio-item/stemmen-uit-het-verleden>    <https://www.meertens.knaw.nl/db/>

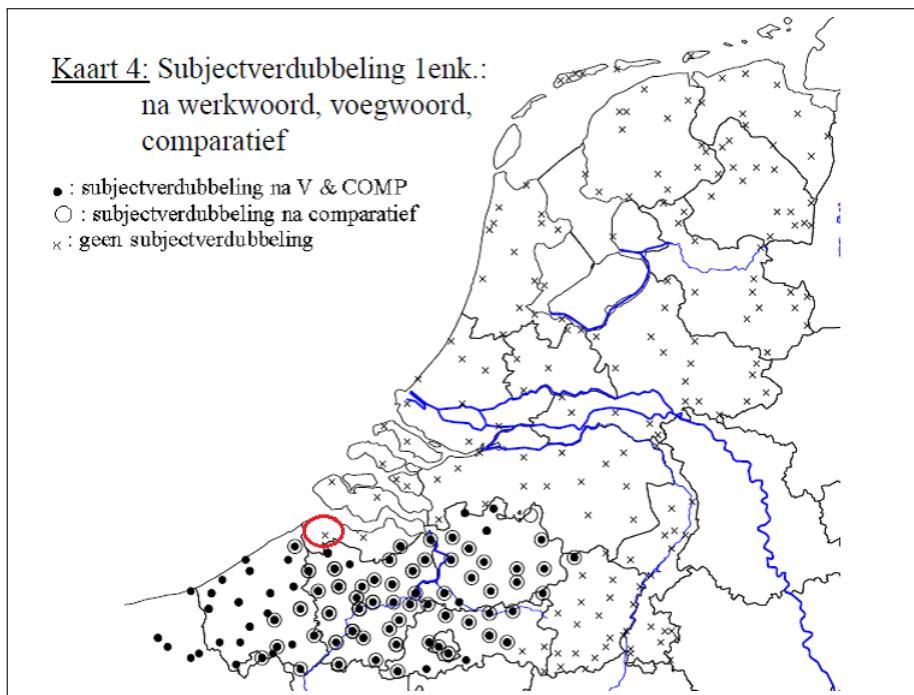


Figure 5: Geographical distribution of subject doubling in 1<sup>st</sup> person singular following a verb, complementiser or comparative in the Dutch-speaking area (Source: De Vogelaer 2008: 243, West Zeelandic Flanders is encircled with red).

plies that the polarity markers *yes* and *no* are followed by subject clitics (represented in bold in the examples (19)–(21)), which refer to the subject of a preceding utterance (underlined in (19)–(21)). The examples in (19) and (20) contain an inflected *yes*-particle referring to a singular, 3<sup>rd</sup> person, male antecedent, example (21) an inflected *yes*-particle referring to a plural, 3<sup>rd</sup> person antecedent.

(19) Speaker 1 (A.)

Die camion die      is kapot      gebrook'n. Jao=j,      die      staot  
that truck that-one is to pieces broken.      yes=he<sub>clitic</sub> that-one stands  
daa nog  
there yet

'That truck is broken to pieces. Yes, it is still standing there.'

- (20) Speaker 1 (A.)

Frans B., **jao=j** die èè nie meer terug gewist  
 Frans B. yes=he<sub>clitic</sub> that-one has not anymore back been  
 ‘Frans B., yes he did not come back anymore.’

- (21) Speaker 1 (A.):

die zoeten drinken, da ging nog dansen ne  
 those-ones sweet liquor drank, that went still dance<sub>inf</sub> TAG  
 ‘The ones that drank sweet liquor still went dancing.’

Speaker 2 (J.):

Dansen **jao=s**  
 dance<sub>inf</sub> yes=they<sub>clitic</sub>  
 ‘Dancing, yes they did.’

Such inflected *yes-* and *no-*particles are nowadays restricted to some (Belgian) Flemish dialects (Paardekooper 1993) (especially West and East Flemish dialects), and take the following forms:<sup>42,43</sup>

Table 1: Subject clitics following yes (/ no)-particle

1SG (ja)=k	1PL (ja)=m
2SG (ja)=g	2PL (ja)=g
3SG masc. (ja)=j	3PL (ja)=s
3SG fem. (ja)=s	
3SG neuter (ja)=t	

De Vogelaer & Devos (2008) discuss the distribution of these forms in the Dutch-speaking area (i.e. the Netherlands and Flanders). Whereas the *yes-*particle is inflected in many West- and East Flemish dialects (Belgium), De Vogelaer does not find any attestations of an inflected *yes-*particle in West Zeelandic Flemish dialects.<sup>44</sup> The distribution of the inflected *yes-*particle for 3SG masculine (as in

<sup>42</sup> Although other forms are possible (see Paardekooper 1993 and De Vogelaer & Devos 2008: 161, 168–170, 179–180, 195–197, 201–202), we restrict ourselves to the forms that are most common in the region surrounding West Zeelandic Flanders.

<sup>43</sup> Since we only found instances of the inflected polarity marker *yes* in the speech of the rusty speakers, we restrict ourselves to the possible forms of the *yes-*particle.

<sup>44</sup> There is one attestation of inflected *yes-*particles for 1SG in the municipality of Terneuzen, which is outside the region of West Zeelandic Flanders (see De Vogelaer 2008: 161).

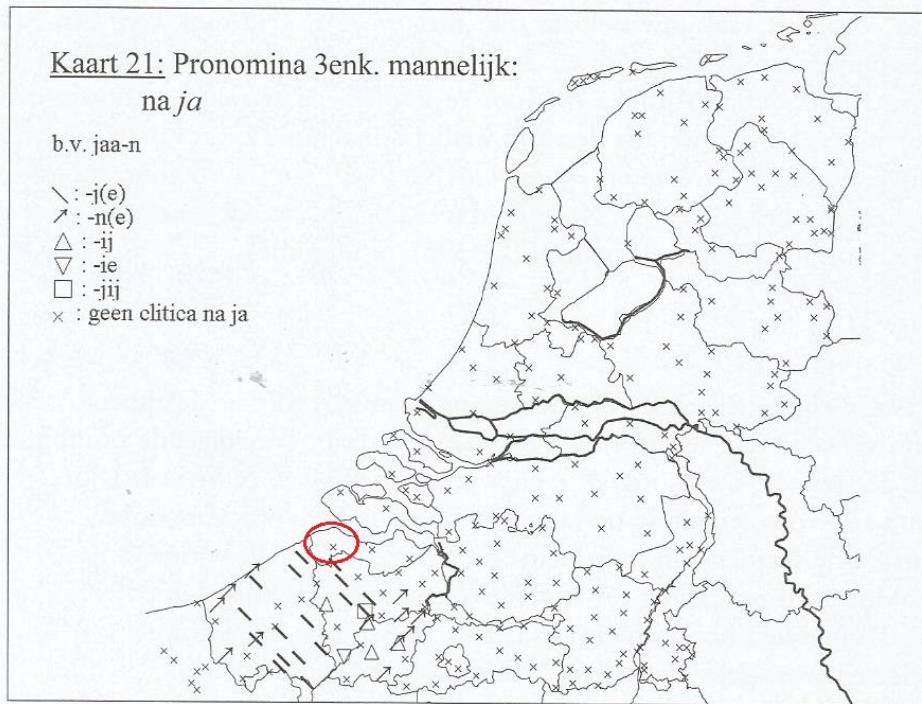


Figure 6: Distribution of the inflected yes-particle for 3SG masculine in the Dutch-speaking area. (Source: De Vogelaer 2008: 161)

the examples (19) and (20)) is represented in Figure 6 (Source: De Vogelaer & Devos 2008: 196).<sup>45</sup> yes-particle for 1SG (p. 161), 1PL (p. 169), 2SG (p. 179), 3SG feminine (p. 201), and 3SG neuter (p. 202). As can be seen on this map, there are no clitics (represented by the symbol x) following *ja* ('yes') in West Zeelandic Flanders (which is encircled in red on Figure 6), although *ja* is inflected (as *ja=j*, symbol: \) in the (Belgian) Flemish region to the south of the border with West Zeelandic Flanders.

The distribution of the inflected yes-particle for 3PL (as in example (21)) is represented in Figure 7 (Source: Barbiers et al. 2006) which shows that there is no inflection (symbolised by a blue square) of the yes-particle in present-day West Zeelandic Flanders (encircled with green), although inflected forms occur in the surrounding (Belgian) Flemish dialects.

De Vogelaer & Devos (2008) do not find any evidence of an inflected yes-particiles in Zeelandic Flemish dialects in whatever context, but according to Barbiers

<sup>45</sup>See De Vogelaer (2008) for maps of subject clitics following the

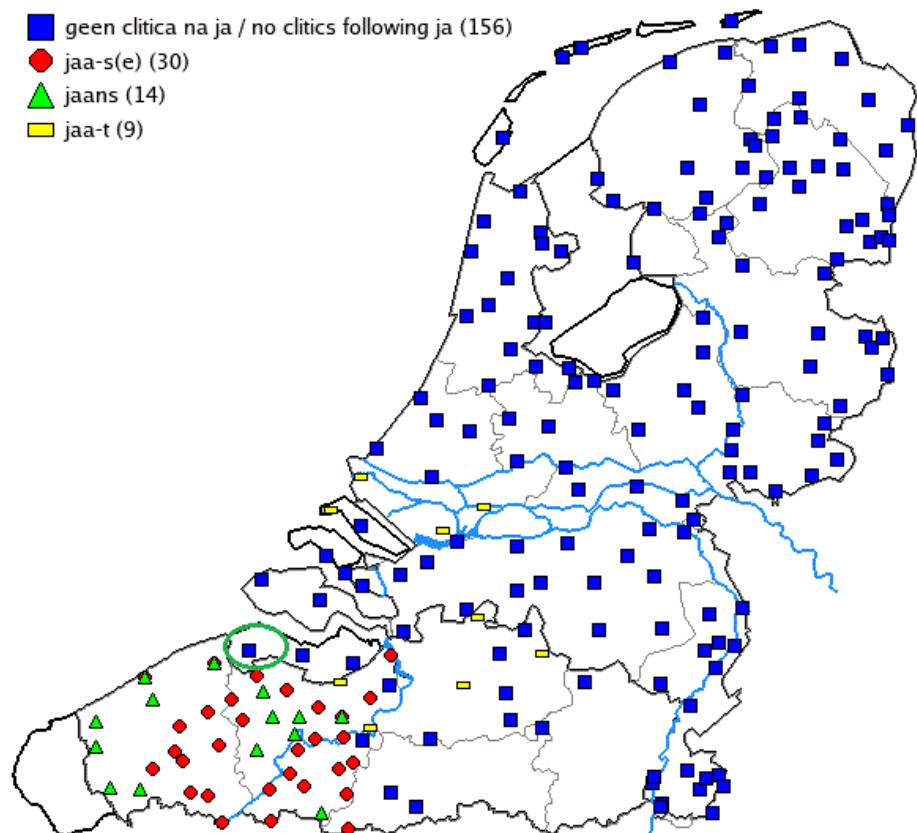


Figure 7: Distribution of the inflected yes-particle for 3PL in the Dutch-speaking area. (Source: Barbiers et al. 2006). Copyright resides with the Meertens Institute.

et al. (2006), one informant from Cadzand (West Zeelandic Flanders) replies in a written questionnaire that the inflected yes-particle is possible – although very rare – in his / her dialect in the case of 3SG feminine in the test sentence in (22):<sup>46</sup>

- (22) a. Question:  
 Gaat ze dansen?  
 goes she dance<sub>inf</sub>  
 'Is she going to dance?'

---

<sup>46</sup>This corresponds to test sentence 354 in Barbiers et al. (2006).

b. Answer: Ja's(e).

yes=she<sub>clitic</sub>

Likewise, one informant from Groede (West Zeelandic Flanders) replies that *ja* 'yes' can be followed by a clitic in his /her dialect in the case of 1SG as in the test sentence in (23)<sup>47</sup>, although its occurrence is very rare.

(23) a. Question:

Wil je nog koffie, Jan?  
want you PART coffee Jan  
'Do you want more coffee, Jan?'

b. Answer:

Ja'k.  
yes=I<sub>clitic</sub>  
'Yes I do.'

To summarise, the fact that the surrounding Flemish dialects have inflection, and that two West Zeelandic Flemish informants indicate that inflected yes-particles can occur, though infrequently, in their dialects, may suggest that inflected particles were more common in the Zeelandic Flemish protolanguage. Their occurrence in the speech of the rusty speakers in Espírito Santo supports this assumption.

#### 5.4 Conclusion

In this section we analysed the elicited speech of four rusty speakers of the Zeelandic Flemish dialect living in Espírito Santo. These speakers are descendants of 19<sup>th</sup> century Dutch immigrants and belong to a very small group of terminal speakers of Zeelandic Flemish in Brazil. They are considered rusty speakers since they acquired the language as their mother tongue, but they have stopped speaking their language due to a lack of contact with other speakers of Zeelandic Flemish. Because of this absence of regular usage of the language, the language of these rusty speakers is only influenced by the multilingual setting to a relatively small extent. As a consequence, many aspects of their language have remained rather similar to the protolanguage of the original immigrants. At the same time, the present-day motherland language, that is, the current Zeelandic Flemish dialect as spoken in the province of Zeeland (the Netherlands), has been subject to

---

<sup>47</sup>This corresponds to test sentence 353 in Barbiers et al. (2006).

processes of convergence to the northern Dutch standard language and dialect levelling in the past few decades. As a consequence, certain archaic linguistic features can be found in the language of the rusty speakers in Brazil, which have long since disappeared from the motherland language.<sup>48</sup> In this section we have argued that we can use these linguistic features to reconstruct the original immigrants' protolanguage. By way of experiment, we started out from a reduced data set, i.e. confined to the modern varieties in Brazil and the Netherlands only.

We have focused on three linguistic features: (1) deletion of /l/ in codas and coda clusters with concomitant compensatory lengthening of the preceding vowel, (2) subject doubling and (3) inflected yes- / no-particles. A similar pattern was found for all three features: whereas the speech of the rusty speakers contains instances of /l/-deletion, subject doubling, and inflected yes-particles, these phenomena hardly occur anymore (except from a few traces) in the contemporary West Zeelandic Flemish dialects, but do occur in the surrounding (Belgian) Flemish dialects. The occurrence of features in the speech of our rusty speakers that are also found in current Flemish dialects, seems to suggest that these features must have been brought along with the original immigrants and thus must once also have occurred in the dialects spoken in the villages of the motherland where these immigrants came from.

On the basis of these results, we can conclude that the speech of rusty speakers of a transplanted language can have historical value for the reconstruction of the original immigrants' protolanguage. Most research on heritage languages focuses on the internal changes the minority language undergoes in the process of language obsolescence and under influence of language contact. For the historical reconstruction of the protolanguage, however, one has to focus on the unchanged features in heritage language speakers' speech. The speech of rusty speakers lends itself best to such analysis, because these speakers use their language to such a small extent that some "original" language features have remained unchanged. However, when using heritage language data for historical reconstruction, one has to take into account the high degree of inter- and intra-speaker variation in the speech of terminal speakers and the possibility that certain features of their language are idiosyncrasies. Therefore, if historical data are available, one should incorporate them in the research, which we will do in Section 6. In addition, when studying speech island varieties that are threatened

---

<sup>48</sup>Our results about the disappearance of dialect features in the current West Zeelandic Flemish dialects are consistent with findings from the (sociolinguistic) literature that the Zeelandic Flemish regiolect is spoken to a much smaller extent than some other regiolects from the Dutch-speaking area (see Rys et al. 2019: 26) and that there has been a considerable amount of dialect loss in the province of Zeeland in the last five decades (Versloot 2021: 11).

with disappearance because they are dominated by other local or immigrant languages, as is the case for Zeelandic Flemish in Brazil, one should ideally always study the data from the perspective of language contact as well. This will be demonstrated in Section 7.

## 6 Reviewing our findings from the perspective of historical linguistics

In order to find out the extent to which rusty speaker data of a transplanted dialect (i.e. a speech island variety) alone suffice to reconstruct the so-called protolanguage of the original immigrants, we confined ourselves to a comparison of modern varieties of Zeelandic Flemish in Brazil and in the present-day motherland. By doing so, we actually ignored available historical data on the three linguistic features discussed. In what follows, we confront our findings with these historical data. This confrontation will demonstrate that reliance on rusty speaker data alone may in some cases (*/l/-deletion*, to be more specific) lead to the wrong conclusions and that historical data about the motherland variety are therefore essential in the evaluation of the data.

### 6.1 Historical data on */l/-deletion*

#### 6.1.1 Winkler 1874

As a matter of fact, there are several historical sources that reveal some information on the status of */l/-deletion* in older stages of the Flemish dialects (i.e. West- and East Flemish dialects as spoken in Belgium, as well as Zeelandic Flemish dialects as spoken in the Netherlands). The oldest source is the so-called *Dialecticon* of Winkler, published in 1874, in which the parable of the Prodigal Son is translated into various dialects of the Dutch-speaking area by speakers of those dialects.<sup>49</sup> Dialects that are included in the *Dialecticon* and which are of interest to our study are those of the East Flemish villages of Maldegem and Kleit and of the West Zeelandic Flemish places Eede, Heille, Aardenburg and Cadzand. In none of these dialects is */l/-deletion* found. More specifically, for all fragments together there are zero cases of */l/-deletion* out of 36 possible cases. Of course, one can question the accuracy of the broad transcription used in the *Dialecticon*. The transcription used by Winkler is not phonemic, but in normal alphabetic spelling, used in such a way that it represents the pronunciation of lexemes. This spelling

---

<sup>49</sup>[https://www.dbln.org/tekst/wink007alge02\\_01/](https://www.dbln.org/tekst/wink007alge02_01/)

is characterised by some limitations, though. A word like *zonen* ‘sons’, for example, is pronounced as [zɔ̃:s] in the present-day dialect of Maldegem, but is represented as *zeuns* in Winkler (1874). Thus, the vocalic phoneme /ø/ is written as *eu*. We do not know for sure, however, whether the /n/ was pronounced or not and – depending on this – whether the vowel was nasalised or not. It might be the case that the alphabetic spelling was not suitable for indicating any degree of nasalisation in this word, although one could imagine that in the case of deletion of /n/ the lexeme would have been represented as *zeu(n)s*, with brackets indicating a certain degree of deletion. Obviously, the same holds for cases of /l/-deletion: we assume that brackets would have been used to indicate (partial) deletion of /l/ or that /l/ would not have been written at all. However, in none of the 36 cases is *l* written between brackets or absent. We thus conclude that /l/-deletion seems to have been absent in the relevant dialects around 1874.

### 6.1.2 Corpus Dialectmateriaal Pieter Willems 1885

Another 19<sup>th</sup>-century source is the data that was collected by the dialectologist Pieter G. H. Willems (*Corpus Dialectmateriaal Pieter Willems*).<sup>50</sup> In 1885 and the following years Willems asked speakers of a large number of dialects to translate a list of more than 2000 lexemes into their dialects. Since Willems was particularly interested in phonological and morphological dialect phenomena, he added a document in which he was very explicit about the way the pronunciation of the lexemes had to be represented by the informants. Dialects that were included and that are of interest to this paper are the East Flemish dialect of Maldegem and the West Zeelandic Flemish dialects of Aardenburg, Zuidzande and IJzendijke. The list of lexemes contains 26 contexts for preconsonantal /l/-deletion and 33 contexts for prepausal /l/-deletion. However, we do not find indications that /l/ was deleted in any of these lexemes for any of these dialects around 1885.<sup>51</sup>

In her attempts to reconstruct the dialect of Maldegem of 100 years earlier, Versieck (1989) uses the *Corpus Dialectmateriaal Pieter Willems*. She evaluates the accuracy of the transcription used by the Maldegem informant in this corpus and concludes that this informant represented certain speech sounds only approximately. She also evaluates the question to which extent phonological processes of the Maldegem dialect (such as /l/-deletion) are manifested in Willems’ material. With respect to /l/-deletion, she reaches the same conclusion as we do:

<sup>50</sup><https://bouwstoffen.kantl.be/CPWNL/CPWNL.xq?browse=s181&act=browse#browse>

<sup>51</sup>Again, we might assume that /l/-deletion would have been indicated by either writing <l> between brackets or not writing it at all.

/l/-deletion is not represented in the *Corpus Dialectmateriaal Pieter Willems*. Versieck observes that there are only a few cases in which the lengthening of the vowel preceding /l/ seems to be indicated, for example in the lexeme *volk* ‘people’ (represented as *vôlk*). Versieck also argues that /l/-deletion is not represented in the relevant extract of Winkler (1874) either. Versieck (1989: 175) concludes that “zowel Willems als Winkler suggereren [...] dat (volledige) l-deletie in het toenmalige Maldegem niet voorkwam” (“Willems as well as Winkler suggest [...] that (complete) l-deletion did not occur in the Maldegem dialect at the time”). Versieck then continues by discussing the data on the Maldegem dialect which can be found in the *Reeks Nederlandse Dialectatlassen* (1935), but this evaluation will be discussed below.

### 6.1.3 Archief Jacobus van Ginneken 1910-1945

At the beginning of the 20<sup>th</sup> century, a Dutch linguist called Jacobus van Ginneken maintained a correspondence with Piet Meertens, dialectologist and first director of the Meertens Institute (Amsterdam). One of the items from this correspondence is Figure 8, which represents the vocalisation of /l/ in the lexemes *half* ‘half’, *kalk* ‘lime’, *kalf* ‘calf’ and *zalf* ‘ointment’. As can be seen from this map, vocalisation of /l/ in these lexemes occurred in an extensive area in West Flanders, a small area in the southwest of East Flanders and in a vast area in the southeast, which covers a part of present-day Vlaams-Brabant and large parts of Limburg (Belgium as well as the Netherlands). There are no indications on this map that vocalisation (or deletion) of /l/ was observed in the region of West Zealandic Flanders or the neighbouring East Flemish places. Of course, this map focuses specifically on vocalisation of /l/ in these four lexemes and does not reveal anything about vocalisation of /l/ in other lexemes. This implies that we cannot rule out the possibility that /l/-deletion was present in these dialects at that time.

### 6.1.4 Reeks Nederlandse Dialectatlassen 1935

A slightly more recent source is the *Reeks Nederlandse Dialectatlassen* (abbreviated *RND*)<sup>52</sup>, compiled by Edgar Blancquaert, of which the part about North East Flanders and Zealandic Flanders was published in 1935. The *RND* consists of 141 sentences that are translated into different dialects of the Dutch-speaking area. A great advantage of this source is that the sentences are transcribed in narrow

---

<sup>52</sup><https://www.dialectzinnen.ugent.be/transcripts/>

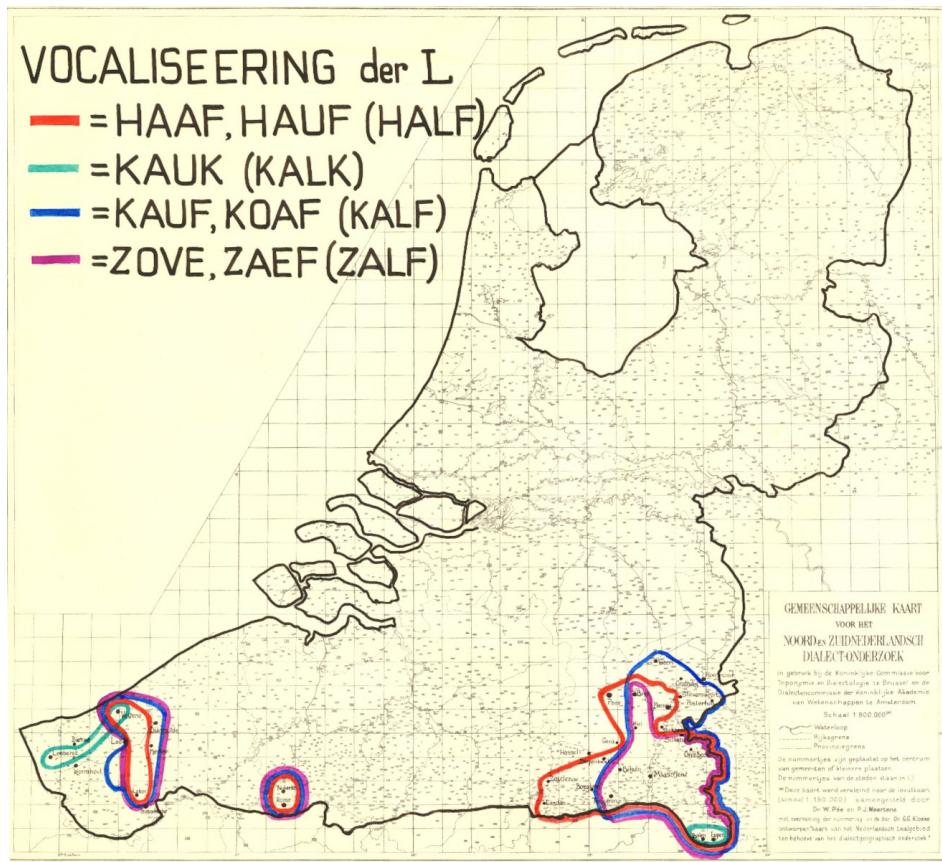


Figure 8: Vocalisation of /l/ in the lexemes *half* ‘half’, *kalk* ‘lime’, *kalf* ‘calf’ and *zalf* ‘ointment’ (Source: Archief Jacobus van Ginneken. Map: Vocaliseering der L. Archief Meertens Instituut, Kaart 19951). Copyright resides with the Meertens Institute.

phonetic transcription. We examined the *RND*-transcriptions for the East Flemish dialects of Maldegem, Kleit and Middelburg and the West Zeelandic Flemish dialects of Aardenburg, Biervliet, Breskens, Cadzand, Groede, Hoofdplaat and Retranchement. A brief discussion of the results for each of these dialects is necessary to gain more insight into the possible development of /l/-deletion.

The dialect showing most instances of /l/-deletion around 1935 is that of Maldegem. In total, we found 17 cases of full deletion of /l/ (in prepausal and preconsonantal contexts), but without the overlong quality of the preceding vowel that is characteristic of the present-day Maldegem dialect (cf. Taelde man 1966, Versieck 1989, and Rys 2007). Only 1 case of /l/-deletion and compensatory lengthening

of the preceding vowel was observed. This occurred in the lexeme *lijnzaadmeel* ‘linseed meal’ (pronounced as [lyzəmɛl]). There are 7 potential cases in which /l/ is not deleted, all of them involving preconsonantal /l/. Finally, the Maldegem dialect displays 2 cases of partial deletion of /l/, indicated by using brackets (i.e. (l)), both of which have preconsonantal contexts. Thus, it seems that /l/-deletion is a phonological process that did not occur in the dialect of Maldegem in the 19<sup>th</sup> century (cf. Winkler 1874 and Corpus Materiaal Willems 1885), but did occur, though not categorically, in 1935. It seems to have been present with the characteristic overlong vowel in prepausal contexts first (cf. *lijnzaadmeel*). Cases of partial deletion likely indicate that /l/-deletion was an ongoing phonological change at that time. This hypothesis is supported by the many cases of /l/-deletion without the compensatory lengthening of the vowel and the cases in which there is no deletion of /l/ at all.

For Kleit, which nowadays has a dialect which closely resembles that of Maldegem (Taeldeman 1966), we found 9 cases of /l/-deletion lacking the compensatory lengthening of the preceding vowel, and we found 1 case with compensatory lengthening, more specifically in the lexeme *spel* ‘game’, which has a prepausal context. Further, we found 9 potential cases in which /l/ is not deleted, all of them involving preconsonantal /l/. Finally, 10 cases of partial deletion were observed. Thus, Kleit shows more or less the same development of /l/-deletion as Maldegem.

The dialect spoken in Middelburg, another sub-municipality of Maldegem, barely showed any cases of /l/-deletion in 1935. There is 1 case of deleted /l/ without compensatory lengthening of the vowel (i.e. in *karnemelk* ‘buttermilk’) and 1 case of partial deletion (in *melkboer* ‘milkman’), but 26 potential cases in which /l/ is not deleted.

All of the West Zeelandic Flemish dialects of the places mentioned above<sup>53</sup> more or less display similar results: potential cases of /l/-deletion in which /l/ is not deleted vary between 30 and 35. Cases of deleted /l/ without compensatory lengthening occur in each of these dialects, but almost always in the lexeme *veel* ‘much, many’, in which /l/ is in a prepausal context.<sup>54</sup> Thus, the process of /l/-deletion in the West Zeelandic Flemish dialects seems to be lexically determined.<sup>55</sup> Occasionally, we do find some other cases: the dialect of Cadzand has

<sup>53</sup> Aardenburg, Biervliet, Breskens, Cadzand, Groede, Hoofdplaat and Retranchement.

<sup>54</sup> However, the /l/-deletion in the lexeme *veel* is not categorical, since we also found cases in which /l/ does not delete.

<sup>55</sup> Recall that in the modern data, more specifically the *Nederlandse Dialectenbank*, /l/-deletion in the West Zeelandic Flemish dialects was also restricted to prepausal contexts in a limited set of frequently occurring lexemes (viz. *veel* ‘much’ (also *zoveel* ‘this much’, *hoeveel* ‘how much’), *wel* ‘PART’, *zal* ‘shall’, *al* ‘already’, and *nogal* ‘quite’).

1 case of /l/-deletion without compensatory lengthening in the word *kelder* ‘cellar’ and the dialect of Biervliet displays partial deletion of /l/ in *kelder* and full deletion (without compensatory lengthening) in *helft* ‘half’ and *wel* ‘PART’.

In her reconstruction of the Maldegem dialect of 100 years earlier, Versieck (1989) also discusses the *RND*-material. Like us, she observes that /l/-deletion with compensatory lengthening is present in one lexeme only (*lijnzaadmeel*), that there are some cases in which /l/ is deleted but the preceding vowel is short and that there are also cases which do not have /l/-deletion (as opposed to the present-day Maldegem dialect). Versieck (1989: 176) argues that it is unlikely that the process of /l/-deletion, which is so characteristic of the contemporary Maldegem dialect, would have taken place in the relatively short period between the publication of Winkler (1874) and the *Corpus Dialectmateriaal Pieter Willem*s (1885) on the one hand and the *RND* (1935) on the other hand.<sup>56</sup> century, did find /l/-deletion in the speech of his informants. She therefore assumes that the informants of these latter two sources were not able to represent the phonetic detail necessary to indicate /l/-deletion. She concludes that because of these reasons it may be assumed that the Maldegem dialect of 1885 was already characterised by deletion of /l/. By way of “evidence”, she mentions the example of the toponym *Eelvelde*, which is pronounced as [e'væ:də] and not as [e:'væ:də]. Versieck (1989: 176) argues that the fact that the vowel of the first syllable of this toponym is “no longer” lengthened indicates that speakers’ awareness of an underlying /l/ is “already completely blurred”, and that this indicates that the process of /l/-deletion must be “very old”. However, there is a fallacy in this argument, since /l/-deletion with compensatory lengthening of the preceding vowel typically affects stressed syllables. The first syllable of *Eelvelde* is not stressed though, which – according to us – explains the absence of lengthening of the vowel.<sup>57</sup>

All in all, we believe that Versieck (1989) does not come up with convincing arguments for the presence of /l/-deletion in the Maldegem dialect around 1874/1885. Based on the *RND*, however, we have reason to assume that /l/-deletion was an ongoing phonological process around 1935, which probably originated in the dialects of Maldegem and Kleit and which may have spread to neighbouring places (e.g. Middelburg, West Zeelandic Flanders) to some extent. Around 1935

<sup>56</sup> As Versieck (1989: 176) points out, the informants of the *RND* were born in the beginning of the 20<sup>th</sup> century (more specifically 1904 and 1911), whereas the informants of Winkler (1874) and of the *Corpus Dialectmateriaal Pieter Willem*s in 1811 and 1855, respectively. She further argues that Taeldeman (1966), who interviewed informants born at the end of the 19

<sup>57</sup> Compare with the lexeme *soldaat* ‘soldier’, in which stress is on the second syllable. Although /l/ may be partially deleted in this word, the preceding vowel is not lengthened either.

the process did not apply categorically, as opposed to the present-day situation (cf. Taeldeman 1966, De Schutter 2005, and Rys 2007).

To conclude, historical sources from some years after the migration of Zeelandic farmers to Brazil (more specifically from 1874 and 1885) do not contain indications that /l/-deletion was present in the language of these West Zeelandic Flemish migrants, nor in the surrounding East Flemish dialects of that time. Around 1935, /l/-deletion seems to be an ongoing process in some of these dialects and modern data show that this process is categorical in present-day Maldegem and Kleit and nearly categorical in the West Zeelandic Flemish village of Eede (and to a smaller extent in Middelburg). This may imply that /l/-deletion is a relatively “new” feature of these dialects and that its occurrence in the language of the Brazilian rusty speakers cannot be considered as an archaic feature of the so-called protolanguage. So, in this case, the reconstruction of the old Zeelandic Flemish language on the basis of rusty speaker data failed. Obviously, one wonders how the same feature of /l/-deletion and compensatory lengthening of the preceding vowel can be present in this speech island variety as well as in some present-day homeland varieties. An alternative explanation will be discussed in Section 7.

## 6.2 Historical data on subject doubling

For the case of subject doubling, there is historical evidence from only a few years after the time the Zeelandic farmers migrated to Brazil. The *Dialecticon* of Winkler (1874) contains a number of fragments which testify to the occurrence of subject doubling in various West Zeelandic Flemish dialects. We found 16 forms of subject doubling for the dialects of Eede/Heille, Aardenburg and Cadzand, such as example (24):

- (24) oeveel errebeiers van m'n voader ææn alles in de vulte, in 'k  
 vergoane-'k-ik van oenger  
 'how many of my father's workmen have everything in abundance, and I  
 am starving.' (Cadzand)

Winkler comments that the “double repetition” of the personal pronoun of 1SG is typically “Vlaamsch” ‘Flemish’, but is used in some Zeelandic Flemish dialects. He argues that it is particularly used to emphasise the personal pronoun and especially in confidential conversations which are characterised by strong emotions like anger or complaint.

Thus, in this case, the rusty speaker data could successfully be used to reconstruct the West Zeelandic Flemish protolanguage. As we inferred from the language of our rusty speakers, and as confirmed by the historical data, subject doubling is an archaic Flemish feature that was present abundantly in the West Zeelandic Flemish dialects in the days of emigration, but has disappeared almost entirely from the contemporary Zeelandic Flemish dialects (see Section 5.2).

### 6.3 Historical data on the inflected polarity markers yes and no

Paardekooper (1993) discusses various sources that provide information on the occurrence of inflected *yes*- and *no*-particles in the dialects of the Dutch-speaking area. The oldest source is a map based on the *Corpus Materiaal Willems* (1885) (see Figure 9). This map demonstrates (by use of the filled circle) that inflection of *ja* ‘yes’ as well as *nee* ‘no’ and in combination with all person features (i.e. 1SG-3SG and 1PL-3PL) occurs in the West Zeelandic Flemish region the immigrants departed from as well as in the neighbouring West- and East Flemish dialects. Thus, the use of inflected polarity markers was omnipresent in the relevant dialects only a couple of decades after the migration to Espírito Santo. This makes it very likely that inflected polarity markers were present in the protolanguage of the Zeelandic migrants as well.

Another, more recent map of the phenomenon is found in the *Archief Jacobus van Ginneken* (1910–1945) (see Figure 10). This map shows that at the beginning of the 20<sup>th</sup> century, the inflection of the polarity marker *ja* was still ubiquitous in the dialects of West Zeelandic Flanders as well as the surrounding West and East Flemish dialects (all delineated by the pink line in Figure (10)).

As we made clear in Section 5.3, according to modern sources (i.e. Barbiers et al. 2006, De Vogelaer & Devos 2008) inflected *yes*- and *no*-particles have disappeared entirely from the West Zeelandic Flemish dialects. On the basis of the Brazilian rusty speaker data, we assumed that the phenomenon must have been present in the dialects of West Zeelandic Flanders in the days of migration. The two historical sources discussed in this section indeed testify to the presence of such inflected polarity markers in the protolanguage of the immigrants.

## 7 Reviewing our findings from the perspective of language contact

In Section 6.1 we discussed the historical data on /l/-deletion and demonstrated that these did not match the conclusions we had drawn on the basis of the rusty

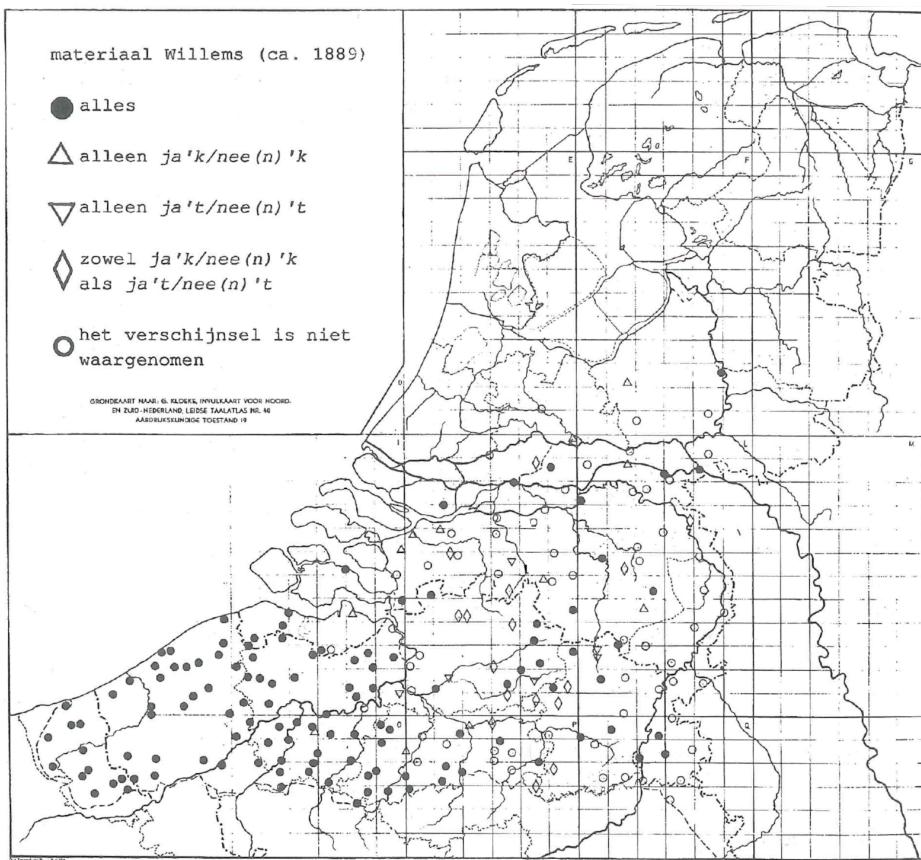


Figure 9: Occurrence of inflected polarity markers *yes* and *no* in the dialects of the Dutch-speaking area based on the *Corpus Materiaal Willem* 1885. Source: Paardekooper (1993)

speaker data. The historical sources which date back to a couple of decades after the migration of Zeelandic farmers to Brazil (1874 and 1885, to be more precise) do not contain indications that /l/-deletion was present in the dialects of these West Zeelandic Flemish migrants, nor in the surrounding East Flemish dialects of that time. This implies that its occurrence in the language of the rusty speakers in Espírito Santo cannot be considered as an archaic feature of the so-called protolanguage. Therefore, we must look for another explanation of this phenomenon in the language of the rusty speakers. We do this by approaching it from the perspective of language contact, a perspective that is traditionally emphasised within the field of language death studies (Dressler 1996, 1972, 1996, Dorian 1977, 1981).

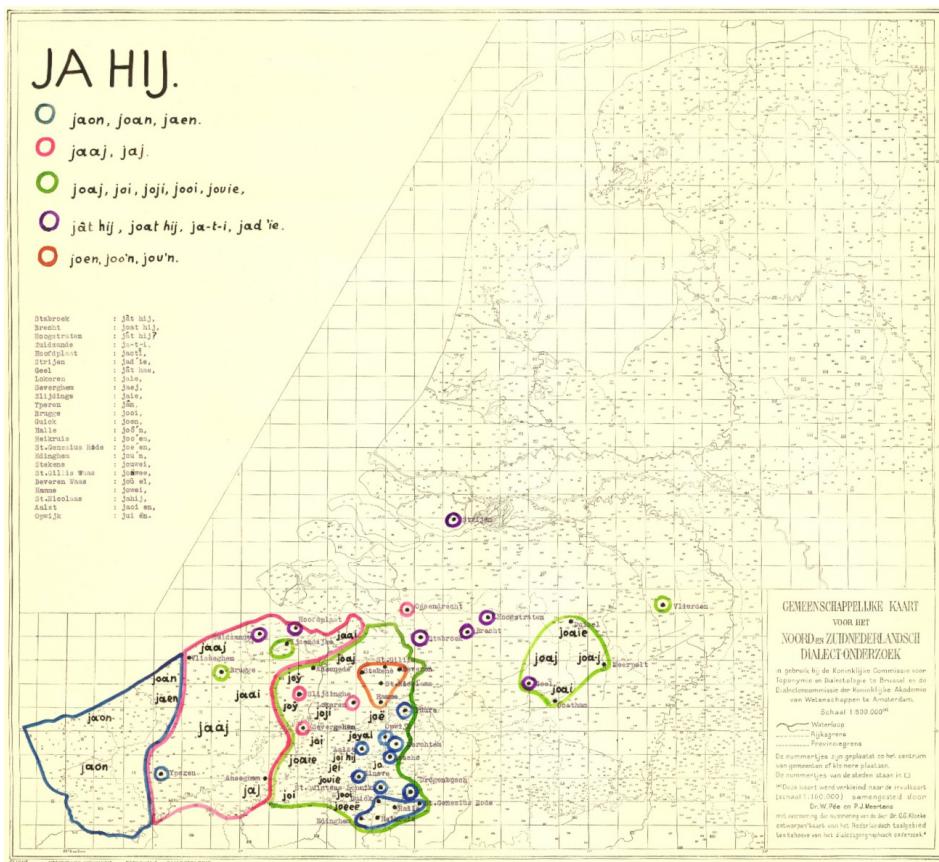


Figure 10: Occurrence of the inflected polarity marker *ja* ‘yes’ in the dialects of the southern part of the Dutch-speaking area (Source: Archief Jacobus van Ginneken. Map: Ja-hij, Kaart20766). Copyright resides with the Meertens Institute.

The two contact languages of Zeelandic Flemish in Espírito Santo are Brazilian Portuguese (i.e. the national language) and Pomeranian, which is another transplanted language. In Schaffel Bremenkamp et al. (2017) it was demonstrated that the Zeelandic Flemish language was heavily influenced by both of these dominant languages (e.g. lexical borrowing, calquing, relative pronoun neutralisation, structural reduction). Therefore, we should also look at Brazilian Portuguese and Pomeranian in order to find out whether the /l/-deletion observed in the speech of the rusty speakers could possibly be related to one of these languages. As a matter of fact, vocalisation of /l/ in the coda is a common feature of Brazilian Portuguese. Barbosa & Albano (2004: 229) observe that “[t]he archiphoneme

/L/, which some generations ago used to be a velarised lateral approximant, is changing into a labial-velar approximant throughout the entire Brazilian territory, producing homophones such as *mau* maʊ̯ ‘bad’ and *mal* maʊ̯ ‘evil.’ Thus, in Brazilian Portuguese, historical [ɫ] (i.e. so-called ‘dark l’, /l/ in the syllable coda) has been vocalised, rendering [ɔ̯].<sup>58</sup> Assuming that Brazilian Portuguese /l/-vocalisation influences the rusty speakers’ pronunciation of Zeelandic Flemish words with coda /l/, the following alternation is plausible:

- (25) *bal-ke* ‘beam’ > /bal.kə/ > [baʊ̯.kə]

Brazilian Pomeranian (as opposed to European Pomeranian) is characterised by a productive process of monophthongisation (Postma 2019: 56), in which /au/ is realised as /a:/, rendering:

- (26) *blaum* > *blaam* ‘flower’ (example from Postma 2019: 56)

If we assume that the Brazilian Portuguese process exemplified in (25) is feeding the Brazilian Pomeranian process illustrated in (26), we get the following alternation:

- (27) *bal-ke* ‘beam’ > /bal.kə/ > [baʊ̯.kə] > [ba:.kə]

This alternation can also be applied to the rusty speaker’s example given in (4), and repeated as (28) below,

- (28) Koeien zo kalvers ['ka:vərs]

‘cows, you know, calves’

in which the alternation would be:

*kal-vers* ‘calves’ > /kal.vərs/ > [kaʊ̯.vərs] > [ka:.vərs]

It might then be the case that this combined process of vocalisation and monophthongisation, which in first instance affected lexemes containing the sequence /al/, was by analogy extended to other words as well (e.g. example (3) *ielk* ‘each one’, example (5) *melken* ‘to milk’).

We can conclude that it is plausible to assume that the occurrences of forms with /l/-deletion in the language of the rusty speakers (see examples (1)–(6)) are

---

<sup>58</sup>Notice that the Brazilian Portuguese step of vocalisation is also present in some Flemish dialects from the relevant region, e.g. the lexeme *balk* ‘beam’ is realised as [bow.ka] in Moerkerke (West Flanders) and Eeklo (East Flanders) (see GTRP-database, De Schutter 2005).

the results of the Brazilian Portuguese process of /l/-vocalisation feeding the Brazilian Pomeranian process of monophthongisation, since the alternative hypothesis, stating that these forms are relics of the Zeelandic Flemish immigrants' protolanguage, did not "survive" the confrontation with the available historical data. This outcome shows the importance of always including the perspective of language contact into the study of linguistic features of declining languages.

## **8 Conclusion**

In this paper we have drawn attention to the condition that transplanted (cf. "di-aspora") languages are not only subject to levelling, koineisation, or other processes of language simplification, but can – in other respects – also be rather conservative in that they can retain archaic features found in the motherland variety. Particularly, when a transplanted language variety is roofed by a language which is structurally very different, archaic features may be expected. We argued that such archaic features are found in the language of a number of so-called rusty speakers of Zeelandic Flemish in Espírito Santo (Brazil). We focused on three linguistic features which we believed to be archaic features of the protolanguage: (1) deletion of /l/ in codas and coda clusters with concomitant compensatory lengthening of the preceding vowel, (2) subject doubling in inversion contexts and (3) inflected polarity markers *yes* and *no*. By way of experiment, we confined ourselves in first instance to a comparison of modern varieties, ignoring any available historical data. In the case of features (2) and (3), our findings demonstrate the potential historical value of transplanted dialects in the reconstruction of the original immigrants' language: a comparison of modern varieties shows that subject doubling as well as inflected *yes*- and *no*-particles are both phenomena that still occur in some Belgian Flemish dialects, but have disappeared from the contemporary variety as spoken in the Dutch province of Zeelandic Flanders. With respect to these cases, historical data bear witness to the presence of these features in the West Zeelandic Flemish dialect in the days of emigration. However, in the case of /l/-deletion, the available historical data do not support our hypothesis that it concerns a relic feature of the protolanguage. Thus, reliance on rusty speaker data alone leads to wrong conclusions in this case. Alternatively, we find an explanation by approaching the phenomenon from the perspective of language contact. We argued that the cases of /l/-deletion in the rusty speakers' speech are probably the result of the Brazilian Portuguese process of /l/-vocalisation feeding the Brazilian Pomeranian process of monophthongisation. Summarizing, we may say that a multidisciplinary perspective is the most preferable approach in the study of declining languages.

## Acknowledgements

Our special thanks go to Lea Busweiler for her help in the transcription of parts of the recordings. We are also grateful to Andrew Nevins and two anonymous reviewers for their comments on earlier versions of this paper.

## References

- Appel, René & Pieter Muysken. 1987. Language maintenance and shift. In René Appel & Pieter Muysken (eds.), *Language contact and bilingualism*, 32–45. London: Edward Arnold.
- Archief Jacobus van Ginneken. 1910–1945. <https://www.meertens.knaw.nl/kaartenbank/>.
- Barbiers, Sjef, Hans Bennis, Gunther De Vogelaer, Magda Devos & Margreet Van der Ham. 2006. *Dynamische syntactische atlas van de Nederlandse dialecten (DynaSAND)*. Amsterdam: Meertens Instituut.
- Barbosa, Plínio A. & Eleonora C. Albano. 2004. Brazilian Portuguese. *Journal of the International Phonetic Association* 34(2). 227–232.
- Botma, Bert & Erik-Jan van der Torre. 2000. The prosodic interpretation of sonorants in Dutch. In Helen de Hoop & Ton van der Wouden (eds.), *Linguistics in the Netherlands 2000*, 17–30. Amsterdam & Philadelphia: John Benjamins Publishing.
- Buyssse, Frans. 1984. *De Zeeuws-Vlaamse gemeenschap van Holanda, Brazilië (1858–1982)*. Nijmegen: Katholieke Universiteit Nijmegen. (MA thesis).
- Campbell, Lyle & Martha C. Muntzel. 1989. The structural consequences of language death. In Nancy C. Dorian (ed.), *Investigating obsolescence: Studies in language contraction and death*, 181–196. Cambridge: Cambridge University Press.
- De Schutter, Georges. 2005. *Morfologische atlas van de Nederlandse dialecten: deel 1: meervoudsvorming bij zelfstandige naamwoorden, vorming van verkleinwoorden, geslacht bij zelfstandig naamwoord, bijvoeglijk naamwoord en bezittelijk voornaamwoord*. Vol. 1. Amsterdam: University Press.
- De Vogelaer, Gunther. 2008. *De Nederlandse en Friese subjectmarkeerders. Geografie, typologie en diachronie*. Ghent: Koninklijke Academie van Nederlandse Taal- en Letterkunde.
- De Vogelaer, Gunther & Magda Devos. 2008. On geographical adequacy, or: How many types of subject doubling in Dutch. *Syntax and Semantics* 36. 251–276.
- De Wulf, Chris, Jan Goossens & Johan Taeldeman. 2005. *Fonologische atlas van de Nederlandse dialecten: Deel IV: de consonanten*. Ghent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde.

- Dorian, Nancy. 1977. The problem of the semi-speaker in language death. *International Journal of the Sociology of Language* 12. 23–32.
- Dorian, Nancy. 1981. *Language death: The life cycle of a Scottish Gaelic dialect*. Philadelphia: University of Pennsylvania Press.
- Dressler, Wolfgang U. 1972. On the phonology of language death. In Paul M. Peranteau (ed.), *Papers from the eighth regional meeting of the Chicago Linguistic Society*, 448–457. Chicago: Chicago Linguistic Society.
- Dressler, Wolfgang U. 1996. Language death. In Rajendra Singh (ed.), *Towards a critical sociolinguistics (current issues in linguistic theory 125)*, 195–210. Amsterdam & Philadelphia: John Benjamins Publishing.
- Fase, Willem, Koen Jaspert & Sjaak Kroon (eds.). 1992. *Maintenance and loss of minority languages* (Studies in Bilingualism 1). Amsterdam: John Benjamins Publishing.
- Feyér, Balint, Peter Racz, Marton Soskuthy & Daniel Szereedi. 2012. *A phonetic study of l-deletion in Hungarian*. [https://s18798.pcdn.co/szeredi/wp-content/uploads/sites/12/labphon\\_poster\\_final.pdf](https://s18798.pcdn.co/szeredi/wp-content/uploads/sites/12/labphon_poster_final.pdf) (3245).
- Field, Thomas T. 1985. Language attitudes and language death in the French Pyrenees. *The SECOL Review* 9. 44–67.
- Haegeman, Liliane. 1991. Subject clitics and clitic doubling in West Flemish. In Henk van Riemsdijk & Luigi Rizzi (eds.), *Clitics and their hosts*, 99–154. Tilburg: Grammatical Models.
- Paardekooper, P. C. 1993. Jaak/neenik enz. *TABU* 23. 143–170.
- Postma, Gertjan. 2014. Complex complementizers and the structural relation with weak T: New (morpho)syntactic data from a Pomeranian language island in Brazil. *STUF – Language Typology and Universals* 67. 627–663. DOI: 10.1515/stuf-2014-0032.
- Postma, Gertjan. 2019. *A contrastive grammar of Brazilian Pomeranian*. (Linguistik Aktuell / Linguistics Today 248). Amsterdam: John Benjamins Publishing.
- Roos, Ton & Margje Eshuis. 2008. *Os capixabas holandeses: Uma história holandesa no Brasil*. Barneveld: BDU Uitgevers.
- Rosenberg, Peter. 2005. Dialect convergence in the German language islands (Sprachinseln). In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect change: Convergence and divergence in European languages*, 221–235. Cambridge: Cambridge University Press.
- Rottet, Kevin J. 1995. *Language shift and language death in the Cajun French-speaking communities of Terrebonne and Lafourche parishes, Louisiana*. Bloomington, IN: Indiana University. (Doctoral dissertation).

- Rys, Kathy. 1999. *Dialect en AN van inwoners van Eede*. Een sociolinguïstisch onderzoek naar resistentie en accommodatie. Ghent: Ghent University MA thesis.
- Rys, Kathy. 2000. Dialect in Eede, een onderzoek naar resistentie en accommodatie. In Véronique De Tier, Magda Devos & Jacques Van Keymeulen (eds.), *Huldealbum Hugo Ryckeboer. Nochtans was scherp van zin*, 349–360. Ghent: Vakgroep Nederlandse Taalkunde.
- Rys, Kathy. 2007. *Dialect as a second language: Linguistic and non-linguistic factors in secondary dialect acquisition by children and adolescents*. Ghent: Ghent University. (Doctoral dissertation).
- Rys, Kathy, Usha Balesar, Johan De Caluwe, Wilbert Heeringa, Frans Hinskens, Sita Doerga Misier, Maarten van der Peet & Margareth Rozenblad. 2019. Onderzoeksrapport staat van het Nederlands 2. Over de taalkeuzes van Surinamers in het dagelijks leven en meer over die van Nederlanders en Vlamingen.
- Sasse, Hans-Jürgen. 1991. *Arvanitika: Die Albanischen Sprachreste in Griechenland. Teil 1: Die Mundart Nordost-Attikobootiens*. Wiesbaden: Harassowitz.
- Sasse, Hans-Jürgen. 1992. Language decay and contact-induced change: Similarities and differences. In Matthias Brenzinger (ed.), *Language death: Factual and theoretical explanations with special reference to East Africa*, 59–80. Berlin: De Gruyter Mouton.
- Schaffel, Elizana. 2010. *Análise sociolinguística do desaparecimento da língua holandesa no espírito santo*. Vitória: Universidade Federal do Espírito Santo MA thesis.
- Schaffel Bremenkamp, Elizana, Kathy Rys, Andrew Nevins & Gertjan Postma. 2017. Zeeuws-Flemish in Brazil: Multilingualism and language decay. *Gragoata* 22. 435–472.
- Taeldeman, Johan. 1966. *Kleit in zijn taallandschap: Een fonetisch-fonologische dialectstudie*. Ghent: Ghent University MA thesis.
- Taeldeman, Johan. 1979. Op fonologische verkenning in Zeeuws-Vlaanderen. *Taal en Tongval* 31. 143–193.
- Tressmann, Ismael. 2006a. *Pomerisch-Portugiesisch Wörbauk – Dicionário Encyclopédico Pomerano-Português*. Santa Maria de Jetibá: Secretaria de Educação.
- Tressmann, Ismael. 2006b. *Upm Land – up pomerisch språk*. Santa Maria de Jetibá: Secretaria de Educação.
- Van Craenenbroeck, Jeroen & Marjo van Koppen. 2002. Pronominal doubling and the structure of the left periphery in Southern Dutch. In Sjef Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic microvariation*, 280–304. Amsterdam: Meertens Institute Electronic Publications in Linguistics.

- van den Broecke-de Man, E. J. 1978. *Dialect in West-Zeeuwsch-Vlaanderen*. Groede: Uitgeverij Pieters.
- Van Reenen, Piet. 1986. The vocalisation of /l/ in standard Dutch: A pilot study of an ongoing change. In Frits Beukema & Aafke Hulk (eds.), *Linguistics in the Netherlands 1986*, 189–198. Dordrecht: Foris.
- Versieck, Sabina. 1989. *Het Maldegemse klanksysteem in het heden en honderd jaar geleden*. Ghent: Ghent University MA thesis.
- Versloot, Arjen. 2021. Streektaaldoed in de lage landen. *Taal en Tongval* 72. 7–16.
- Von Tschudi, Johann Jakob. 2004. *Viagem à província do espírito santo. Imigração e Colonização Suíça 1860*. Vitória: Arquivo público do estado do Espírito Santo.
- Will, George. 2004. Zeggen k-ik da(t) zô? Reduplicatie van het subjectspronomen in Zeeuws-Vlaanderen. *Taal en Tongval* 56. 187–211.
- Willems, Pieter. 1885. *Corpus dialectmateriaal Pieter Willems*. <https://bouwstoffen.kantl.be/CPWNL/>.
- Winkler, Johan. 1874. *Algemeen Nederduitsch en Friesch dialecticon* (deel 2). The Hague: Martinus Nijhoff.

# Name index

- Łepkowski, Józef, 271
- Aalberse, Suzanne, 29, 48
- Adelaar, Willem F. H., 82, 90, 108
- Aikhenvald, Alexandra, 213–215
- Åkermark, Sia Spiliopoulou, 64
- Albano, Leonora C., 378
- Albó, Xavier, 82
- Allard, Real, 64
- Anders, Heinrich, 229
- Andrade Ciudad, Luis, 82, 83
- Andrade, Luis, 83
- Andrason, Alexander, 213, 215, 218–232, 234, 241, 242, 248, 276
- Andriani, Luigi, 10, 39
- Ansaldo, Umberto, 214
- Appel, René, 346
- Archief Jacobus van Ginneken, 372, 378
- Assouline, Dalit, 145, 146, 161, 172, 175, 176
- Austin, Peter, 214
- Babel, Anna, 83, 111
- Babel, Anna M., 83
- Bacher, Josef, 240
- Badii, Remo, 218
- Bagna, Carla, 16
- Bąk, Piotr, 238–240
- Baker, Mark C, 316
- Bakker, Peter, 214
- Bakker, Wiljo, 206–208
- Balowska, Grażyna, 262
- Barbiers, Sjef, 362, 365–367, 376
- Barbosa, Plínio A., 378
- Bargiełowna, Maria, 245
- Barnes, Michael, 244
- Bayer, Josef, 310, 313
- Belk, Zoë, 145, 151, 156, 158, 172, 176, 179, 182, 184
- Bell, Allan, 134, 139
- Belletti, Adriana, 39
- Benincà, Paola, 37
- Bennis, Hans, 313, 326
- Berleant, Arnold, 285
- Bert, Michel, 214
- Besch, Werner, 243
- Bettoni, Camilla, 19
- Biale, David, 144
- Bickerton, Derek, 305
- Biedrzycki, Leszek, 241
- Bloch-Rozmaj, Anna, 239
- Blommaert, Jan, 124
- Bonney, Rick, 191
- Bosque, Ignacio, 81
- Bossong, Georg, 39
- Botma, Bert, 352
- Braber, Natalie, 218, 221–224, 226–228, 230
- Brandi, Luciana, 37
- Brandner, Ellen, 310, 313
- Brandstätter, Julia, 229, 231
- Breuker, Pieter, 193, 203
- Briggs, Lucy T, 82

## *Name index*

- Britain, David, 307  
Brockhaus, 309  
Brunelle, Marc, 190  
Buyssse, Frans, 343  
Bybee, Joan, 80  
  
Cameron, Deborah, 136  
Campbell, Lyle, 213, 214, 346  
Caratini, Emilie, 221–223, 226, 229,  
    231, 240  
Cardinaletti, Anna, 165, 167  
Casalicchio, Jan, 10, 37  
Cavaioli, Frank J., 18  
Cayetano Choque, Miriam, 109  
Cerrón-Palomino, Rodolfo, 82  
Cenni, Franco, 17  
Čermák, Radek, 288  
Cerrón-Palomino, Rodolfo, 82, 107,  
    108  
Chambers, Jack K., 306  
Chang, Lidia, 83  
Chappell, Whitney, 79  
Chromik, Bartłomiej, 215, 217, 267–  
    271, 275  
Cilliers, Paul, 219  
Clements, Nick, 232  
Cole, Peter, 108  
Coler, Matt, 81, 82, 85, 88, 97–99, 102,  
    103, 107  
Condemi, Filippo, 137  
Cordin, Patrizia, 37  
Córdova, Gavina, 108  
Cornips, Leonie, 51, 307  
Costa, James, 130, 131  
Coulson, Michael, 286  
Crespo del Río, Claudia, 82  
Crevels, Emily I., 82  
Cubberley, Paul, 220, 238, 239, 241,  
    243–245  
  
Czerwionka, Lori, 105  
D’Alessandro, Roberta, 10, 39  
Dahl, Östen, 215–217  
Dalton-Puffer, Christiane, 324  
Dankel, Philipp, 82–85, 89, 91, 94–96,  
    101, 105  
de Boor, Helmut, 221–226, 228, 231  
De Decker, Paul, 191  
De Fina, Anna, 19  
De Haan, Germen J., 193  
De Schutter, Georges, 356, 375, 379  
de Tschudi, Johann Jakob, 320  
De Vogelaer, Gunther, 359, 360, 362–  
    365, 376  
De Vries, Nic J., 191  
De Wulf, Chris, 355, 357–359  
Defensoria del Pueblo, 107  
DeGraff, Michel, 214  
Dejna, Karol, 238, 245  
DeLancey, Scott, 88  
Demenko, Grażyna, 239  
Demske, Ulrike, 311, 316  
DeSchutter, Georges, 240  
Deutscher, Guy, 216, 217, 219  
Devos, Magda, 359, 360, 364, 365, 376  
Diez, Friedrich Christian, 39  
Dillard, Joey Lee, 306  
Długosz-Kurczabowa, Krystyna,  
    238, 245  
Dodd, Bill, 218, 221–223, 226  
Dogil, Grzegorz, 241  
Dolatowski, Marek, 217  
Donaldson, Bruce, 240, 244  
Dorian, Nancy, 75, 131, 192, 213, 214,  
    346, 347, 377  
Dressler, Wolfgang U., 345, 346, 377  
Drinka, Bridget, 79  
Droogers, André, 314

- Dubisz, Stanisław, 238, 242, 245  
Dukiewicz, Leokadia, 246  
Dziubalska-Kołaczyk, Katarzyna, 245
- Edmonds, Bruce, 218  
Ehala, Martin, 64  
Eisenberg, Pater, 218, 221–224, 226, 228, 233, 235  
Elspaß, Stephan, 191  
Escobar, Anna María, 82  
Eshuis, Margje, 342–344, 354  
Estrada Fernández, Zarina, 61  
Evers, Arnold, 313
- Fagan, Sarah, 218, 221–224, 226–230, 233, 235, 240  
Falcone, Giuseppe, 123  
Faller, Martina, 82, 94  
Fanciullo, Franco, 123  
Fase, Willem, 346  
Faßke, Helmut, 285  
Feitsma, Tony, 193  
Feke, Marilyn Suzanne, 83  
Fellin, Luciana, 19  
Ferrari, Andrea, 16  
Feyér, Balint, 352  
Field, Thomas T., 346  
Filip, Elżbieta Teresa, 270, 272, 276  
Filipović, Luna, 213, 214  
Fishman, Joshua A, 126  
Fleischer, Jürg, 144  
Flores Farfán, José Antonio, 62, 75  
Foley, James, 232  
Fortson, Benjamin W., 285–287  
Fourquet, Jean, 229  
Fox, Anthony, 218, 221–224, 226–228, 231, 233, 235  
Franceschetto, Cilmar, 320
- Frasson, Alberto, 10, 37  
Frings, Theodor, 305
- García Tesoro, Ana Isabel, 82  
Gell-Mann, Murray, 215, 217  
Gibson, Charles, 59  
Goblirsch, Kurt, 221, 229, 231  
González Luna, Ana María, 61  
Grageda Bustamante, Arón, 61  
Grández Ávila, Magaly, 105  
Granzow, Klaus, 311  
Grinevald, Colette, 75, 214  
Guardiano, Cristina, 39  
Gumperz, John J., 306  
Gusinde, Konrad, 241, 242, 245  
Gussmann, Edmund, 238–245
- Haboud, Marleen, 82  
Haegeman, Liliane, 359  
Haider, H., 313  
Haimovich, Gregory, 104  
Halevy, Rivka, 177  
Hall, Tracy, 221–226  
Hall, Tracy Alan, 221, 228, 233, 235  
Haller, Hermann W., 19, 20, 42  
Hamann, Silke, 240, 241  
Hammarström, Harald, 214, 217, 219  
Harbert, Wayne, 220, 221, 231–233, 240, 243, 244, 246, 247  
Hardman, Martha J., 82, 96  
Harshav, Benjamin, 144  
Haspelmath, Martin, 316  
Haupt, Leopold, 285, 289, 292, 294  
Heath, Shirley Brice, 61, 62  
Heine, Bernd, 215  
Hennings, Thordis, 221, 230, 231  
Herzog, Marvin, 240  
Hill, Jane H., 72  
Hill, Kenneth C., 72

## *Name index*

- Hilton, Nanna Haug, 194, 195, 201–203, 208  
Hinskens, Frans, 306  
Hoekstra, Erik, 193  
Hoekstra, Jarich, 240, 244, 308, 310, 313, 320  
Hoekstra, Teun, 313, 317  
Höhle, Tilman N., 310  
Hoops, Johannes, 290  
Horn, Alexander, 192  
Howard, Rosaleen, 82  
Huddleston, Rodney, 177  
Ingham, Richard, 324  
Isaac, Barry L., 60–62  
Iverson, Gregory, 221, 231, 247  
Jacobs, Neil G., 145, 147–150, 154, 156, 160–162, 172, 175, 176, 178, 220, 240, 244  
Janaš, Pětr, 285  
Jankuhn, Herbert, 290  
Jarvis, Scott, 84  
Jassem, Wiktor, 238–241  
Jeanneret, René, 2  
Jenč, Rudolf, 288  
Jessen, Michael, 229, 231, 247  
Johanson, Lars, 83, 84, 86, 106  
Johnson, Sally, 218, 221–224, 226–228, 230  
Juvonen, Päivi, 214  
Kahn, Lily, 146  
Kalt, Susan E., 83  
Karanastasis, Anastasios, 137  
Karaś, Mieczysław, 240  
Katsoyannou, Marianna, 137  
Katsoyannou, Marianne, 135  
Katz, Dovid, 149, 150, 162, 174–176, 178  
Kaufman, Terence, 324  
Kayne, Richard S., 313  
Kayser, Wolfgang, 289  
Keating, Patricia, 241  
Kijak, Artur, 239, 245  
Klausmann, Hubert, 320  
Kleczkowski, Adam, 217, 221, 223–230, 241, 242, 244, 245, 274  
Klee, Carol A., 82  
Kloss, Heinz, 263  
Kluge, Friedrich, 292  
Knauthe, Christian, 298  
Kolly, Marie-José, 191, 199  
Koster, Jan, 313  
Krefeld, Thomas, 320  
Krogh, Steffen, 146, 172, 175, 176, 180  
Król, Tymoteusz, 215, 219, 221, 223–225, 227–231, 274, 276  
Kučera, Henry, 233, 246  
Kusters, Wouter, 214–216, 219  
Kuteva, Tania, 110, 215  
Kuznitz, Cecile Esther, 144  
Labov, William, 305  
Ladefoged, Peter, 241  
Landry, Rodrigue, 64  
Lasatowicz, Maria K., 220–223, 225, 227, 231  
Lass, Roger, 243  
Łatosiński, Józef, 223, 228, 230  
Laude, Robert, 321  
Ledgeway, Adam, 39  
Lee, Tae Yoon, 82  
Leemann, Adrian, 191, 192, 199, 243  
Lehmann, Christian, 95  
Leivada, Evelina, 14  
Leonetti, Manuel, 39  
Leskién, August, 287  
Lewaszkiewicz, Tadeusz, 285

- Lind, Fabienne, 192  
Lindström, Eva, 216  
Lipski, John, 80  
Llanos, Primitivo Nina, 82  
Lloyd, Seth, 215, 217  
Llwyd, Alan, 289, 291  
Lockhart, James, 59  
Louden, Mark, 318, 325
- Macha, Jürgen, 240  
Maddieson, Ian, 241  
Madejowa, Maria, 240  
Maiello, Giuseppe, 288  
Mamani Lopez, Marcelino, 110  
Manley, Marilyn S., 82  
Mannheim, Bruce, 82  
Manzini, Maria Rita, 39  
Mark, Yudel, 173  
Marquis, Yan, 130, 131  
Martínez Vera, Gabriel A., 82, 88  
Martino, Paolo, 121, 124, 126, 130, 137  
Masor, Alyssa, 175, 176  
Matras, Yaron, 214  
Mattheier, Klaus J., 264  
Matthews, Stephen, 214  
Maurizio, Roxana, 15  
May, Robert, 313  
Mazin, R., 174  
McWhorter, John, 214, 217  
Meakins, Felicity, 213–215  
Mendoza, José G, 82  
Meo Zilio, Giovanni, 16  
Merma-Molina, Gladys, 82  
Messing, Jacqueline, 63  
Mesthrie, Rajend, 213, 214  
Mętrak, Maciej, 266  
Meyer-Lübke, Wilhelm, 39  
Meyerhoff, Miriam, 190  
Michalk, Siegfried, 285
- Miestamo, Matti, 215–217, 219  
Mitchell, Melanie, 217, 218  
Mojmir, Herman, 221, 223–230, 241  
Möller, Robert, 191  
Monroe, George K., 233, 246  
Moosmüller, Sylvia, 229, 231  
Moravcsik, Edith A., 39  
Morciniec, Norbert, 217, 267, 276  
Morris-Jones, John, 291  
Moseley, Christopher, 215  
Mühlhäusler, Peter, 214, 215  
Müller, R., 320, 327  
Muntzel, Martha C., 346  
Mustanoja, Tauno F., 317, 324  
Muysken, Pieter, 48, 82, 108, 346  
Muysken, Pieter C., 82
- Nagy, Naomi, 190  
Narrog, Heiko, 105  
Nedo, Pawoł, 285, 288, 289, 292, 293,  
299  
Neeleman, Ad, 164  
Neels, Rinaldo, 215, 276  
Nesvig, Martin, 59  
Nevins, Andrew, 2  
Newmeyer, Frederick J., 218  
Nicholas, Alexandre, 215  
Nicholas, Joe, 192  
Niebaum, Hermann, 240  
Nitsch, Kazimierz, 242  
Noglo, Kossi, 190  
Nota, Amber, 194  
Nove, Chaya, 154  
Nunes, Jairo, 328  
Nutini, Hugo H., 60–62  
Nycz, Jennifer, 191
- O'Brien, Mary, 218, 221–224, 226,  
228, 230, 233

## *Name index*

- O'Grady, William, 35  
O'Rourke, Bernadette, 130, 131  
O'Rourke, Bernadette, 192  
O'Shannessy, Carmel, 190  
Ocampo, Alicia M., 82  
Okrutny, Iosef, 161  
Olbertz, Hella, 83, 94  
Olko, Justyna, 75, 217, 267, 269, 270,  
    272, 276  
Orzechowska, Paula, 245, 246  
  
Paardekooper, P. C., 364, 376, 377  
Padgett, Jay, 241  
Pagel, Steve, 85, 91, 101, 105  
Palacios, Azucena, 82  
Palosaari, Naomi, 213, 214  
Parkvall, Mikael, 214, 215, 217  
Parodi, Claudia, 59  
Pascual y Cabo, Diego, 30  
Paul, Hermann, 221–226, 228–231,  
    234, 235  
Pavlenko, Aneta, 84  
Peliti, Luca, 218  
Pellegrino, Manuela, 2, 121, 125–127,  
    130, 136, 137  
Peralta Zurita, Elvira, 107, 108  
Pesetsky, David, 304, 313, 326, 334  
Petropoulou, Christina, 128, 137  
Pfänder, Stefan, 82–84, 110, 111  
Pipyrou, Stavroula, 126, 130  
Plaza Martínez, Pedro, 107  
Poletto, Cecilia, 37, 51  
Polinsky, Maria, 35, 36  
Politi, Antonio, 218  
Postma, Gertjan, 307, 313, 314, 334,  
    344, 379  
Preston, Dennis, 190  
Preston, Laurel B., 218  
Protze, Helmut, 264  
  
Putschke, Wolfgang, 215  
Pütz, Martin, 213, 214  
Py, Bernard, 2  
  
Quartararo, Geraldine, 82  
Quelca Huanca, Heriberto, 91  
  
Ramallo, Fernando, 130, 131, 192  
Raupp, Jan, 290  
Rawp, Jan, 288, 289, 293  
Reinhart, Tanya, 164, 308  
Reinke, Kristin, 18, 19  
Rescher, Nicholas, 218  
Reuland, Eric, 308  
Ritchie, Carlo J. W., 215, 220, 222, 267  
Rizzi, Luigi, 37, 43  
Roos, Ton, 342–344, 354  
Rosenberg, Peter, 305, 306, 346, 348  
Rospond, Stanisława, 241  
Rothstein, Robert, 220, 239, 241, 244  
Rothwell, William, 324  
Rottet, Kevin J., 345, 347  
Rowicka, Grace J., 239  
Rubach, Jerzy, 239  
Russ, Charles V. J., 218, 221, 222, 224,  
    226, 228, 229, 231, 233, 235,  
    240  
Ruys, Eddy G., 164  
Ryckeboer, Hugo, 267, 276  
Rys, Kathy, 352, 354, 359, 368, 372,  
    375  
  
Saab, Andres, 39  
Sabel, Joachim, 313  
Sadock, Benjamin, 175, 176  
Sakel, Jeanette, 214  
Sallabank, Julia, 130, 131, 213, 214  
Salmons, Joseph, 221, 231, 247  
Sánchez, Liliana, 82

- Sandoval Arenas, Carlos O., 63  
Sasse, Hans-Jürgen, 134, 345, 346  
Savoia, Leonardo Maria, 39  
Schaarschmidt, Gunter, 287, 293  
Schabus, Wilfried, 321  
Schaffel Bremenkamp, Elizana, 314, 325, 344–346, 350, 353, 378  
Schaffel, Elizana, 343–345, 349  
Schlegel, Matthias, 296, 297  
Schmidt, Hans, 222–226, 243  
Schnibben, Frank, 321  
Schumacher de Peña, Gertrud, 82  
Schuster-Šewc, Heinz, 296, 297  
Schwaller, John, 59  
Sebregts, Koen, 205  
Seibel, Ivan, 325  
Seiler, Guido, 225, 231, 319  
Selmer, Carl, 242, 243  
Seuren, Pieter A. M., 214  
Shalizi, Cosma, 217, 218  
Shameem, Nikhat, 192  
Shing, Han-Chin, 192  
Shtiebel, Kave, 171, 172, 174  
Silva-Corvalán, Carmen, 84  
Siegel, Jeff, 214  
Sierra Silva, Pablo Miguel, 61  
Silva-Corvalán, Carmen, 213, 214  
Silverstein, Michael, 43  
Simmler, Franz, 221–225, 228, 229, 231  
Sinnemäki, Kaius, 215–217  
Sjölin, Bo, 193  
Slaviūnas, Zenonas, 289, 294  
Slobin, Dan I., 84, 106, 111  
Smith, John Charles, 39  
Smoleř, Jan Arnošt, 285, 289, 292, 294  
Sorgini, Luana, 10  
Soto Rodríguez, Mario, 82, 84, 89, 90, 94, 96, 109, 113  
Sotto-Santiago, Sylk, 2  
Spencer, Andrew, 241  
Squillaci, Maria Olimpia, 124, 126, 129, 134, 137  
Stadthagen-González, Hans, 35, 36  
Stamuli, Francesca, 124  
Stanford, James N., 190  
Starke, Michal, 165, 167  
Steemers, Patrick, 194  
Stefan, Nika, 194, 203  
Steiner, B. Devan, 324  
Stern, Dieter, 263  
Stevenson, Patrick, 240  
Stieber, Zdzisław, 241  
Stirling, Lesley, 177  
Stone, Gerald, 282, 288, 297, 299  
Stowell, Tim, 317  
Stratford, Dale, 82  
Stritzel, Herbert, 321  
Strutyński, Janusz, 238–241, 244  
Sussex, Roland, 220, 238, 239, 241, 243–245  
Swiggers, Pierce, 214  
Taeldeman, Johan, 307, 354, 359, 372–375  
Taylor, Gerald, 82  
Terenghi, Silvia, 10  
Thelander, Mats, 306  
Thomason, Sarah, 324  
Práinsson, Höskuldur, 240  
Tierstna, Peter M., 240, 244  
Torrego, Esther, 304, 313, 326, 334  
Tressmann, Ismael, 314, 344  
Trudgill, Peter, 214, 215, 306  
Trunte, Hartmut, 287

## *Name index*

- Urbańczyk, Stanisław, 239, 242, 244, 245  
Urton, Gary, 82  
Van Bezoijken, Renee, 194, 201, 205–208  
van Bree, Cor, 193  
van Coetsem, Frans, 305  
Van Craenenbroeck, Jeroen, 359  
Van de Craats, Ineke, 305  
van de Weijer, Joost M., 239  
van den Broecke-de Man, E. J., 353, 354, 359, 362  
van der Hoek, Michel, 221, 230, 244  
van der Torre, Erik-Jan, 352  
van Gelderen, Elly, 318  
van Koppen, Marjo, 193, 359  
van Ness, Silke, 240  
van Oostendorp, Marc, 221, 233–235, 246  
Van Reenen, Piet, 352  
Vanden Wyngaerd, Guido, 313  
Vaux, Bert, 191  
Veith, Werner, 233  
Velasco Murillo, Dana, 61  
Velasco Rojas, Pedro, 107  
Velupillai, Viveka, 214, 215  
Versieck, Sabina, 354, 370–372, 374  
Versloot, Arjen, 368  
Villata, Bruno, 18, 26  
Villavicencio Zarza, Frida, 61  
Visser, F. Th., 317  
von Eckhart, Johann Georg, 290  
Von See, Klaus, 290  
Von Tschudi, Johann Jakob, 343  
von Unwerth, Wolf, 242, 244, 245  
Vulpiani, Angelo, 218  
Wągiel, Marcin, 238, 239, 243  
Wagner, Richard Ernst, 269  
Walde, Martin, 285  
Waniek, Gustav, 242  
Weckwerth, Jarosław, 220, 222, 223, 225  
Weddige, Hilkert, 231  
Weening, Joke, 194, 201–203, 208  
Weinreich, Max, 240  
Weinreich, Uriel, 154, 156, 190, 274, 275  
Wekker, Herman C., 214  
Wenker, Georg, 191  
Werner, Eduard, 282  
Werner, Edward, 282  
Weyhe, Eivind, 244  
Wicherkiewicz, Tomasz, 215, 217, 220–223, 225, 227–229, 231, 263, 267, 269, 270, 272, 275, 276  
Wierzchowska, Bożena, 241  
Wiese, Richard, 218, 221, 222, 224, 226, 228, 231, 233, 235  
Wiesinger, Peter, 215, 217, 263  
Will, George, 362  
Wilson, Robert, 306  
Winkler, Johan, 370, 371, 373–375  
Wisniewski, Roswitha, 221–226, 228, 231  
Woehrling, Jean-Marie, 3  
Wolf, Meyer, 147, 156  
Wrede, Fred, 310  
Wright, Joseph, 221–226, 228, 229, 231  
Wright, Susan, 109  
Yannanakis, Yanna, 59  
Yelin, Boris, 105  
Zabrodskaja, Anastassia, 64

Żak, Andrzej, 217, 220, 227, 241–245,  
276

Zariquiey, Roberto, 108

Zieniukowa, Jadwiga, 217, 220–223,  
225

Zimmermann, Klaus, 60, 61

Zydrorowicz, Paulina, 245, 246

Zygis, Marzena, 241





# Contemporary research in minoritized and diaspora languages of Europe

This volume provides a collection of research reports on multilingualism and language contact ranging from Romance, to Germanic, Greco and Slavic languages in situations of contact and diaspora. Most of the contributions are empirically-oriented studies presenting first-hand data based on original fieldwork, and a few focus directly on the methodological issues in such research. Owing to the multifaceted nature of contact and diaspora phenomena (e.g. the intrinsic transnational essence of contact and diaspora, and the associated interplay between majority and minoritized languages and multilingual practices in different contact settings, contact-induced language change, and issues relating to convergence) the disciplinary scope is broad, and includes ethnography, qualitative and quantitative sociolinguistics, formal linguistics, descriptive linguistics, contact linguistics, historical linguistics, and language acquisition. Case studies are drawn from Italo-Romance varieties in the Americas, Spanish-Nahuatl contact, Castellano Andino, Greko/Griko in Southern Italy, Yiddish in Anglophone communities, Frisian in the Netherlands, Wymysioryś in Poland, Sorbian in Germany, and Pomeranian and Zeelandic Flemish in Brazil.

ISBN 978-3-96110-404-8



9 783961 104048