

Interpreting language-learning data

Edited by

Amanda Edmonds

Pascale Leclercq

Aarnes Gudmestad

Eurosla Studies



No series description provided

Interpreting language-learning data

Edited by

Amanda Edmonds

Pascale Leclercq

Aarnes Gudmestad


Edmonds, Amanda, Pascale Leclercq & Aarnes Gudmestad (eds.). 2020.
Interpreting language-learning data (Eurosla Studies). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/278>

© 2020, the authors

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

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

ISBN: no digital ISBN

no print ISBNs!

ISSN: 2626-2665

no DOI

Source code available from www.github.com/langsci/278

Collaborative reading: paperhive.org/documents/remote?type=langsci&id=278

Cover and concept of design: Ulrike Harbort

Fonts: Libertinus, Arimo, DejaVu Sans Mono, Source Han Serif

Typesetting software: Xe_{La}TeX

Language Science Press

Xhain

Grünberger Str. 16

10243 Berlin, Germany

langsci-press.org

Storage and cataloguing done by FU Berlin

Freie Universität



Berlin

Contents

1	Introduction: Reflecting on data interpretation in SLA	1
	Amanda Edmonds, Pascale Leclercq & Aarnes Gudmestad	
2	L2 acquisition in a rich dialectal environment: some methodological considerations when SLA meets dialectology	9
	Linda Evenstad Emilsen & Åshild Søfteland	
3	Comparing ERPs between native speakers and second language learners: Dealing with individual variability	39
	Maud Péliissier	
4	Replication: Measuring the influence of typologically diverse target language properties on input processing at the initial stages of acquisition	71
	Marzena Watorek, Rebekah Rast, Xinyue Cécilia Yu, Pascale Trévisiol, Hedi Majdoub, Qianwen Guan & Xiaoliang Huang	
5	On the relationship between epistemology and methodology: A reanalysis of grammatical gender in additional-language Spanish	111
	Aarnes Gudmestad	
6	Analysing interaction in primary school language classes: Multilevel annotation and analysis with EXMARaLDA	141
	Heather E. Hilton & John Osborne	
7	Transcribing interlanguage: The case of verb-final [e] in L2 French	171
	Pascale Leclercq	
8	Potential pitfalls of interpreting data from English-French tandem conversations	199
	Sylwia Scheuer & Céline Horgues	
	Index	235

Chapter 1

Introduction: Reflecting on data interpretation in SLA

Amanda Edmonds

Université Paul-Valéry Montpellier 3

Pascale Leclercq

Université Paul-Valéry Montpellier 3

Aarnes Gudmestad

Virginia Polytechnic Institute and State University

The past decade has seen a growing number of publications that urge researchers in the field of second language acquisition (SLA) to engage more directly and more critically with questions of research methodology. These include, among many others, [Plonsky \(2014\)](#), who makes clear recommendations for quantitative second-language (L2) research and issues a call for change, [Leclercq et al. \(2014\)](#), who call for more transparency in the assessment of L2 proficiency, [Marsden et al. \(2016\)](#), who make a strong case for the importance of replication in moving the field forward, [Gudmestad & Edmonds \(2018\)](#), who showcase different ways to bring critical reflections on method to the fore, and [Ortega \(2014\)](#), who draws attention to the need to move beyond a native-speaker bias in L2 research. Although diverse in aim and scope, these endeavours and others like them share a strong interest in moving methodological practices forward. [Byrnes \(2013: 825\)](#) goes so far as to characterise this increasing interest as a “methodological turn” within our field. SLA research that has come out of this turn has led to numerous advances. To take but a few examples, underlying concepts and constructs have been (re)defined (e.g., [Pallotti 2009](#) on the construct of complexity-



Amanda Edmonds, Pascale Leclercq & Aarnes Gudmestad

accuracy-fluency), certain well-established ways of doing things have been questioned (e.g., [Plonsky & Oswald's](#) (2017) plea to move away from ANOVA), and new approaches have been developed and championed (see the numerous recent special issues devoted to both wide and narrow methodological issues: [Norris et al. 2015](#); [Choi & Richards 2016](#); [De Costa et al. 2017](#); [Edmonds et al. forthcoming](#)). As a result, the methodological landscape in SLA is arguably more diverse than ever before, with [Ortega \(2013: 5\)](#) identifying the increase in “research methodological prowess” as one of the noticeable trends in SLA research.

According to [King & Mackey \(2016: 214\)](#), the field of SLA is in its prime. It has left behind the largely unproductive, so-called ‘paradigm wars’ between those supporting quantitative and qualitative approaches. Both cognitively and socially oriented researchers are showing greater awareness of the importance of incorporating a range of perspectives. The field is pushing methodological boundaries in many directions.

The pushing referred to by King and Mackey is taking many forms, including cross-disciplinary pollination and collaboration (e.g., [Duff & Byrnes 2019](#)), the use of mixed-methods, leading to an attitude of “methodological inclusivity” ([Römer 2019: 478](#)), and a growing number of scientific publications that tackle methodological issues head on. In this final category, researchers generally aim to stimulate discussion and potentially initiate change, be this through discussion papers, such as [Group \(2016\)](#) or [Young \(2018\)](#), or with empirical studies (often through reanalyses of previous published data or meta-analyses), which serve to concretely demonstrate the import and impact of methodological choices ([Santos et al. 2008](#); [Leeser & Sunderman 2016](#); [Edmonds & Gudmestad 2018](#); [Solon 2018](#)).

With the current collected volume, we aim to contribute to this focus on methodological issues. Specifically, we bring together a collection of seven chapters, each of which provides a new angle on the treatment or interpretation of language-learning data, a crucial issue in the building of knowledge in the field of SLA. Three main lines of reflection are pursued in these chapters.

The first concerns the question of how comparisons to a baseline norm can be carried out in L2 research, as well as what norm might be best adopted. In the present volume, this question is addressed from two novel standpoints: the question of how to identify interlanguage forms in the dialect-rich environment of Norway, which provides many different input forms for the same concept (Evenstad Emilsen & Søfteland) and the questioning of a general native baseline in event-related potential (ERP) studies (Pélessier).

The second line of reflection, broadly speaking, concerns epistemological stance in research design. By epistemological stance, we refer to a researcher’s view

1 Introduction: Reflecting on data interpretation in SLA

about what constitutes knowledge in a given field. One common epistemological tension in the field of SLA opposes two visions of language learning: “Is learning like acquiring stuff or is it like doing things?” (Young 2018: 45). These two visions lead to different positions on how to study language learning and even as regards to what is ultimately worthy of study. Issues connected to the role of epistemological stance are visible in two chapters. Whereas Watorek, Rast, Yu, Trévisiol, Majdoub, Guan & Huang reflect on how to carry out a conceptual replication, thereby holding constant their epistemological understanding of the phenomenon under study (namely, L2 acquisition from the very initial stages), Gudmestad purposefully sets out to follow to its logical conclusion a shift in epistemological stance.

The third question grapples with more technical issues surrounding the annotation, coding, and interpretation of data, especially when faced with ambiguous interlanguage forms. Issues identified involve both multimodal data (Hilton & Osborne; Scheuer & Horgues) and difficulties specific to the transcription of oral data (Leclercq).

In the first chapter, Evenstad Emilsen & Søfteland offer reflections on SLA in a dialect-rich environment. Such environments have received little explicit attention in the research literature, and yet they entail challenges for both learners and for researchers. For learners, the co-existence of multiple dialects provides an arguably more complex input, one in which numerous forms co-exist to express the same function. For researchers, making data coding and analysis decisions about learner production is particularly challenging, as forms found in interlanguage use may not correspond to the dominant dialect, but may be present in other varieties. The authors detail the challenges facing researchers, providing several examples. They highlight the difficulties inherent in determining whether forms produced by learners are evidence of sociolinguistic variation (i.e., variation present in the input) or instances of interlanguage variation.

Pélissier’s contribution questions the comparison of native and non-native performance in online processing studies involving ERPs. The author shows that although a large body of research into native language processing has identified a biphasic ERP pattern when native speakers are asked to process syntactic violations, recent research has called this pattern into question, showing instead that there is substantial inter-individual variability among native speakers. More specifically, when it comes to syntactic violations, most individuals show only one of the two components of the biphasic pattern. For the field of SLA, traditionally preoccupied with comparing native and non-native performance, this finding begs the question of how we might meaningfully compare learners and

Amanda Edmonds, Pascale Leclercq & Aarnes Gudmestad

native speakers. Pélissier explores two approaches that hold some promise insofar as they allow researchers to account for individual variability: the Response Magnitude Index and the Response Dominance Index (Tanner et al. 2014). The target structure in Pélissier's study is past-tense morphology with auxiliaries in English. Results show that the Response Dominance Index, but not the Response Magnitude Index, is useful in accounting for the data analysed.

In the third chapter, Watorek and colleagues provide a detailed presentation of the ambitious VILLA project (*Varieties of Initial Learners in Language Acquisition: Controlled classroom input and elementary forms of linguistic organisation*). This project seeks to provide insight into language acquisition in the first hours of exposure to a new language. In the original VILLA project, Polish is the target language, with learners having either Dutch, English, French, German or Italian as their native language. The contribution included in this volume reflects on three conceptual replications of the VILLA project, in order to study the initial acquisition of Modern Standard Arabic, Mandarin Chinese, and Japanese by native French speakers. The goal of the conceptual replications is to contribute additional insight into language learning starting from first exposure, but with typologically diverse languages. This diversity requires the authors to reconsider the target of learning (nominal morphology, in the original project), the variables controlled for (transparency and frequency), as well as the way of assessing learning. The reflection offered by the authors raises the intriguing question of comparability when transposing research design and questions to study new language combinations.

Gudmestad's chapter directly addresses the oft-ignored issue of epistemological stance. In other words, she engages with "what counts" as knowledge in SLA. Using the concrete example of grammatical gender in L2 Spanish, she highlights the fact that there exist (at least) two different epistemological understandings as to what production of gender-marked modifiers reveals about interlanguage. One position (exemplified in Gudmestad's previous work) considers all instances of gender marking to reveal the same underlying process, regardless of whether the modifier in question is an adjective or a determiner. The second position sees two different processes at work: on the one hand, the gender marked on determiners is thought to reflect the gender attributed by the speaker to the noun in question (a lexical property) and, on the other, gender marked on adjectives reveals the speaker's ability to compute morphosyntactic agreement. In her chapter, Gudmestad departs from her original stance in order to "try on" the second position in a reanalysis of data originally published in Gudmestad et al. (2019). She thereby explores what is gained by adopting new ways of seeing data. In so doing, Gudmestad essentially participates in what King & Mackey (2016: 214)

1 Introduction: Reflecting on data interpretation in SLA

term “layering”: “Layering involves considering theory as well as practice and, in particular, considering varied epistemological stances every time one looks at a traditional problem.”

The next chapter provides a concrete and critical reflection on how the tool EXMARaLDA can be profitably used to carry out multi-tiered annotation of classroom data. Hilton and Osborne report on part of an exploratory study that took place in English classes held in two French elementary schools. After detailing the development of their multi-layered approach to transcribing and annotating three weeks of language lessons, the authors focus on data from one lesson from each classroom in order to demonstrate how conducting analyses at different levels of annotation may lead to the identification of the differences in the two learning environments that triggered different learning outcomes for the students (regarding memorization of new vocabulary and utterance construction). Although the authors highlight that the analyses are limited in scale and thus cannot be used to suggest pedagogical implications, they demonstrate that the two classrooms are not equally effective, which is visible, for example, in the organisation of pupil and teacher talk.

Chapter six focuses on how theory, data coding, and data transcription intersect. To accomplish this goal, Leclercq uses the example of verb-final [e] in L2 French. Verb-final [e] in French can correspond to the infinitive form (*parler* ‘to speak’), imperfective forms (e.g., *parlais* ‘(you) speak’, *parlait* ‘(s/he) speaks’), the first-person simple past form (*parlai* ‘(I) spoke’), and various forms of the past participle (*parlé*, *parlés*, *parlée*, *parlées*). In other words, one spoken form – [paʁle] – is highly homophonous. This leads to a clear challenge for any researcher working on oral productions in L2 French. How does one transcribe a form like [paʁle] when produced by a learner? Leclercq takes up this thorny issue and critically details how other studies in SLA research have dealt with it. She concludes by showing that some transcription choices result from a premature categorisation of the data, often reflecting theoretical positioning and potentially introducing interpretative bias.

The volume closes with a chapter devoted to identifying and reflecting on potential pitfalls involved in analysing data from English-French tandem conversations. Scheuer and Horgues report on data collected from 21 tandem pairs during a semester-long programme at a French university. Each tandem is made up of a native speaker of French and of English and was recorded on two occasions (once at the beginning and once at the end of the semester). For each recording, approximately half of the speaking time is in each of the two languages. The authors use these data to explore corrective feedback and communication

Amanda Edmonds, Pascale Leclercq & Aarnes Gudmestad

breakdowns, addressing, among other things, which member of the tandem initiated the feedback or signalled the breakdown and what type of issue (lexis, pronunciation, syntax, etc.) led to the feedback or breakdown. The authors offer a thought-provoking discussion on the difficulties involved in determining both what constitutes corrective feedback or comprehension breakdowns and in pinpointing what linguistic issue was the cause (or causes) for either. They thus provide clear and concrete examples of dealing with ambiguity in learner data in an L2 analysis.

The seven chapters brought together in this volume offer original and timely contributions on the role of (native-speaker) norms in L2 analyses, on the impact of epistemological stance, and on the challenges of transcription and annotation of language-learning data. In addressing these issues, the researchers rely on a variety of methodological practices and highlight in their chapters the import of methodological choice. These choices have a far-reaching impact, as they constrain and orient what observations can be made in research and what conclusions are ultimately drawn. We hope to have demonstrated with this volume that reflecting on these decisions – making them explicit and holding them up to study – is indeed a valuable enterprise.

References

- Byrnes, Heidi. 2013. Notes from the editor. *The Modern Language Journal* 97(4). 825–827.
- Choi, Seongsook & Keith. (eds.) Richards. 2016. *Innovation in research methods*. Applied Linguistics [Special issue], 37(1).
- De Costa, Peter, Hima I. Rawal & Irina Zaykovskaya (eds.). 2017. *Study abroad in contemporary times: Toward greater methodological diversity and innovation*. System [Special issue], 71.
- Duff, Patricia A. & Heidi Byrnes. 2019. SLA across disciplinary borders: New perspectives, critical questions, and research possibilities. *The Modern Language Journal* 103 (Suppl. 1).
- Edmonds, Amanda & Aarnes Gudmestad. 2018. Operationalizing variables: The case of future-time expression in L2 French. In Aarnes Gudmestad & Amanda Edmonds (eds.), *Critical reflections on data in second language acquisition*, 125–148. Amsterdam: Benjamins.
- Edmonds, Amanda, Pascale Leclercq & Nathalie Auger (eds.). Forthcoming. *Methodological challenges in SLA: A focus on French*. *Journal of French Language Studies* (30) [Special issue].

1 Introduction: Reflecting on data interpretation in SLA

- Group, The Douglas Fir. 2016. A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal* 100(Suppl.) 19–47. DOI: [10.1111/modl.12301](https://doi.org/10.1111/modl.12301).
- Gudmestad, Aarnes & Amanda Edmonds (eds.). 2018. *Critical reflections on data in second language acquisition*. Amsterdam: Benjamins.
- Gudmestad, Aarnes, Amanda Edmonds & Thomas Metzger. 2019. *Using variationism and learner corpus research to investigate grammatical gender marking in additional-language Spanish*. Language Learning.
- King, Kendall A. & Alison Mackey. 2016. Research methodology in second language studies: Trends, concerns, and new directions. *The Modern Language Journal* 100((Suppl.)). 209–227.
- Leclercq, Pascale, Amanda Edmonds & Heather Hilton (eds.). 2014. *Measuring L2 proficiency: Perspectives from SLA*. Clevedon, UK: Multilingual Matters.
- Leeser, Michael J. & Gretchen L. Sunderman. 2016. Methodological implications of working memory tasks for L2 processing research. In Gisela Granena, Daniel O. Jackson & Yucel Yilmaz (eds.), *Cognitive individual differences in second language processing and acquisition*, 89–104. Amsterdam: Benjamins.
- Marsden, Emma, Alison Mackey & Luke Plonsky. 2016. The IRIS repository: Advancing research practice and methodology. In Alison Mackey & Emma Marsden (eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages*, 1–21. New York: Routledge.
- Norris, John M., Steven J. Ross & Rob Schoonen (eds.). 2015. *Improving and extending quantitative reasoning in second language research*. Language Learning [Special issue], 65, S1.
- Ortega, Lourdes. 2013. *SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn*. Language Learning, 63(Suppl. 1). 1–24.
- Ortega, Lourdes. 2014. Ways forward for a bi/multilingual turn in SLA. In S. May (ed.), *The multilingual turn: Implications for SLA, TESOL, and bilingual education*, 32–53. New York: Routledge.
- Pallotti, Gabriele. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics* 30(4). 590–601.
- Plonsky, Luke. 2014. Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *Modern Language Journal* 98. 450–470.
- Plonsky, Luke & Frederick L. Oswald. 2017. Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition* 39. 579–592.

Amanda Edmonds, Pascale Leclercq & Aarnes Gudmestad

- Römer, Ute. 2019. Corpus research for SLA: The importance of mixing methods. In Viola Wiegand & Michaela Mahlberg (eds.), *Corpus linguistics, context and culture*, 467–481. Berlin: De Gruyter.
- Santos, Denise, Suzanne Graham & Robert Vanderplank. 2008. Second language listening strategy research: Methodological challenges and perspectives. *Evaluation & Research in Education* 21(2). 111–133.
- Solon, Megan. 2018. Novel sounds: What L2 phonetic data might be telling us that we do not always hear. In Aarnes Gudmestad & Amanda Edmonds (eds.), *Critical reflections on data in second language acquisition*, 89–124. Amsterdam: Benjamins.
- Tanner, Darren, Kayo Inoue & Lee Osterhout. 2014. Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition* 17(2). 277–293.
- Young, Richard F. 2018. Habits of mind: How do we know what we know? In Aek Phakiti, Peter De Costa, Luke Plonsky & Sue Starfield (eds.), *The Palgrave handbook of applied linguistics research methodology*, 31–53. New York: Palgrave Macmillan.

Chapter 2

L2 acquisition in a rich dialectal environment: some methodological considerations when SLA meets dialectology

Linda Evenstad Emilsen

Østfold University College

Åshild Søfteland

Østfold University College

This chapter discusses how interlanguage variation and dialectal variation in the target language appear homophonic in Norwegian. We demonstrate that this may pose challenges for the interpretation of second-language data. In societies with a high degree of variation in spoken vernaculars (or written norms), second-language learners are likely to be exposed to a great deal of variation and possibly conflicting features. The Norwegian language situation is a case in point: dialects have a neutral or high status and most people speak their local dialect in a variety of settings, both formal and informal. In this chapter, we review empirical and theoretical studies on second-language acquisition, focusing on the predictions they make for interlanguage variation. We then compare the findings of these studies to spontaneous speech data obtained from *The Nordic Dialect Corpus* and first-language studies of Norwegian. We demonstrate that it can be hard or impossible to distinguish between targetlike dialect variation and nontargetlike interlanguage variation. This has implications for the coding and interpretation of data. Our investigation seeks to raise awareness of the methodological issues related to differentiation between target-language variation and interlanguage variation and to stimulate further discussion on the topic.



Linda Evenstad Emilsen & Åshild Sjøteland

1 Introduction

One of the fundamentals in a lot of second-language (L2) research is the distinction between interlanguage (or nontargetlike) and targetlike variation (Gass & Madden 1985). In differentiating between targetlike and nontargetlike variation, it is imperative that any relevant comparisons be made with an appropriate baseline, if a comparison is needed. However, it is not always straightforward what the adequate baseline is.

Until recently, the norm in both second language acquisition (SLA) research and in additional-language teaching has been to compare bi- and multilingual speakers with an idealised first-language (L1) monolingual speaker. Even early on, researchers expressed concerns about the appropriateness of this (see for instance Bley-Vroman 1983; Klein 1998). Yet, the norm persisted until very recently. However, now even the concept of an “L1 monolingual speaker” is strongly contested and debated, and the L1 monolingual comparison is meeting strong criticism (see Group 2016 for an update on the debate).

One point of criticism against L1 monolinguals as a baseline for L2 acquisition, is that the concept of a monolingual speaker is an abstraction and idealisation. For instance, an L1 monolingual speaker is often associated with a standard language, and dialectal variation is not taken into account. Why this is problematic can be exemplified with the following: if an L2 English learner receives as input mostly a variety of Scottish English, that learner will start acquiring English based on the input received. It would be inadequate to compare the interlanguage of that learner exclusively to the grammar of an L1 speaker of Oxford English, as many aspects of the grammars in Scottish and Oxford English diverge. A comparison with Oxford English would exclude features that are present in the Scottish English input if they are not present in Oxford English, and vice versa: include features that are present in Oxford English even if they are not present in Scottish English. Needless to say, this is highly problematic from a scientific point of view.

In addition, language learners may receive input from several dialects at once, thus being exposed to potentially diverging linguistic systems. Input from different spoken varieties poses extra challenges in establishing both the exact input and the baseline.¹ It also makes it difficult to determine what grammatical features a language learner is expected to acquire. Input consisting of multiple

¹In this study, we will not discuss issues related to quantity of input, including how much input is needed for something to be acquired. We will leave this question open and set no threshold for the quantity of input. We take the stance that if something is present in the input, no matter to what extent, it is relevant to the current discussion.

2 L2 acquisition in a rich dialectal environment

varieties leads to ambiguity in output analyses, making it difficult to determine if an utterance is targetlike. This is an important methodological challenge related, fundamentally, to how we interpret all kinds of language acquisition/development data.

Variationist approaches to the acquisition of sociolinguistic variation deal with issues like these rather extensively (see for instance [Geeslin 2011](#)). However, this kind of a methodological challenge applies to research on L2 acquisition and bi- or multilingualism in any speech community characterised by a high degree of variation and goes beyond the boundaries of acquisition of sociolinguistic variation. The issue is also relevant to language teachers and others working with language assessment, as the differentiation between targetlike and nontargetlike is important in those contexts.

Even though challenges related to variation in the input apply to the entire field of SLA, it remains neglected in much research literature. Some studies investigate the L2 acquisition of dialects or of variation in the target language (TL) (see for instance [Geeslin & Gudmestad 2008](#); [Schmidt 2011](#); [Geeslin et al. 2012](#); [Rodina & Westergaard 2015a](#)). Much of the literature, however, does not address explicitly how variation in L2 learners' input affects the interpretation of L2 data. The main aim of the present study is to enhance the discussion on this topic and show that the issue is relevant for multiple research traditions; we aim to expand this discussion beyond variationist and sociolinguistic literature and into the whole field of SLA, focusing especially on grammatical aspects.

We seek to highlight methodological issues related to the presence of more than one variety in the input in additional-language acquisition. We do this by exploring one of the challenges caused by variation in the TL: empirical observations and theoretical approaches to SLA describe interlanguage variation that is coinciding with features regarded as characteristic of dialectal variation. In other words, we show that variation in L2 learners' grammars may look both like interlanguage and like TL variation, making it difficult, even impossible, to distinguish between the two analyses. By comparing TL dialects with interlanguage variation described by earlier studies on L2 acquisition (see [Section 3](#) for relevant references), we hope to demonstrate how complex the interpretation of linguistic data is when L2 learners are exposed to several varieties in the input.

Our study focuses on the Norwegian language situation. We compare spontaneous speech data from different dialects of L1 Norwegian excerpted from a spoken language corpus (*The Nordic Dialect Corpus* (NDC), [Johannessen et al. 2009](#)) with empirical observations and theoretical predictions about L2 interlanguage from SLA studies on Norwegian and other languages (see [Sections 3.1](#) and [3.2](#)).

Linda Evenstad Emilsen & Åshild Sjøteland

By examining how dialectal variation and interlanguage variation may coincide, the second aim of our study is to bridge the gap between SLA and dialectology.

Section 1.1 provides a brief note on terminology. Section 2 contains a description of the background for our study, focusing mainly on the Norwegian language situation (Section 2.1), the role of an idealised or monolingual norm in assessing L2 use and development (Section 2.2) and earlier research on target-language variation in SLA research (Section 2.3). Section 3 explores specific grammatical features described as interlanguage variation in the SLA literature that is homophonic with Norwegian dialect variation: Section 3.1 deals with morphology in the determiner phrase (DP) and Section 3.2 with finiteness and verb second (V2) constructions. Section 4 summarises the chapter and presents a few suggestions for addressing the methodological challenges identified.

1.1 A note on terminology

In our chapter, we seek to have a general approach: We do not focus on the order of acquisition of different languages, we do not distinguish between formal and informal learning, or between learning and acquisition. Further, we do not distinguish between bi- and multilingualism. We therefore use *L2 speaker/listener/learner* as an umbrella term for bi- and multilingualism and use the terms *learning* and *acquisition* interchangeably, unless otherwise specified.

For pragmatic reasons, we use the terms *variety* and *TL variation* to include all kinds of dialectal variation: geographically induced (*geolects*), sociolinguistic (*sociolects*) and spoken-language variation often described as multi-ethnolectal (*ethnolects*). Unless otherwise specified, we include all kinds of (oral) registers and inter- and intra-individual variation. In descriptions of the Norwegian language, the terms *dialect* and *spoken language variety* are both used to describe the same kinds of variation, and we will also use them as synonymous.

In general, we consider transcription, coding and other analyses as part of the *interpretation of data*. Still, our main focus here is the interface between coding and overall grammatical analyses, i.e. the interpretation of authentic utterances as targetlike or not.

2 Background

2.1 Language diversity: The Norwegian context

Norwegian is part of the Scandinavian dialect continuum, where dialects differ extensively in phonology, morphology and syntax, but are mutually intelligible

2 *L2 acquisition in a rich dialectal environment*

both inside and across national borders. Within Norway, most dialects have high or neutral status, and there is high acceptance for the use of dialects in most contexts – including the media, university lectures and parliament (Røyneland 2009; Sandøy 2009a). Most Norwegians would agree that it is important to keep using dialects (Røyneland 2009), and dialectal variation is officially recognised and protected in a variety of ways (e.g., Trudgill 2002: 31). One important language policy document is the “Dialect paragraph” (*Talemålsparagrafen*) in the School Law (Lovdata No date), introduced in 1878, stating that teaching should take place in the children’s own dialect. Hence, teachers have never been officially instructed to teach in a standard language, rather the contrary. The official phrasing today is that students and teachers can decide what spoken language variety to use, but that teachers and school leaders shall take the students’ dialects into consideration as much as possible (Lovdata, no date).

Norway has two official written standards (Bokmål and Nynorsk), but no official spoken standard. The Oslo dialect, which also is close to the written standard “Bokmål”, has a more neutral status than other varieties and could to some extent be considered an unofficial standard (Mæhlum 2009; Røyneland 2009). This variety is also the most common in oral media, and it is spreading in Southeast Norway (i.e., the Oslo circumference, Mæhlum 2009). In Norwegian sociolinguistic research, this spoken variety is often referred to as Standard Eastern Norwegian”, and we use this term in this chapter.² Nevertheless, local dialects have high status and are widely used, including on national TV and radio (Røyneland 2009; Sandøy 2009a). There is also a great deal of mobility in Norway, especially into the Oslo area, but also in other directions (Stjernholm 2013), and most people continue to speak their original dialect if they move to another part of the country (Jahr 1990: 7). Furthermore, many language learners will hear dialectal variation associated with multi-ethnolectal style, i.e. a dialect shared by people from several minority groups and some of their majority group friends (Svendsen & Røyneland 2008; Opsahl & Nistov 2010).³ The status of these varieties seems to be rising.

²The reader should still keep in mind that this is not an official standard. Also, the term “Standard Eastern Norwegian”, and the existence of a standard spoken language in Norway is disputed by researchers in Norwegian dialectology (cf. Mæhlum 2009 vs. Sandøy 2009b).

³Svendsen & Røyneland define *multi-ethnolect* and *ethnolect* as follow: “Whereas *ethnolects* might be conceived of as “varieties of a language that mark speakers of ethnic groups who originally used another language or distinctive variety” (Clyne 2000: 86), *multiethnolects* are characterised by their use by *several* minority groups “collectively to express their minority status and/or as a reaction to that status to upgrade it” (Clyne 2000: 87). When majority speakers come to share a multiethnolect with minorities, we see an expression of a new form of group identity” (Svendsen & Røyneland 2008: 64).

Linda Evenstad Emilsen & Åshild Søfteland

In summary, one must say that all learners of Norwegian will be exposed both to local dialects and to “standard Eastern Norwegian”, and most language learners will also encounter many dialects from other parts of the country and/or multi-ethnolects. This entails that the input for both L1 and L2 learners, children and adults, is characterised by variation. It is from this complex input learners of Norwegian develop the rules that make up their interlanguage grammar, and the kind of input they encounter is of course important for further language development (see 2.3 for more details).

Our work on L2 acquisition started with the project MultiCKUS – *MULTIlingual Children from Kindergarten to Upper Secondary school* (Arntzen 2012). This is a longitudinal research project following a group of L2 children from kindergarten to high school. The project consists of a variety of data, including spontaneous speech, where the children play or talk with each other and/or with a teacher or a researcher. The two authors of this chapter were especially responsible for developing a transcription and annotation standard for the spontaneous speech part of the project. This is when we first encountered examples like the ones we discuss in this chapter and we had to make explicit decisions about their interpretation.⁴ (1a) and (2a) shows two of them.

MultiCKUS was carried out in a city in Southeast Norway. Because of the proximity to Oslo, Standard Eastern Norwegian has a strong influence in the area. Still, the local dialect is also in use and some local dialect features are especially common, e.g. parts of the pronominal system (cf. Stjernholm & Søfteland 2019). In 3rd person plural, the local dialect subject form can be homophonic with the object form in standard Eastern Norwegian (and written Bokmål).⁵ How, then, should we interpret situations like in (1), where (1a) is an utterance by an (early) L2 learner, (1b) is the equivalent in Standard Eastern Norwegian, and (1c) is a local form? “OBJ.” in the glossing marks when *dem* would be analysed as targetlike in the subject position.

- (1) a. (Actual utterance from L2 data)
 dem gikk hjem
 they.”OBJ.”/them.OBJ. walked home
 ‘They/Them walked home.’
 b. (Standard Eastern Norwegian)

⁴See Johannessen (2017) and Søfteland (2018) for thorough discussions on annotation processes for Norwegian/Scandinavian spoken language.

⁵This is shown in (1c), and is indicated in the glossing where the (subject) pronoun is marked “OBJ” (compare with the glossing of 1b).

2 L2 acquisition in a rich dialectal environment

- de gikk hjem
 they.SUBJ. walked home
 ‘They walked home.’
- c. (Local dialect form)
 dem gikk hjem
 they.”OBJ.” walked home
 ‘They walked home.’

The form *dem* ‘them’ used in a subject position, as in (1a), is targetlike when compared to the local dialect (1c). /*dem*/ or /*døm*/ is the regular form in this area, both in subject and object position. In Standard Eastern Norwegian the most frequent subject form is *de* ‘they’ (1b). Considering the language situation in the area where MultiCKUS took place, we can be sure that the children have heard both *de* ‘they’ and *dem* ‘them’ in subject position, but we cannot know how much. Thus, we must consider both (1b) and (1c) targetlike. However, it is not possible to determine if (1a), the actual utterance by an L2 speaker, is dialectal or interlanguage variation. Within sociolinguistics and research on language change, this kind of ambiguity has sometimes been referred to as an *isomorphic crux* (Hårstad 2009): the finishing point (the cross, or “crux”) of a specific linguistic change can be traced back to two different origins, with both (diachronic) processes ending in the same homonymous (“isomorphic”) forms. If a researcher, student or teacher is supposed to judge whether an L2 learner utterance is targetlike or not, how can they make adequate decisions about examples like (1a)? Our concern is that if (1a) was uttered by an L1 speaker it might be judged as targetlike, while if it was uttered by an L2 speaker it might be judged as nontargetlike. This of course affects analyses of data.

The interpretation of the feminine pronoun form *henne* ‘her’ in subject position, as in (2a), is even trickier. In Standard Eastern Norwegian *henne* would be the object form and *hun* the subject form (as shown in 2b).

- (2) a. (Actual utterance from L2 data)
 da skrek henne
 then screamed she.”OBJ.”/her.OBJ.
 ‘Then she/her screamed.’
- b. (Standard Eastern Norwegian)
 da skrek hun
 then screamed she.SUBJ.
 ‘Then she screamed.’

Linda Evenstad Emilsen & Åshild Sjøteland

- c. (Attested among young L1 speakers in adjacent dialect areas)
 da skrek henne
 then screamed she.”OBJ.”
 ‘Then she screamed.’

The original object form *henne* in subject position is not known to be part of the local dialect traditionally, but it might be targetlike if compared to the dialect of the L2 learners’ L1 Norwegian classmates: *henne* is used in subject position by adolescents elsewhere in Eastern Norwegian, in and around Oslo (cf. 2c), but we do not know if this linguistic change has appeared in these learners’ linguistic environments.⁶ The only way we can decide whether *da skrek henne* (2a) is targetlike, is to find out if L2 speakers encounter this in their input (and to what extent). This exemplifies some of the complexities of data interpretation in our project.

These are just two examples from an area close to Oslo, but they motivated us to investigate potential ambiguities between TL and interlanguage variation more systematically, as we find little discussion of this in the research literature. This issue is relevant to teachers as well. Despite the “Dialect paragraph” and its long history, there is reason to believe that many teachers are unconscious of dialectal variation, both in L1 teaching classrooms (Jahr 1992; Jansson et al. 2017) and in L2 classrooms and learning materials (Husby 2009; Heide 2017).

2.2 The role of an idealised or monolingual norm in assessing L2 use and development

As mentioned in the introduction, SLA research and additional-language teaching have been criticised for having used L1 monolingual norms when assessing L2 use and development. Cook & Newson (2007: 221) even go so far in their criticism as to claim that “the only true knowledge of the language is taken to be that of the adult monolingual native speaker”, suggesting that interlanguages have been of less importance in linguistics. Many researchers (see for instance Saniei 2011) connect this norm with Chomsky (1965: 3) saying, “linguistic theory is concerned with an ideal speaker-listener in a completely homogeneous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance.” This quote has often been

⁶The glossing in (2) marks *henne* as “OBJ.” with quotation marks when we analyse it as a subject.

2 L2 acquisition in a rich dialectal environment

taken to mean that researchers should only study idealised L1 speakers. At the time of Chomsky's statement, the field of study was so new that investigating idealisations was complicated enough. As the field developed and insights and methodological developments accumulated, researchers started to go beyond the ideal speaker–listener. SLA research is a good example: for decades now, investigations have described and explained variation in the interlanguage of language learners (see for instance [Corder 1967](#); [Selinker 1972](#)), i.e. the grammars of “non-ideal” language learners have also been studied. However, many SLA studies still use L1 native speakers as a baseline for L2 learners/speakers and take L1 monolingual speakers as “the golden standard” ([Amaral & Roeper 2014](#): 29).

Using L1 monolinguals as the standard for L2 learners has been contested as it raises a number of issues (pointed out by [Cook & Newson 2007](#); [Amaral & Roeper 2014](#); [Slabakova 2016](#); [Group 2016](#); [Ortega 2019](#)). Idealisations can be useful when a research field is so new that there are too many unknown factors and no established methodology; a more streamlined approach is needed to gain the first insights. However, idealisations are always problematic as they are not accurate depictions of reality. Some of the issues when working with idealisations relate to ontological status, others to theoretical approaches, (unconscious) attitudes, and methodological issues, especially how we handle data. In what follows, we will describe the issues most relevant to our study and draw explicit connections to the Norwegian language context.

First, we may ask who are the monolingual L1 speakers? This is an empirical question. In today's globalised society, L1 speakers are often not monolingual. [Cook & Newson \(2007: 6\)](#), for instance, ask if “the issue is really whether it is proper to set universal bilingualism to one side in linguists' descriptions of competence or whether it should in effect form the basis of the description from the beginning”; the norm should rather be an L1 bilingual speaker than an L1 monolingual.

Further, *what is an L1?* It often seems as if an L1, at least on a societal level, only consists of *one grammar* and is unchanging and stable. However, this is an idealisation and a simplification; in reality, all languages will have some amount of dialectal variation. Furthermore, often “L1 grammar” seems to be considered as equivalent to a standardised norm when used as a baseline for L2 acquisition. Given all this, an idealised L1 grammar is insufficient to have as the regular starting point for a scientific approach: many L2 learners worldwide receive input from varieties other than the standard language. Also, using a language in different contexts often includes using different registers. Seen this way, speakers can have *parallel grammars* (cf. [Eide & Sollid 2011](#)), which differ from other individuals' grammars (still speaking the same variety). In addition, a language is in

Linda Evenstad Emilsen & Åshild Søfteland

continual change both in the individual and across individuals, which may cause synchronic variation and generational variation to exist at the same time. Several studies indicate that cross-linguistic influence may affect the L1 (see for instance Cook & Newson 2007), so the typical idealisation needs to be questioned considering this as well. When considering cross-linguistic influence effects on the L1, the typical idealisation needs to be questioned, as it highlights that the L1 is not one fixed grammar, but varies synchronically and is changeable over time both across and within individuals.

Working with an L1 idealisation may also harm research on *L1 acquisition*: Child L1 acquisition is characterised by variation that differ from the grammar of adult L1 speakers. We may call this *L1 interlanguage variation*. If an L1 child meets dialectal variation that is homophonic to L1 interlanguage in the input, this may be wrongly analysed as developmental variation in a child learning the language and not as dialect variation. On a more ideological level, in light of the critique of idealisations, one could ask if not L1 interlanguage variation also should be classified as L1 variation at the same level as dialect variation. Even if the child's L1 interlanguage grammar is different from that of an adult L1 speaker and undergoing rapid development, it is still an L1 grammar just as much as the adult's L1 grammars are. It is the children's L1.

To sum up, the concept of L1 is clearly not straightforward, and, in many respects, perhaps not appropriate as a baseline for L2 acquisition – unless all relevant L1 variation is taken into account.⁷ The existence of L1 variation – be it language change, dialectal variation, social registers or something else – raises the following questions: What grammar do L2 learners actually acquire? And how do we handle variation in the input methodologically? Our main methodological question in the present study concerns decisions on what can be considered targetlike when there is potentially extensive variation in L2 learners' input.

2.3 TL variation in SLA

TL variation in L2 acquisition is not a widely covered topic. However, there is a growing body of research on variation in American English (e.g., Eisenstein 1986), Spanish (e.g., Gudmestad 2012), Cypriote and Standard Greek (Leivada et al. 2017), and Norwegian (Rodina & Westergaard 2015a). Much of the research focuses on acquisition of sociolinguistic variation. Some studies examine L2 learners' exposure to linguistic variation in the input (including but not limited to so-

⁷This then again raises the questions: What is relevant input and when is input not relevant, if it were to be not relevant at any points? We do not address this any further, but the questions highlight the complexity of the topic of input and baseline.

2 L2 acquisition in a rich dialectal environment

ciolinguistic variation), how this is constrained by internal and external factors, and whether learners acquire this variation. Others investigate how L2 learners acquire specific dialectal features. Such work underscores that acquiring a TL means acquiring variation and demonstrates that it may be hard to idealise what an L1 is. What is less studied is how L2 learners navigate in language environments where they may come in contact with extensive – and potentially conflicting – input from a different variety than the one they may primarily be thought to acquire. To our knowledge, few studies contextualise how this kind of TL variation poses methodological challenges, with “isomorphic cruxes” where interlanguage variation is homophonic with TL variation (but see Emilsen (in preparation) and short comments as in Cornips (2018: 17)). In this section, we focus on Norwegian and show that this aspect is ignored in the research literature, in language learner corpora for research, and in textbooks for teachers.

In Section 2.1, we described the Norwegian language context with its linguistic variation and the high status of dialects. Due to this situation, a language learner of Norwegian in Norway may, and is likely to, get input from

- a. the local dialect,
- b. dialects from other parts of the country,
- c. the (unofficial) standard Eastern Norwegian
- d. two different written standards (Bokmål and Nynorsk),
- e. multi-ethnolects, and
- f. L2 Norwegian variation from other L2 learners of Norwegian.

Still, in acquisition research from the Norwegian language community, we find a range of examples of Norwegian being treated as one uniform variety, disregarding the variation in the input. For instance, Glahn et al. (2001) study agreement in nominal phrases and the placement of negation in subordinate clauses in adult L2 acquisition of Norwegian, Swedish and Danish. They use elicitation tasks to find out if the participants follow a specific acquisition trajectory of the tested features. One major issue with this study is that they appear to take for granted that learners of all Scandinavian varieties can be compared; little consideration is given to variation within the languages. In so doing, they also imply that the input the different language users have access to is comparable. We argue that comparing language learners without considering different dialect backgrounds is problematic. For example, placement of negation can vary in both subordinate

Linda Evenstad Emilsen & Åshild Sjøteland

and main clauses in Norwegian dialects (see Bentzen 2007), and apocope can lead to invisible agreement marking (see Section 3.1). This should be highly relevant variation for Glahn et al., but they barely mention it.

Similarly, Ragnhildstveit (2017) claims there is a strong correlation between assigned gender and declension of nouns in Norwegian, but relates this only to written norms. This is problematic, as the learners may also have oral input, and the oral input may – and likely does – diverge from the written systems. There is no discussion of the fact that the L2 learners in her study may have much oral input diverging from the written systems described. She does, however, describe both written standards of Norwegian – Bokmål and Nynorsk – thus acknowledging some variation in Norwegian. The lack of discussion of the variation between oral and written language in Ragnhildstveit’s (2017) study, is especially problematic since several recent studies have attested an ongoing change in some Norwegian dialects (including the dialects in and around Oslo), where the three-gender system is reduced to a two-gender system (see Lødrup 2011; Rodina & Westergaard 2015b; Busterud et al. 2019) and where the definite suffixes are affected differently. As pointed out by Emilsen (In preparation), in several dialects there is now a clear discrepancy between the definite singular suffix and gender agreement, and different systems co-exist, resulting in a weakening of the link between gender and suffix. This means that the gender agreement/definite suffix system is less transparent, making it less evident what system language learners acquire. For instance, if L2 learners of Norwegian produce a two-gender system, is that due to a two-gender system in the input or is it interlanguage variation?

We also find a lack of acknowledgement and discussion of variation in Norwegian in the extensive ASK corpus developed at the University of Bergen (*Norsk AndreSpråksKorpus*, ‘Norwegian second language corpus’). ASK consists of data collected from written exams by adult L2 learners, testing their competence in Norwegian. These exams are annotated with the learners’ L1 and general linguistic background, other personal data, level on the test-exam, and *feilanalyser*, ‘error analyses’. The corpus is searchable for a variety of linguistic features, both in the students’ original submissions and in the “correct-marked” corpus. On the corpus website, the guidelines for error-annotation include some important methodological considerations, but looking through the listed error examples, it still seems as if potential (written) language variation is ignored, for instance in the placement of negation. Searches in the written language corpus *Leksikografisk bokmålskorpus* (Knudsen & Fjeld 2013) show that several of the “errors” identified for L2 texts are common in L1 text production.

If we turn to the pedagogical literature for teachers, we find that, in a number of cases, Norwegian is treated as a uniform variety. Heide (2017) points out

2 L2 acquisition in a rich dialectal environment

that L2 textbooks mostly describe typical pronunciation of the written standard Bokmål and only mention dialectal variation briefly, while the research literature most often excludes it completely. Husby (2009) describes a situation where the usual language of instruction for adult L2 teachers of Norwegian is a “Bokmål-influenced spoken language with some dialectal variation” (Husby 2009, our translation). Husby says this is problematic because such a variety rarely exists outside the classroom. He further explains that L2 children and L2 adults can have different primary sources of input: Children often have more access to local (and other) dialects through their peers in kindergartens and schools, while adults often primarily encounter Norwegian through Norwegian courses for adults. Here different dialects are much less present, superseded by the “Bokmål-influenced” speech. It is also not unusual for minority language families to speak the majority language at home (e.g., Berggreen & Latomaa 1994; Kulbrandstad 1997; Mancilla-Martinez & Kieffer 2010; Karlsen & Lykkenborg 2012; Fulland 2016). This entails that family members will be part of the (L2) input of other family members – at the same time as each family member may have diverging input from L1 Norwegian sources. This again makes it hard to pinpoint what grammatical system we could expect the learners to acquire, i.e. what an adequate baseline would be.

The teacher textbook *God nok i norsk?* (‘Good enough in Norwegian?’) (Berggreen et al. 2012) is one example that shows how common it is to treat Norwegian as one variety when analysing the L2 learner’s acquisition process. This book has L2 writing as the main subject and relies mostly on the authors’ research on L2 students’ texts, but the book still has many generalizing quotes about Norwegian grammar, such as this one:

In Norwegian, subordinate words in the noun phrases, such as determinatives and adjectives, must adjust to the noun. [...] The adjective shall indicate whether the noun it belongs to is singular or plural, definite or indefinite, neutral or common-gender. (Berggreen et al. 2012: 80, our translation; see Section 3.1 for details on the grammatical features)

Statements like this appear to discuss the Norwegian language in general, not only written standards, thus failing to acknowledge the variation in the input that L2 learners may have been exposed to. In combination with the findings from Heide (2017) and Husby (2009) this strengthens our assumption that both researchers and teachers may be unconscious about dialectal variation in the L2 learners’ input.

3 Interlanguage variation or targetlike (dialect) variation

So far, we have claimed that dialectal variation in the TL may lead to challenges in the interpretation of L2 data. More specifically, we claimed that interlanguage variation may coincide with TL variation, making it difficult to distinguish between the two. In this section, we present data and analyses supporting this statement, by comparing empirical descriptions from a range of L2 studies with authentic spontaneous speech data from the L1 corpus *The Nordic Dialect Corpus* (NDC) (Johannessen et al. 2009).

First, we review some relevant features of the Norwegian nominal phrase, i.e. how it is described in the literature and in descriptive reference grammars. These descriptions are heavily influenced by the written standards. Then, we present an overview of SLA literature on L2 acquisition of the relevant linguistic features, focusing on nontargetlike variation. This is followed by a description of certain dialect phenomena that may cause the realisation of the relevant features in Norwegian dialects to be homophonic with the described interlanguage variation (i.e. “isomorphic cruxes”, as mentioned in Section 2.1). The description of the phenomena is accompanied by authentic spontaneous speech data from different Norwegian dialects. After this investigation of the nominal phrase in Section 3.1, we do the same for finiteness and V2 constructions in Section 3.2.

3.1 Morphology in the noun phrase

3.1.1 Typical description of the Norwegian nominal phrase

Norwegian noun phrases are often described as being inflected for definiteness and number, as in *The Norwegian Reference Grammar* (Faarlund et al. 1997). Some also say that nouns are inflected for gender since the definite singular suffix often correlates with the gender of the noun (e.g., Johannessen & Larsson 2015). According to *The Norwegian Reference Grammar*, adjectives and determiners agree with the noun in gender, number and definiteness. When an attributive adjective is present, a prenominal determiner is often obligatory, and definite contexts with attributive adjectives give rise to the construction often labelled as double definiteness or compositional definiteness (Julien 2005; Baal 2018); there is both a definite suffix on the noun and a definite determiner present. Table 1 shows a typical paradigm for Norwegian noun phrases with attributive adjectives.

2 *L2 acquisition in a rich dialectal environment*

Table 1: A typical paradigm for Norwegian nouns modified with an attributive adjective

	Indefinite singular			Definite singular		
M.SG.	en	stor bil		den	stor-e	bil-en
	M.INDEF.SG	big car		M.DEF.SG	big-DEF.SG	car-M.DEF.SG
		‘a big car’				‘the big car’
F.SG.	ei	stor dukke		den	stor-e	dukk-a
	F.INDEF.SG	big doll		F.DEF.SG	big=DEF.SG	doll-F.DEF.SG
		‘a big doll’				‘the big doll’
N.SG.	et	stor-t	hus	det	stor-e	hus-et
	N.INDEF.SG	big-N.INDEF.SG	house	N.DEF.SG	big-DEF.SG	house-N.DEF.SG
		‘a big house’				‘the big house’
M.PL.	stor-e	bil-er		de	stor-e	bil-ene
	big-INDEF.PL	car-INDEF.PL		DEF.PL	big-DEF.PL	car-DEF.PL
		‘big cars’				‘the big cars’
F.PL.	stor-e	dukk-er		de	stor-e	dukk-ene
	big-INDEF.PL	doll-INDEF.PL		DEF.PL	big-DEF.PL	doll-DEF.PL
		‘big dolls’				‘the big dolls’
N.PL.	stor-e	hus		de	stor-e	hus-ene
	big-INDEF.PL	house-Ø		DEF.PL	big-DEF.PL	house-DEF.PL
		‘big houses’				‘the big houses’

3.1.2 Nominal morphology in SLA

Acquisition and use of nominal morphology in L2 has been extensively investigated across languages: Bruhn de Garavito & White (2002) for L2 Spanish, Hawkins & Franceschina (2004) for L2 Spanish and L2 French, Trenkic (2007) for L2 English, Glahn et al. (2001) for L2 “Mainland Scandinavian”, and Jin (2007); Jin et al. (2009); Anderssen & Bentzen (2013); Rodina & Westergaard (2013; 2015a); Emilsen & Søfteland (2018); and Emilsen 2019; In preparation) for L2 Norwegian. All of these studies report nontargetlike variation at some point in the acquisition of L2 nominal morphology. A frequently observed pattern is omission of agreement or prenominal determiners in contexts where they are expected in the TL. Another frequent pattern is substitution of phonological forms: the overt marking is realised by a morphological form other than the one predicted in the TL. A third pattern, albeit rare, is the use of a morphological marking in contexts

Linda Evenstad Emilsen & Åshild Søfteland

where it is not expected in the TL.

This brief overview shows that (nontargetlike) variation in the realisation of the nominal phrase is attested and predicted in L2 acquisition. However, Norwegian dialects vary greatly in the way nominal morphology is realised, and some of this variation is homophonic with variation predicted for interlanguage grammars, as we show in Section 3.1.3.

3.1.3 Nominal morphology in Norwegian dialects

Nominal morphology is subject to variation due to apocope in many Norwegian dialects. Apocope may be defined as the loss of unstressed word final vowels (e.g., Mæhlum & Røyneland 2012: 76, 106). The examples show apocopation in authentic spontaneous speech data from South-Western Norwegian (*Fusa*) in (3) and North-Western Norwegian (*Aure*) in (4) (data from NDC, phonetic transcription). As a consequence of apocope, the unstressed *-e* on adjectives is missing, and the nominal phrases look, on the surface, as if they lacked agreement for plural (3) or definiteness (4).⁸

- (3) veldi go vænna ⇒ *go-e*
 very good friend.PL good.PL.
 ‘very good friends’ ‘good’
- (4) n ræu rinngen ⇒ *ræu-e*
 DEF.SG.COM red ring.DEF.SG.M red.DEF.
 ‘this red ring’ ‘red’

Apocope is an established dialect feature in Norwegian, the core geographical area for it being the northern and middle part of Norway (cf. Mæhlum & Røyneland 2012: 76). Apocope is also frequent in fast spontaneous speech across all spoken varieties in typically unstressed words or contexts. Apocope is, in other words, rather widespread. That increases the likelihood of L2 learners (and L1 learners) receiving (extensive) input where the nominal phrase may be considered targetlike even though overt agreement marking is not present. As previously mentioned, this kind of morpho-phonological realisation of the nominal phrase is also found in L2 interlanguage variation, creating a potential ambiguity – an isomorphic crux – between interlanguage variation and dialectal, targetlike variation when coding and interpreting L2 data.

⁸This is marked in the glossing with an arrow pointing to how it would look if it were not apocopised, i.e. if the agreement was spelled out phonologically.

2 L2 acquisition in a rich dialectal environment

A second challenge is caused by what on the surface may look like substitution in the prenominal determiner: the prenominal determiner /de/ (*det*, ‘the’) associated with singular definite neuter, is substituted for the masculine/feminine form /den/ (*den* ‘the’) and the plural form /di/ (*de* ‘the’), as seen in (5) (Mid-Norwegian dialect, Eide et al. 2017: 46) and (6) (Northern Norwegian dialect, Sollid 2014). This leads to phrases that look like they are violating the agreement criteria often found in typical descriptions of Norwegian. If an L2 speaker produced phrases like (5–6) it is likely that the definite article *de* would be analysed as a neuter singular form and not as a masculine (5) or plural (6) form.⁹

- (5) **de** grønn-e pærme-n ⇒ *den*
 DEF.”SG.N.” green.DEF.SG. portfolio.DEF.SG.M DEF.SG.M.
 ‘the green portfolio’
- (6) **de** her flertallord-an ⇒ *di*
 DEF.”SG.N.” here pluralword.PL.DEF DEF.PL
 ‘these plural words’

This kind of agreement may however be targetlike, as it is attested in certain Norwegian dialects, at least *Fosen* (Middle Norwegian, (5)) and *Reisadalen* (Northern Norwegian, (6)), as the examples show.

A third challenge related to nominal morphology is the loss of final /r/ in certain frequent word types, such as indefinite plural nouns and present tense verbs. This is often labelled *r-bortfall* ‘r-loss’, and is common in many dialects (cf. Mæhlum & Røynealand 2012: 53). R-loss may cause nouns to look as if they lack plural declension, as in (7) (*Herøy*, North-Western Norwegian) and (8) (*Evje*, South-Norwegian) (data from NDC, phonetic transcription).¹⁰

- (7) e lika gått **jænnte** i bikini ⇒ jænnte-r
 I like well girl in bikini girls.PL.
 ‘I like girls in bikini very much’
- (8) dæi va som **pærrle** ⇒ pærrle-r
 they were like pearl pearl.PL.
 ‘They were like pearls’

⁹We mark this by glossing the definite article in these phrases as “NEUT” and “SG” with quotation marks even if they are masculine or plural forms in the dialect data. The arrows point to what the form would look like if it was spelled out with unambiguous masculine or plural agreement.

¹⁰The arrow points to what the forms would look like if there were no r-loss, i.e. if the plural marking was spelled out phonologically. The written standard forms would be *jente/perle* (INDEF.SG) and *jenter/perler* (INDEF.PL).

Linda Evenstad Emilsen & Åshild Søfteland

Omission of declension is a common feature of L2 interlanguage at some point during acquisition (e.g. White 2003; Trenkic 2007; 2009; Goad & White 2009; Emilsen & Søfteland 2018; Emilsen 2019), making it potentially hard to differentiate between the two: If an L2 learner produces an utterance such as (7) or (8), is it targetlike or is it interlanguage variation?

3.2 Finiteness and V2

3.2.1 Typical description of finiteness and V2 in Norwegian

Norwegian is often described as a V2-language: every main clause needs a subject and a finite verb, where the finite verb is in second position in declarative sentences (see *The Norwegian Reference Grammar*). In sentences with topicalisation (of phrases other than the subject), the verb and subject *invert*, i.e. the verb moves in front of the subject, as in (9b):

- (9) a. S V ADV
 Linda danser hver kveld
 Linda dances every night
 ‘Linda dances every night.’
 b. ADV V S
 Hver kveld danser Linda
 Every night dances Linda
 ‘Every night Linda dances.’

A paradigm of Norwegian verb tenses is given in (10), based on descriptions from descriptive grammars. Especially important in our case is that present tense is regarded as finite and that *-er* is a frequent present tense suffix.

- (10) Infinitive Present Preterit Perfect
 kjør-e kjør-**er** kjør-te kjør-t
 drive-INF drive-PRS drive-PRET drive-PERF
 dans-e dans-**er** dans-a/-et dans-a/-et
 dance-INF dance-PRES dance-PRET dance-PERF

3.2.2 V2 and finiteness in SLA

It is well-attested that both finiteness and V2 may pose challenges for L2 learners: e.g., Prévost & White (2000) for L2 French and German, and Hagen (2001; 2005); Mosfjeld (2017) and Gujord et al. (2018) for L2 Norwegian. L2 acquisition is often

2 L2 acquisition in a rich dialectal environment

characterised by a period of nontargetlike finiteness marking either because of substitution of morphological marking or because of omission of marking and overuse of the infinitival form. Adult L2 learners of V2 languages are found to lack inversion of the verb and subject in contexts where this may be expected, giving raise to V3 word order, e.g., Bohnacker (2010) on L2 Swedish, Bohnacker (2006) on L2 German, and Mosfeld (2017) on L2 Norwegian.

However, as shown for the nominal phrases, V2 and finiteness are not uniform features across all spoken varieties of Norwegian, which poses a challenge for interpreting language learner data.

3.2.3 Finiteness in Norwegian spoken varieties

As noted above, present tense is regarded as finite, and *-er* is a frequent present tense suffix. However, the aforementioned loss of /r/ in final position in Norwegian also impacts verb morphology, often making infinitive and present tense homophonic in productive inflectional classes. Two of many examples from NDC are shown in (11) (*Volda*, North-Western Norwegian) and (12) (*Ballangen*, Northern-Norwegian). In dialects where r-loss is attested, non-overt finiteness marking like this is targetlike.¹¹

- (11) møtje fállk så **tjøpe** físskekort dær ⇒ *tjøpe-r*
 much people who buy fish-cards there buy-PRS
 ‘(There are) much people who buy fishing licences there.’

- (12) æ **digge** tran ⇒ *digge-r*
 I love cod_liver_oil love-PRS
 ‘I love cod liver oil.’

Apocope also affects verbs, and in many dialects both the infinitive suffix and the present tense suffix are apocopated, making these forms homonymous, as in (13) and (14) (*Mo i Rana*, Northern-Norwegian, examples from NDC).

- (13) næi de **spis** e ikke ⇒ *spis-er*
 no that eat I not eatPRS
 ‘No, I don’t eat that’.

¹¹The arrow points to what the forms would look like if there were no r-loss, i.e. if the present tense marking was spelled out. The written Bokmål forms would be *kjøpe/digge* (INF.) and *kjøper/digger* (PRES.).

Linda Evenstad Emilsen & Åshild Sjøfteland

- (14) de bruk e ikke å spis \Rightarrow spis-e
 that use I not to eat eat-INF
 ‘I usually don’t eat that.’

There are quite a few dialects with no overt distinction between the infinitival form and the finite present tense: either both end in an unstressed *-e* (or *-a*) or they only consist of the stem of the verb due to apocope. Since these features are common, it is likely that L2 learners of Norwegian encounter them in the input.

3.2.4 V3 in Norwegian spoken varieties

V2 is often presented as a consistent rule in general descriptions of the Norwegian grammar, but variation related to the V2-rule is widely discussed in recent literature on dialect syntax. Westergaard (2008) shows that word order varies in wh-questions, depending on the length of the wh-element and different information structural aspects. A national data collection of grammaticality judgments of syntactic spoken language variables, *Nordic Syntactic Judgment Database* (Lindstad et al. 2009), also documents non-V2 in wh-questions in large parts of the country (e.g., Vangsnes & Westergaard 2014). Furthermore, lack of inversion is a grammatical feature of the Oslo multi-ethnolect (cf. Svendsen & Røyneland 2008; Opsahl & Nistov 2010), making many declarative sentences V3 (examples 15 and 16 from Opsahl & Nistov 2010):

- (15) egentlig alle kan bidra
 actually everyone can contribute
 ‘Everyone can actually contribute.’
- (16) hvis noen står og breaker alle stopper opp
 if some stands and breaks all stop up
 ‘I someone is breakdancing, everyone stops.’

Opsahl & Nistov (2010) show that lack of inversion is a signature of multi-ethnolectal style among adolescents in Oslo, but the use varies both inter- and intra-individually and there are sociolinguistic limitations on the variation between XVS (inversion) and XSV (non-inversion). They also point out that lack of inversion is more frequent after certain adverbials, such as *uansett* and *egentlig*. Later research on the same and similar data also finds pragmatic limitations for the use (see Freywald et al. (2015) for a comparative study of this in Norwegian and other North-Germanic languages).

2 *L2 acquisition in a rich dialectal environment*

Svendsen & Røyneland (2008) also discuss lack of inversion in the same language group. They add an important methodological detail: utterances with an Adverbial (X) right before Subject+Verb (SV) do not necessarily entail lack of inversion (XSV). If the initial adverbial has “a break after” (Svendsen & Røyneland 2008: 75) in the pronunciation of the utterance, it should be interpreted as extraposed, not topicalised. The adverbial must then be considered external to the main clause, the Subject is still in first position and there is no lack of inversion. Without access to the sound recording, an example like (15) is ambiguous: V3 with *egentlig* ‘actually’ analysed as a topicalized adverbial (and a regular main clause pronunciation pattern), or V2 with *egentlig* analysed as an extraposed adverbial.

We studied relevant utterances in the *NoTa-Oslo Corpus* (an Oslo dialect corpus, with audio and video), and found that there are gradual transitions between these two analyses. Listening to the prosody of each utterance – accent, stress, pauses – makes it possible to tease apart the interpretations for many examples. In some, like (17) and (18), it is still impossible to decide, meaning that for some Oslo adolescents, both analyses are possible: an interpretation of *uansett* ‘anyway’ and *faktisk* ‘actually’ as extraposed adverbials, followed by V2 syntax (18a), or an interpretation of the same adverbials as topicalised, with lack of inversion (18b), i.e. V3:

- (17) **uansett** alt er bedre enn Norge altså V3?
 anyway all is better than Norway then
 ‘Anyway everything is better than Norway.’
- (18) **faktisk** jeg har aldri sett en hel episode av Glamour V3?
 actually I have never seen a whole episode of Glamour
 ‘Actually I have never seen a whole episode of Glamour.’
- a. (extrap.) S Vaux A V DO >V2
 faktisk (.) Jeg har aldri sett en hel episode
 [actually] (.) I have never seen a whole episode
- b. A S Vaux A V DO >V3
 faktisk jeg har aldri sett en hel episode
 actually I have never seen a whole episode

In the corpus (*NoTa-Oslo*), it looks like V2/V3 ambiguity can appear independently of the adolescents’ reported linguistic background (reported L1 or L2 parents), geographical background (East or West) and social background (parents’ education level). We do not know what kind of linguistic variation these speakers

Linda Evenstad Emilsen & Åshild Sjøteland

encountered in the input when they learned Norwegian, as L1 or L2, but multi-ethnolects appear to be widespread in urban areas such as Oslo, and minority-language speaking families also use L2 Norwegian at home (cf. Section 2.3). Thus, the likelihood of encountering lack of inversion in the input is high, at least in urban areas.

In sum, the issue of V2 is multifaceted in Norwegian, with a potentially large amount of variation in the input of language learners. The placement of the verb depends on a number of conditions. In the acquisition process, language learners must navigate between marginal differences in information structure and word length/complexity to acquire targetlike verb placement. L1 children appear sensitive to these patterns from an early age (Westergaard 2008: 1854). Given the variation discussed in this section, we can conclude that V3 among adolescents in Oslo can have multiple sources. The variation and ambiguity in interpretation of utterances demonstrate the complexity of working with these syntactic phenomena in language learner data.

4 Conclusion

In this chapter we described the highly varied language situation in Norway, where any language learner is likely to be exposed to different dialects and also different written norms of the same language. This means that the language learner has to navigate between potentially diverging linguistic systems in the input, which has substantial implications for how we interpret L2 data. Since the learner may be extracting grammatical information from different systems, it becomes less transparent what an adequate baseline is. This supports the criticism of the L1 monolingual idealisation that has prevailed in SLA.

Even though Norwegian is far from being one variety with one grammar, either within or between individuals, it often is treated as a single variety – in research literature on SLA, in an L2 corpus and in textbooks for (future) L2 teachers. This is a highly problematic approach to interpreting L2 data. We discussed several dialect phenomena – apocope, agreement variation, r-loss and lack of inversion/non-V2 – that can give rise to ambiguity when compared to descriptions of interlanguage variation in the SLA literature. We referred to Hårstad (2009) and his use of the term *isomorphic crux* to describe when, in sociolinguistic analyses of language change, it is impossible to determine where a linguistic form stems from. This is exactly what we see from our methodological point of view. If an L2 learner had uttered the examples in (3–8) and (11–16), we would not be able to determine if the morphological forms or syntactic features in use is dialectal or interlanguage variation.

2 L2 acquisition in a rich dialectal environment

Some SLA research considers that language learners may be acquiring a local dialect and/or investigates the acquisition of a specific local dialect; nevertheless, the potential influence from diverging dialectal systems is rarely thematised and discussed. Our study has shown that a range of constructions considered typical for L2 acquisition are homophonic with targetlike variation if the language learner is receiving input on it. Descriptions of *only* the local dialect, or *one* of the written standards, for example, would not be sufficient to determine whether a produced construction is targetlike or not, since the construction may have occurred in the learner's input from other dialects, multi-ethnolects and/or written varieties. Our study also shows that it is imperative to strive for an updated description of the variety/varieties in question; relying solely on older descriptions of dialects and/or abstractions from the written systems is insufficient.

We have claimed that working with data from language situations characterised by extensive variation pose methodological challenges for the interpretation of data. The challenges we describe may be impossible to solve fully, but it is important that we acknowledge and take into consideration that there might be targetlike variation homophonic to interlanguage variation, and this then raises a need to know more about the input of the learners.

Some of these challenges are probably present to a certain degree for most SLA researchers. Even so, our study highlights the relevance of and need for detailed information about exposure to different varieties, both qualitative and quantitative. Some important considerations are

- a. how much input is needed to acquire a feature, i.e. when a feature can be expected to be acquired, and, hence, when different sources of input are necessary to include,
- b. when faced with diverging input, what determines which specific features the language learner acquires, and
- c. how factors such as saliency, frequency and transparency affect the process (see for instance [Sun 2008](#)).

These considerations may shed more light upon the nuances in methodological challenges as those we describe in this chapter.

Another step on the way, focusing on the methodological considerations alone, is by sharing our data. We acknowledge, of course, that there may be ethical considerations concerning the public sharing of data, especially when children are involved, but open access/open data should be the general goal. This will not resolve the challenges we have described, but it will allow others to make

Linda Evenstad Emilsen & Åshild Søfteland

their own judgements about the data and help bring transparency to the analytic choices we have made. As is clear from our chapter, we cannot offer any single, fixed solution to the challenges we have posed, but awareness is a first step.

References

- Amaral, Luiz & Tom Roeper. 2014. Multiple grammars and second language representation. *Second Language Research* 30(1). 3–36. DOI: [10.1177/02679017](https://doi.org/10.1177/02679017).
- Anderssen, Merete & Kristine Bentzen. 2013. Cross-linguistic influence outside the syntax-pragmatics interface: a case study of the acquisition of definiteness. *Studia Linguistica* 67(1). 82–100. DOI: [10.1111/stul.12011](https://doi.org/10.1111/stul.12011).
- Arntzen, Ragnar. 2012. Tospråklige barn fra barnehage til skole: Et prosjekt i språkstimulering og språkkartlegging av andrespråket 2008–2011. In Ragnar Arntzen, Susanne Duek & Arnstein Hjelde (eds.), *Flerspråklig oppvekst i barnehage, førskole og skole: Artikler av lærere og forskere i det nordiske nettverket TOBANO*, 27–74. Halden: Østfold University College.
- Baal, Yvonne van. 2018. Compositional definiteness in heritage Norwegian: Production studied in a translation experiment. In Jan H. Petersen & Karoline Kühl (eds.), *Selected proceedings of the 8th workshop on immigrant languages in the Americas (WILA 8)*, 9–17. Somerville, MA: Cascadilla Proceedings Project.
- Bentzen, Kristine. 2007. *Order and structure in embedded clauses in Northern Norwegian*. Tromsø: University of Tromsø. (Doctoral dissertation).
- Berggreen, Harald, Kjartan Sørland & Vigdis R. Alver. 2012. *God nok i norsk? språk- og skriveutvikling hos elever med norsk som andrespråk*. Oslo: Cappelen Damm Akademisk.
- Bley-Vroman, Robert. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33. 1–17.
- Bohnacker, Ute. 2006. When Swedes begin to learn German: From V2 to V2. *Second Language Research* 22(4). 443–486.
- Bohnacker, Ute. 2010. The clause-initial position in L2 Swedish declaratives: Word order variation and discourse pragmatics. *Nordic Journal of Linguistics* 33(2). 105–143.
- Berggreen, Harald & Sirkku Latomaa. 1994. Språkbytte og språkbevaring blant vietnamesere i Bergen og Helsinki. In Sally Boyd, Anne Holmen & J. Normann. Jørgensen (eds.), *Sprogbrug og sprogvalg blandt indvandrere i Norden: Bind 1* (Københavnstudier i tosprogethed). København: Danmarks Lærerbhøjskole.

2 L2 acquisition in a rich dialectal environment

- Bruhn de Garavito, Joyce & Lydia White. 2002. The second language acquisition of Spanish DPs: The status of grammatical features. In Ana Teresa Pérez-Leroux & Juana Muñoz Licerias (eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection*, 143–160. New York: Bilingual Press. DOI: [10.1007/978-94-010-0291-2_6](https://doi.org/10.1007/978-94-010-0291-2_6).
- Busterud, Guro, Terje Lohndal, Yulia Rodina & Marit Westergaard. 2019. The loss of feminine gender in Norwegian: A dialect comparison. *The Journal of Comparative Germanic Linguistics* 22. 141–167. DOI: [10.1007/s10828-019-09108-7](https://doi.org/10.1007/s10828-019-09108-7).
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Clyne, Michael. 2000. Lingua franca and ethnolects in Europe and beyond. *Sociolinguistica* 14(1). 83–89. DOI: [10.1515/9783110245196.83](https://doi.org/10.1515/9783110245196.83).
- Cook, Vivian J. & Mark Newson. 2007. *Chomsky's universal grammar: An introduction*. Malden, MA: Blackwell Publishing.
- Corder, Stephen P. 1967. The significance of learner errors. *International Review of Applied Linguistics* 5. 161–170. DOI: [10.1515/iral.1967.5.1-4.161](https://doi.org/10.1515/iral.1967.5.1-4.161).
- Cornips, Leonie. 2018. Bilingual child acquisition through the lens of sociolinguistic approaches. In David Miller, Fatih Bayram, Jason Rothman & Ludovica Serratrice (eds.), *Bilingual cognition and language: The state of the science across its subfields*, 15–36. Amsterdam: John Benjamins.
- Eide, Kristin M., Arnold Dalen, Tor Erik Jenstad & Ola Stemshaug. 2017. *Fosenmåla*. Trondheim: Museumsforlaget.
- Eide, Kristin M. & Hilde Sollid. 2011. Norwegian main clause declaratives. *Variation within and across grammars: Linguistic universals and language variation* 231. 327–360.
- Eisenstein, Miriam R. 1986. Target language variation and second-language acquisition: Learning English in New York City. *World Englishes* 5. 31–46. DOI: [10.1111/j.1467-971X.1986.tb00638.x](https://doi.org/10.1111/j.1467-971X.1986.tb00638.x).
- Emilsen, Linda E. 2019. Acquisition of nominal morphology in Norwegian L2: Trends and tendencies. In Ragnar Arntzen, Gisela Håkansson, Arnstein Hjelde & Jörg-U. Keßler (eds.), *Teachability and learnability across languages*, 163–181. Amsterdam: John Benjamins.
- Emilsen, Linda E. In preparation. *Acquiring a moving target*. Manuscript.
- Emilsen, Linda E. & Åshild Søfteland. 2018. Andrespråksanalyse av barns spontantale. *NOA Norsk som andrespråk* 1-2/2018. 128–157.
- Faarlund, Jan T., Svein Lie & Kjell I. Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Freywald, Ulrikke, Leonie Cornips, Natalia Ganuza, Ingvild Nistov & Toril Opsahl. 2015. Beyond verb second: A matter of novel information-structure effects? Evidence from Norwegian, Swedish, German and Dutch. In Jacomine

Linda Evenstad Emilsen & Åshild Sjøteland

- Nortier & Bente A. Svendsen (eds.), *Language, youth and identity in the 21st century: Linguistic practices across urban spaces*, 73–92. Cambridge: Cambridge University Press.
- Fulland, Helene. 2016. *Language minority children's perspectives on being bilingual: On "bilinguals" and their sensitivity towards complexity*. Oslo. (Doctoral dissertation).
- Gass, Susan M. & Carolyn G. Madden. 1985. Introduction. In Susan M. Gass & Carolyn G. Madden (eds.), *Input in second language acquisition*, 3–12. Boston, MA: Heinle & Heinle.
- Geeslin, Kimberly L. 2011. Variation in L2 Spanish: The state of the discipline. *Studies in Hispanic and Lusophone Linguistics* 4. 461–517. DOI: [10.1515/shll-2011-1110](https://doi.org/10.1515/shll-2011-1110).
- Geeslin, Kimberly L., Lorenzo García-Amaya, Maria Hasler-Baker, Nicholas Henriksen & Jason Killam. 2012. The L2 acquisition of variable perfective past time reference in Spanish in an overseas immersion setting. In Kimberly L. Geeslin & Manuel Díaz-Campos (eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium*, 197–213. Somerville, MA: Cascadia Press.
- Geeslin, Kimberly L. & Aarnes Gudmestad. 2008. The acquisition of variation in second-language Spanish: An agenda for integrating studies of the L2 sound system. *Journal of Applied Linguistics* 5(2). 137–157.
- Glahn, Esther, Gisela Håkansson, Björn Hammarberg, Anne Holmen, Anne Hvenekilde & Karen Lund. 2001. Processability in Scandinavian second language acquisition. *Studies in Second Language Acquisition* 23(3). 389–416.
- Goad, Heather & Lydia White. 2009. Articles in Turkish/English interlanguage revisited: Implications of vowel harmony. In María del P. García Mayo & Roger Hawkins (eds.), *Second language acquisition of articles: Empirical findings and theoretical implications*, 201–232. Amsterdam: John Benjamins.
- Group, The Douglas Fir. 2016. A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal* 100(Suppl.) 19–47. DOI: [10.1111/modl.12301](https://doi.org/10.1111/modl.12301).
- Gudmestad, Aarnes. 2012. Acquiring a variable structure: An interlanguage analysis of second language mood use in Spanish. *Language Learning* 62. 373–402. DOI: [10.1111/j.1467-9.00696.x](https://doi.org/10.1111/j.1467-9.00696.x).
- Gujord, Ann-Kristin H., Randi Neteland & Magnhild Selås. 2018. Framvekst av morfologi, syntaks og ordforråd i tidleg andrespråkutvikling: Ein longitudinell kasusstudie. *NOA Norsk som andrespråk* 1-2/2018. 93–127.
- Hagen, Jon E. 2001. Finittkategoriens kritiske karakter i norsk som andrespråk. In Anne Golden & Helene Uri (eds.), *Andrespråk, tospråklighet, norsk: festskrift til Anne Hvenekilde*, 45–57. Oslo: Unipub.

2 L2 acquisition in a rich dialectal environment

- Hagen, Jon E. 2005. Refleksjoner gjennom andrespråksprismet. In Marte Nordan-ger, Silje Ragnhildstveit & Aasne Vikøy (eds.), *Fra grammatikk til språkpolitikk: Utdrag fra Jon Erik Hagens forfatterskap*, 156–174. Oslo: Novus.
- Hårstad, Stian. 2009. Kommer ikke alt godt fra oven? Et forsøk på å se utviklingstrekk i trønderske talemål i et standardiseringsperspektiv. *Norsk Lingvistisk Tidsskrift* 27(1). 133–143.
- Hawkins, Roger & Florencia Franceschina. 2004. Explaining the acquisition and non-acquisition of determiner–noun gender concord in French and Spanish. In Philippe Prévost & Johanne Paradis (eds.), *The acquisition of French in different contexts: Focus on functional categories*, 175–205. Amsterdam: John Benjamins.
- Heide, Eldar. 2017. Det norske språkmangfaldet og opplæringa i norsk som andrespråk. *Norsklæreren* 3. 16–27.
- Husby, Olaf. 2009. Talt bokmål: Målspåk for andrespråksstudenter? 25(2). 9–40.
- Jahr, Ernst H. 1990. *Den store dialektboka*. Oslo: Novus.
- Jahr, Ernst H. 1992. *Innhogg i nyare norsk språkhistorie*. Oslo: Novus.
- Jansson, Benthe K., Christian B. Bjerke, Åshild Søfteland & Karine Stjernholm. 2017. *Dialekt og skole i østfold: Kva seier lærarar på barnesteget om tilnærming til lokal dialekt i klasserommet*. Presentation at SON E-konferansen2017 – “Språk og skole”, NTNU Trondheim, April 2017.
- Jin, Fufen. 2007. *Second language acquisition and processing of Norwegian DP internal agreement*. Trondheim: The Norwegian University of Science & Technology. (Doctoral dissertation).
- Jin, Fufen, Tor A. Åfarli & Wim A. van Dommelen. 2009. Variability in the L2 acquisition of Norwegian DPs: An evaluation of some current SLA models. In María del P. García Mayo & Roger Hawkins (eds.), *Second language acquisition of articles: Empirical findings and theoretical implications*, 175–200. Amsterdam: John Benjamins.
- Johannessen, Janne B. 2017. Annotations in the Nordic Dialect Corpus. In Nancy Ide & James Pustejovsky (eds.), *Handbook of linguistic annotation*, 1303–1321. Dordrecht: Springer.
- Johannessen, Janne B. & Ida Larsson. 2015. Complexity matters: On gender agreement in heritage Scandinavian. *Frontiers in Psychology* 6. 1842. DOI: [10.3389/fpsyg.2015.01842](https://doi.org/10.3389/fpsyg.2015.01842).
- Johannessen, Janne B., Joel J. Priestley, Kristin Hagen, Tor A. Åfarli & Øystein A Vangsnes. 2009. The Nordic dialect corpus: An advanced research tool. In Kristiina Jokinen & Eckhard Bick (eds.), *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA 2009)* (NEALT Proceedings Series Volume 4).

Linda Evenstad Emilsen & Åshild Søfteland

- Julien, Marit. 2005. *Nominal phrases from a Scandinavian perspective* (Linguistik aktuell / Linguistics Today 87). Amsterdam: John Benjamins.
- Karlsen, Jannicke & Marta Lykkenborg. 2012. Språkbruksmønstre i norsk-pakistanske familier. *NORDAND Nordisk tidsskrift for andrespråksforskning* 7(1). 53–83.
- Klein, Wolfgang. 1998. The contribution of second language acquisition research. *Language Learning* 48. 527–550. DOI: [10.1111/0023-8333.00057](https://doi.org/10.1111/0023-8333.00057).
- Knudsen, Rune L. & Ruth V. Fjeld. 2013. LBK2013: A balanced, annotated national corpus for Norwegian Bokmål. In Lars Borin, Ruth Vatvedt Fjeld, Markus Forsberg, Sanni Nimb, Pierre Nugues & Bolette Sandford Pedersen (eds.), *Proceedings of the Workshop on Lexical Semantic Resources for NLP (NODALIDA)* (NEALT Proceedings Series Volume 19). Oslo.
- Kulbrandstad, Lars A. 1997. *Språkportretter. studier av tolv minoritetselevs språkbruksmønstre, språkholdninger og språkferdigheter*. Vallset: Oplandske Bokforlag.
- Leivada, Evelina, Elena Papadopoulou, Maria Kambanaros & Kleanthes K. Grohmann. 2017. The influence of bilingualism and non-standardization on the perception of native grammatical variants. *Frontiers in Psychology* 8. 205. DOI: [10.3389/fpsyg.2017.00205](https://doi.org/10.3389/fpsyg.2017.00205).
- Lindstad, Arne M., Anders Nøklestad, Janne B. Johannessen & Øystein A. Vangsnes. 2009. The Nordic dialect database: Mapping microsyntactic variation in the Scandinavian languages. In Kristiina Jokinen & Eckhard Bick (eds.), *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA 2009)* (NEALT Proceedings Series Volume 4). Northern European Association for Language Technology (NEALT).
- Lødrup, Helge. 2011. Hvor mange genus er det i oslo-dialekten? [‘how many genders does the oslo dialect have?’] *Maal og Minne* 2. 120–136.
- Lovdata. No date. *Talemålsparagrafen*.
- Mæhlum, Brit. 2009. Standardtalemål? Naturligvis! En argumentasjon for eksistensen av et norsk standardtalemål. *Norsk Lingvistisk Tidsskrift* 27(1). 7–26.
- Mæhlum, Brit & Unn Røyneland. 2012. *Det norske dialektlandskapet: Innføring i norske dialekter*. Oslo: Cappelen Damm Akademisk.
- Mancilla-Martinez, Jeannette & Michael J. Kieffer. 2010. Language minority learners’ home language use is dynamic. *Educational Researcher* 39(7). 545–546.
- Mosfjeld, Inger M. 2017. Startalder og andrespråkstilegnelse av finittet. *NOA Norsk som andrespråk* 2. 65–87.
- Opsahl, Toril & Ingvild Nistov. 2010. On some structural aspects of Norwegian spoken among adolescents in multilingual settings in Oslo. In Pia Quist &

2 L2 acquisition in a rich dialectal environment

- Bente A. Svendsen (eds.), *Multilingual urban Scandinavia: New linguistic practices*, 49–64. Bristol: Multilingual Matters.
- Ortega, Lourdes. 2019. SLA and the study of equitable multilingualism. *The Modern Language Journal* 103. 23–38. DOI: [10.1111/modl.12525](https://doi.org/10.1111/modl.12525).
- Prévost, Philippe & Lydia White. 2000. Missing surface inflection or impairment in second language? Evidence from tense and agreement. *Second Language Research* 16(2). 103–133.
- Ragnhildstveit, Silje. 2017. *Genus og transfer når norsk er andrespråk: Tre korpus-baserte studier*. Bergen: University of Bergen. (Doctoral dissertation).
- Rodina, Yulia & Marit Westergaard. 2013. The acquisition of gender and declension class in a nontransparent system: Monolinguals and bilinguals. *Studia Linguistica* 67(1). 47–67.
- Rodina, Yulia & Marit Westergaard. 2015a. Grammatical gender in bilingual Norwegian-Russian acquisition: The role of input and transparency. 20(1). 197–214. DOI: [10.1017/S1366728915000668](https://doi.org/10.1017/S1366728915000668).
- Rodina, Yulia & Marit Westergaard. 2015b. Grammatical gender in Norwegian: Language acquisition and language change. *Journal of Germanic Linguistics* 27(2). 145–187.
- Røyneland, Unn. 2009. Dialects in Norway: Catching up with the rest of Europe? *International Journal of the Sociology of Language* 196/197. 7–30.
- Sandøy, Helge. 2009a. Language culture in Norway: A tradition of questioning standard language norms. In Nikolas Coupland & Tore Kristiansen (eds.), *Standard languages and language standards in a changing Europe*, vol. 27, 119–126. Oslo: Novus.
- Sandøy, Helge. 2009b. Standartalemaal? Ja, men...! *Norsk Lingvistisk Tidsskrift* 27(1). 27–47.
- Saniei, Andisheh. 2011. *Who is an ideal native speaker?! 2011 International Conference on Languages, Literature and Linguistics IPEDR*. Vol. 26. Singapore: IACSIT Press.
- Schmidt, Lauren B. 2011. *Acquisition of dialectal variation in a second language: L2 perception of aspiration of Spanish /s/*. Bloomington: Indiana University. (Doctoral dissertation).
- Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics* 10. 209–231. DOI: [10.1515/iral.1972.10.1-4.209](https://doi.org/10.1515/iral.1972.10.1-4.209).
- Slabakova, Roumyana. 2016. *Second language acquisition*. Oxford: Oxford University Press.
- Søfteland, Åshild. 2018. Nordisk dialektkorpus: Introduction aux recherches sur les parlers dans le Norden. *Nordiques* 35. 95–115.

Linda Evenstad Emilsen & Åshild Søfteland

- Sollid, Hilde. 2014. Kvensk påvirkning på norsk i nord-troms. *Ottar – Populærvitenskapelig tidsskrift fra Tromsø museum – Universitetsmuseet* 5-2014. 41–47.
- Stjernholm, Karine. 2013. *Stedet velger ikke lenger deg, du velger et sted: Tre artikler om språk i Oslo*. Oslo: University of Oslo. (Doctoral dissertation).
- Stjernholm, Karine & Åshild Søfteland. 2019. Talemål i sør-østfold. *Ideologi, struktur og praksis. Målbryting* 10. 101–131. DOI: [10.7557/17.4947](https://doi.org/10.7557/17.4947).
- Sun, Yayun A. 2008. Vol. 8. Input processing in second language acquisition: A discussion of four input processing models.
- Svendsen, Bente A. & Unn Røyneland. 2008. Multiethnolectal facts and functions in Oslo, Norway. *International Journal of Bilingualism* 12. 63–83.
- Trenkic, Danijela. 2007. Variability in second language article production: Beyond the representational deficit vs. processing constraints debate. *Second Language Research* 23(3). 289–327. DOI: [10.1177/0267658307077643](https://doi.org/10.1177/0267658307077643).
- Trenkic, Danijela. 2009. Accounting for patterns of article omissions and substitutions in second language production. In María del P. García Mayo & Roger Hawkins (eds.), *Second language acquisition of articles: Empirical findings and theoretical implications*, 115–143. Amsterdam: John Benjamins.
- Trudgill, Peter. 2002. *Sociolinguistic variation and change*. Edinburgh: Edinburgh University Press.
- Vangsnes, Øystein A. & Marit Westergaard. 2014. Ka korpuse fortæll? Om ordstilling i hv-spørsmål i norske dialekter. In Janne B. Johannessen & Kristin Hagen (eds.), *Språk i norge og nabolanda: ny forskning om talespråk*, 133–151. Oslo: Novus.
- Westergaard, Marit. 2008. Acquisition and change: On the robustness of the triggering experience for word order cues. *Lingua* 118(12). 1841–1863. DOI: [10.1016/j.lingua.2008.05.003](https://doi.org/10.1016/j.lingua.2008.05.003).
- White, Lydia. 2003. Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition* 6. 129–41. DOI: [10.1017/S1366728903001081](https://doi.org/10.1017/S1366728903001081).

Chapter 3

Comparing ERPs between native speakers and second language learners: Dealing with individual variability

Maud Péliissier

University of Agder

Event-related potentials (ERPs) are of great interest in second language acquisition research, as they allow us to examine online language processing and to compare the mechanisms that are engaged to process a first and second language. A long history of research into native language processing has taught us to expect a biphasic pattern in response to syntactic violations, reflecting mechanisms involved first in the automatic and implicit detection of the incongruity and then in the reanalysis and repair of the ungrammatical sentence. However, recent studies show that there is a large degree of individual variability even among native speakers: Instead of this biphasic pattern, most people exhibit one or the other of the two components. This raises an interesting question for second-language research: How do we compare learners and native speakers if there is no unique native-speaker model to compare learners to? In this chapter, I explore two measures that have been put forward to characterise individual variability among native speakers and language learners, the Response Magnitude Index and the Response Dominance Index (Taner et al. 2014), and I show an example of their application to a study comparing native-speaker and non-native-speaker processing of morphosyntactic violations using auditory stimuli instead of visual stimuli.

ERPs; individual differences; second language learners; RMI; RDI

1 Introduction

A large part of research in second language acquisition (SLA) is devoted to comparing learners' performance to native speakers' performance, including measures of online and offline production and perception of a variety of more or



Maud Pélissier

less complex language structures, in order to see how learners may differ at various levels of proficiency. One of the fundamental questions in SLA research is whether learners can and do process their second language (L2) like native speakers (i.e., by engaging the same cognitive and cerebral mechanisms). The development of affordable imagery techniques like electroencephalography (EEG), which records the electric activity of the brain at the surface of the scalp, has given researchers a window into language processing in real time, as opposed to the more indirect measures provided by behavioural experiments. There is an abundant literature on whether L2 learners can eventually recruit the same cognitive mechanisms – as reflected by different event-related potential (ERP) components – as native speakers in order to process syntax in particular, but no definitive answer has been agreed upon. Some researchers claim that syntactic processing in the L2 will never be as automatic and implicit as in the first language (L1) because adults do not have the same access to procedural learning as children before the age of five or six do (e.g., [Birdsong 2006](#); [Clahsen & Felser 2006](#); [2018](#); [Paradis 2009](#)), while others claim that grammatical processing can eventually recruit the same mechanisms when learners attain high proficiency ([Steinhauer et al. 2009](#)). This question has been rendered even more difficult by recent research showing that native speakers do not all use the same mechanisms to process syntax ([Tanner et al. 2013](#); [2014](#); [Tanner & van Hell 2014](#); [Tanner 2019](#)). “Shallow” parsing, where language users do not build a deep syntactic hierarchical structure in real time but instead use lexico-semantic clues to process grammatical information, is not uniquely characteristic of L2 processing (as had been previously hypothesised, [Clahsen & Felser 2006](#)) but also applies to some native speakers. Since then, several studies have attempted to determine what causes this individual variability among native speakers, finding some interesting leads.

In this chapter, I will first give an overview of what event-related potentials are and how they have been used in SLA research. I will then review how individual differences in ERPs have been characterised among native speakers and learners. Finally, I apply those different measures to the analysis of native-speaker and L2 data.

2 The research paradigm in electro-encephalography experiments

2.1 What are ERPs?

ERPs are voltage changes in the electric activity of the brain that are linked to the occurrence of specific events, such as the presentation of a word, sound,

3 *Comparing ERPs between native speakers and second language learners*

or, in many experiments, a grammatical violation (Fabiani et al. 2007; van Hell & Tokowicz 2010). They are obtained from the analysis of EEG data, which is recorded from a number of electrodes placed at different locations on the scalp. The signal is then averaged across many trials to cancel out the unwanted noise due to non-experiment-related brain activity but also to other sources of electrical activity, such as muscle movements, skin potentials, or electrical appliances in the room. ERPs are post-synaptic potentials happening across millions of neurons at the same time. Not all cognitive processes have an ERP signature: To be visible, the activity needs to come from a large number of neurons oriented in the same direction, which most often happens in the pyramidal cells of the cortex (Osterhout et al. 2004; Luck 2014). ERPs are a series of negative and positive peaks over time that are characterised as components depending on their polarity (positive/negative), latency (in milliseconds), and distribution (on the surface of the scalp). These components are thought to reflect cognitive processes.

ERPs are frequently used to study language processing for several reasons (see Kaan 2007; Luck 2014). EEG enables the recording of continuous data, from before the stimulus is presented, until after a response is given. This means that data are recorded during stimulus processing, instead of just after the response to the stimulus as is the case in behavioural experiments. This gives the researcher a window into the processes of interest instead of only their consequences. ERPs also have an excellent temporal resolution, as a sample is recorded every 1 or 2 ms, which makes them particularly suited to study fast online processing, such as spoken-language processing. This excellent timing also makes it possible to see different processes happening simultaneously and to target one in particular through experimental manipulations. Another advantage is that a behavioural response is not needed, although it is standard to obtain one. This means that studies can be conducted with populations from whom it is difficult to get a response (e.g., newborns or patients), or when this would affect treatment, for instance in studies focusing on attention.

Several ERP components are of particular interest for the study of (second) language processing. The first one is the Left-Anterior Negativity or LAN. This negative deflection peaks between 300 and 500 ms after the violation and is maximal at anterior sites, most often bilaterally, but sometimes lateralised to the left. It has mostly been observed in response to word-category violations (e.g., Weber-Fox & Neville 1996; Isel et al. 2007; Bowden et al. 2013) or morphosyntactic violations in sentence contexts (e.g., Ojima et al. 2005; Rossi et al. 2006; Chen et al. 2007; Gillon-Dowens et al. 2010; Molinaro et al. 2011; Alemán Bañón et al. 2014). The LAN is thought to reflect the automatic detection of rule-governed morphosyntactic violations (Gunter et al. 2000; Morgan-Short et al. 2015). Some

Maud Pélissier

claim that it may sometimes reflect working memory load (Kaan 2007), although others argue that this component differs from working memory-related negativities (Martín-Loeches et al. 2005). The LAN is, however, not elicited systematically, and has not been found in contexts where it was expected, for instance with subject-verb, number or gender agreement violations (Bond et al. 2011; Foucart & Frenck-Mestre 2012).

A second important component for language processing is the N400, a large centro-parietal negativity peaking around 400 ms after the violation. It is usually associated with lexico-semantic processing. It follows all lexical words, but is larger for words that are hard to predict or integrate in the context (Kutas & Hillyard 1980; Federmeier 2007; Kutas et al. 2011). However, it can also be elicited by a large range of syntactic incongruities, such as violations of word category (Weber-Fox & Neville 1996; Guo et al. 2009; Kotz 2009), and subject-verb (Xue et al. 2013; Tanner & van Hell 2014; Tanner et al. 2014) or number agreement (Osterhout et al. 2006; Batterink & Neville 2013). The N400 reflects the semantic integration of a word in its context, pre-semantic processing and access to semantic knowledge (Morgan-Short et al. 2015; Isel 2017). It could be related to the retrieval of information from declarative memory. The fact that it follows syntactic violations suggests that some language users rely on lexico-semantic cues rather than rule-based strategies to process syntax.

The final major component is the P600. It is triggered by a large variety of phenomena. The P600 is a positive deflection, maximal at parietal electrodes between 600 and 900 ms, or as early as 500 ms with auditory stimuli (Osterhout & Holcomb 1992; Qi et al. 2017). It follows word-category violations (e.g. Friederici 2002; Pakulak & Neville 2010; Batterink & Neville 2013; Bowden et al. 2013) and all types of agreement violations (e.g., Tokowicz & MacWhinney 2005; Osterhout et al. 2006; Gillon-Dowens et al. 2011; Batterink & Neville 2013; Tanner et al. 2014; Alemán Bañón et al. 2017). The P600 is influenced by several factors (Morgan-Short et al. 2015). Its amplitude is reduced when violations are more frequent in the input (Sassenhagen et al. 2014), and it only appears when attention to form is necessary for the task at hand. The P600 reflects late and controlled analysis and repair processes that follow the detection of an anomaly, which makes a word difficult to integrate in the current structure (Friederici 2002; Kaan 2007; Caffarra et al. 2015; Morgan-Short et al. 2015). It is also associated with the costs of monitoring, checking and reprocessing the input (van de Meerendonk et al. 2009).

(Morpho)syntactic violations usually elicit a biphasic pattern among native speakers: A LAN followed by a P600 (Friederici 2002). This pattern is hypothesised to reflect the succession of two distinct stages of syntactic processing: (1)

3 Comparing ERPs between native speakers and second language learners

the automatic, implicit detection of the morphosyntactic incongruity and (2) the more conscious and controlled reanalysis processes engaged to repair the input for interpretation.

2.2 How are ERPs used in SLA research?

In SLA research, ERPs are generally used to compare native speakers to L2 learners with specific characteristics. Many studies have looked at how the age of acquisition of the L2 (Weber-Fox & Neville 1996; Hakuta et al. 2003) or proficiency (Ojima et al. 2005; Rossi et al. 2006; Steinhauer et al. 2009; Tanner et al. 2009; 2013; 2014; McLaughlin et al. 2010) impact the different ERP components. The effects of the similarity between L1 and L2 and of potential transfer effects have also been extensively studied (Tokowicz & MacWhinney 2005; Chen et al. 2007; Foucart & Frenck-Mestre 2010; 2012; Gillon-Dowens et al. 2010).

ERPs are time-locked to a specific event that is used to synchronise electrical activity across trials. In SLA research, this event is usually a type of syntactic incongruity, such as a violation of phrase structure (**I have many run to miles this week*, e.g., Rossi et al. 2006; Kotz et al. 2008; Bowden et al. 2013), gender agreement (Gillon-Dowens et al. 2010; Foucart & Frenck-Mestre 2012), number or person agreement (e.g., Rossi et al. 2006; Tanner et al. 2009; Tanner & van Hell 2014; Alemán Bañón et al. 2014; 2017). It can also be a semantic incongruity, when a word that is implausible or incoherent (*She slept in my *law that night*) is integrated into a sentence context (Kutas & Hillyard 1980; Friederici et al. 1993; Astésano et al. 2004; Ojima et al. 2005; Weiss et al. 2005; DeLong et al. 2014; Foucart et al. 2014; Schneider et al. 2016).

To compare the ERPs elicited by violations in native and non-native speakers, researchers look at two parameters. The first one, more qualitative, is the absence or presence of certain components. For instance, many studies have found that violations do not elicit a LAN for lower intermediate learners (Ojima et al. 2005; Hahne et al. 2006; Rossi et al. 2006; Chen et al. 2007). Sometimes, even the P600 is missing. For instance, Foucart & Frenck-Mestre (2010) found that violations of noun-adjective gender agreement in the plural in French, which do not exist in their participants' L1 (German¹), did not elicit the expected P600 in learners, whereas violations of a common structure (determiner-noun gender agreement) triggered a similar P600 in native and non-native speakers. Structures relying on cues that conflict with each other across the learners' L1 and

¹Although noun-adjective gender agreement does exist in German, all gender distinctions for adjectives and determiners are neutralised in the nominative plural case, which was used in this experiment.

Maud Pélissier

L2 (e.g., a different word order) have also been found to trigger an N400 instead of a P600 (Foucart & Frenck-Mestre 2012), just like agreement violations do in beginners as opposed to more advanced learners (Osterhout et al. 2006; McLaughlin et al. 2010). Osterhout et al. (2006) conducted a longitudinal study over one academic year with English learners of French. They tested learners' processing of agreement violations such as *Tu adores*/**adorez le français* ('You love_{2ND-PERSON SING. INFORMAL}/*love_{2ND-PERSON SING. FORMAL OR PLURAL French}') after one, four and eight months of university classroom instruction. They found that the initial N400 elicited by the violations evolved into a P600 when proficiency increased – after a relatively short time of instruction.

The second, more quantitative, parameter of interest is the amplitude and latency of the components. The P600 is often delayed and smaller among less proficient learners (Rossi et al. 2006; McLaughlin et al. 2010; White et al. 2012; Batterink & Neville 2013; Tanner et al. 2014). If the P600 is similar when structures work in an equivalent way in participants' L1 and L2 (Tokowicz & MacWhinney 2005; Foucart & Frenck-Mestre 2010), its distribution can change from posterior to anterior when the structure is specific to the L2 (Foucart & Frenck-Mestre 2012). Many studies have thus shown that the electrophysiological correlates of language processing are or can be different in an L2 and in an L1, especially at low levels of proficiency.

2.3 Individual variability in ERPs

In native-language processing, syntactic violations are expected to reliably elicit a biphasic pattern: A LAN followed by a P600 (Friederici 2002). This pattern has indeed been observed among native speakers for different sorts of syntactic incongruities and in a variety of languages (Ojima et al. 2005; Chen et al. 2007; Mueller et al. 2007; Newman et al. 2007; Molinaro et al. 2008; Bowden et al. 2013), even though the negativity is sometimes bilateral (Isel & Kail 2018) or posterior and more N400-like (Zawiszewski et al. 2011). However, recent research shows that this pattern is not found in all native speakers. Pakulak & Neville (2010) investigated language processing among a more variable population than the college students who usually participate in experiments. They found that native speakers who were less literate had a more bilateral LAN and a reduced P600 to syntactic violations.

Osterhout (1997); Tanner et al. (2013; 2014) and Tanner & van Hell (2014) have shown that there are individual differences even among highly literate native speakers and that these differences go beyond dissimilarities in amplitude and latency. Their data reveal that the traditionally expected biphasic pattern is not

3 Comparing ERPs between native speakers and second language learners

characteristic of most participants' response. Instead, most native speakers exhibit either a P600 or an N400-like response. They suggest that the presence of an anterior negativity at the level of the group is in fact an artefact due to the occurrence, at the same time, of a posterior P600 and a largely distributed N400 across participants, as the P600 has already started in the N400/LAN time window (300–500 ms after the violation). Tanner and his colleagues found a reliable negative correlation between N400 and P600 effects, revealing that most native speakers show one or the other but not both components. These four studies used the traditional visual method of stimuli presentation – the Rapid Serial Visual Presentation or RSVP – in which a word is presented on the screen for a short time (usually around 350 ms) and followed by a blank screen (usually for 100 ms). As this reading paradigm is not very ecological, Tanner (2019) reproduced earlier studies with a self-paced reading task, in which the participant reads a sentence word by word but decides when to move on to the next word, and found the same neurocognitive individual differences. Tanner (2019: 232) thus notes that the successive biphasic pattern “cannot necessarily be taken as strong evidence for serial, stage-based processes of agreement comprehension in the broader population”. Instead, readers seem to adopt different processing strategies. Those who exhibit an N400-dominant response may rely more on word-based predictions of upcoming words, while a P600-dominant response could reflect the engagement of combinatorial mechanisms (Tanner & van Hell 2014).

If there is such variability among native speakers, then there is no consistent native model to compare learners to, and exhibiting only an N400 or a P600 in response to syntactic violations cannot be considered the mark of low proficiency or of deficient processing. How can we then compare native speakers and non-native speakers?

3 Characterising individual differences among native speakers

The first step is to adequately characterise individual differences among native speakers and to determine what causes them. To that effect, Tanner et al. (2014) introduced two new measures: The Response Magnitude Index (RMI) and the Response Dominance Index (RDI).

3.1 Effect magnitude: The Response Magnitude Index

A first way to characterise individual differences is to look at correlations between the amplitude of the effect, whichever its direction (positive or negative),

Maud Pélissier

and other predictors such as proficiency. The RMI captures the size of the effect, and reflects the listener's sensitivity to the critical violation. Larger RMI values indicate greater neural response and thus higher sensitivity. The RMI is computed according to the formula in , where $N400_{\text{Gram}}$ and $P600_{\text{Gram}}$ refer to the mean amplitude in the chosen time window after grammatical stimuli and $N400_{\text{Ungram}}$ and $P600_{\text{Ungram}}$ to the mean amplitude following ungrammatical stimuli. For both effects, the amplitudes are averaged over a centro-parietal region of interest (ROI; C3, Cz, C4, P3, Pz, P4² in [Tanner et al. 2014](#)). In [Tanner et al. \(2014\)](#)'s study, the critical time windows were 400–500 ms for the N400 effect and 500–1000 ms for the P600 effect. The details of which time windows and electrodes were chosen for RMI and RDI analyses by the different studies that have used these measures is reported in Table 1.

$$(1) \quad \sqrt{(N400_{\text{Gram}} - N400_{\text{Ungram}})^2 + (P600_{\text{Ungram}} - P600_{\text{Gram}})^2}$$

The RMI has mostly been used to look at L2 learners, and as a consequence there are no real data on what influences the magnitude of the overall response among native speakers. However, several studies have found correlations between the amplitude of one effect (N400 or P600) and different proficiency measures. [Pakulak & Neville \(2010\)](#), who investigated participants with a broad range of literacy levels, found that the amplitude of the P600 and the laterality of the LAN in response to phrase structure violations correlated with proficiency in the L1. [Mehravari et al. \(2017\)](#) also observed a correlation between the amplitude of the P600 and measures of reading skills. However, [Tanner et al. \(2013\)](#) failed to find a significant correlation between the amplitude of the P600 and sensitivity indexes (d') scores³ on a grammaticality judgment task (GJT) among native speakers.

The magnitude of the N400 in response to semantic anomalies has also been found to correlate with proficiency measures ([Newman et al. 2012](#)) and to reflect

²These identify individual electrodes. The letter correspond to the position of the electrode(C: Central, CP: Centro-Parietal, P: Parietal), and the number refers to the laterality: z Electrodes are on the central line, smaller numbers are closer to the midline and larger numbers to the ears. Odd numbers are on the left side.

³The sensitivity index is used in signal detection theory to provide a measure of how sensitive someone is to the presence of the signal to be detected, independently of individual participants' strategies such as always replying "correct". It is the standardized difference between the means of the False Alarm and Hit rates. The Hit rate is the probability of correctly detecting the signal (here, accepting grammatically acceptable sentences) and the False Alarm rate is the probability of incorrectly detecting the signal when it is not present (here, rejecting grammatically correct sentences).

3 Comparing ERPs between native speakers and second language learners

Table 1: Parameters used to compute the RDI and RMI in language studies

Paper	Index	Type of violation	N400 window (ms)	P600 window (ms)	ROI
Tanner et al. (2012)	RDI & RMI	Subject-verb agreement	400–500	500–1000	Midline electrodes
Tanner et al. (2014)	RDI & RMI	Subject-verb agreement	400–500	500–1000	C3, Cz, C4, P3, Pz, P4
Tanner & van Hell (2014)	RDI	Subject-verb agreement and verb tense	300–500	500–800	C3, Cz, C4, CP1, CP2, P3, Pz, P4
Faretta-Stutenberg & Morgan-Short (2018)	RDI & RMI	Phrase structure	300–500	600–900	C3, Cz, C4, P3, Pz, P4
Grey et al. (2017)	RDI	Subject-verb agreement and verb tense	300–500	500–800	C3, Cz, C4, CP1, CP2, P3, Pz, P4
Tanner (2019)	RDI	Subject-verb agreement	300–500	500–800	C3, Cz, C4, CP1, CP2, P3, Pz, P4

Maud Pélissier

the lexical and semantic predictability of an item (Federmeier & Kutas 1999; DeLong et al. 2005; Federmeier 2007). These correlations, however, affect the semantic N400 rather than the centro-posterior negativity found after some syntactic violations among native speakers. It is thus relatively unclear what determines the amplitude of the effect among native speakers, although proficiency does seem to play a role.

3.2 Effect dominance: The Response Dominance Index

A second way to look at individual differences is to focus on the direction of the effect, whatever its size. The RDI captures the polarity of the effect and gives information about response dominance and therefore possibly about the type of cognitive mechanisms recruited to process the incongruity. The RDI is computed according to the formula in , where $N400_{\text{Gram}}$ and $P600_{\text{Gram}}$ refer to the mean amplitude in the chosen time window after grammatical stimuli and $N400_{\text{Ungram}}$ and $P600_{\text{Ungram}}$ to the mean amplitude following ungrammatical stimuli (Tanner et al. 2014). RDI values close to zero signal equal-sized N400 and P600 effects, whereas negative and positive values reflect larger negative or positive effects in both time windows, respectively.

$$(2) \quad \frac{(P600_{\text{Ungram}} - P600_{\text{Gram}}) - (N400_{\text{Gram}} - N400_{\text{Ungram}})}{\sqrt{2}}$$

The different parameters that might influence response dominance are of great interest in the study of individual differences and have been the focus of several studies. A first possible candidate is proficiency. However, this factor does not seem to have a sizeable impact on the RDI – the ERP components elicited by the violations in Tanner (2019) varied, even though the 114 participants were all highly literate monolingual English speakers and similarly proficient in their L1. There was therefore no direct link between proficiency and the type of component elicited by the violation.

Another parameter that has attracted a lot of attention is working memory. Nakano et al. (2010) found that working memory capacity influenced listeners' response to animacy violations in the manipulation of thematic roles (*The dog/*the box bit the mailman*). Verbal working memory was negatively correlated with N400 amplitude but positively correlated with P600 amplitude. Similarly, Kim et al. (2018) exposed participants to semantic anomalies and observed that higher verbal working memory capacities were associated with larger P600 effects and smaller N400 effects, when controlling for spatial working memory and language experience. This is consistent with the observation that learners often exhibit an

3 Comparing ERPs between native speakers and second language learners

N400 where a P600 is expected at the initial stages of learning, when their working memory capacity in the L2 is reduced. This suggests that verbal working memory abilities are positively correlated with the recruitment of computation, reanalysis and repair processes – mechanisms associated with the P600. However, in his large-scale study of highly literate monolinguals, [Tanner \(2019\)](#) did not find a significant correlation between verbal working memory and agreement processing, casting some doubt on the predictive power of that factor for response dominance.

Very recent studies also suggest that response dominance could be largely influenced by familial sinistrality. Familial sinistrality refers to the number of someone's close blood relatives (parents, siblings, grandparents) that are left-handed. [Tanner & van Hell \(2014\)](#) first suggested that this parameter was of importance when it was found to be the only significant predictor in a model including operation span measures, cognitive control measures, proficiency scores, lexical processing speed, and familial left-handedness as explanatory variables and the RDI as the dependent variable. More recently, [Grey et al. \(2017\)](#) extended these findings by focusing specifically on the impact of familial sinistrality on agreement processing. They investigated 60 monolingual English speakers while they read sentences containing subject-verb agreement (*The clerk at the clothing boutique was/*were severely underpaid and unhappy*) and verb tense violations (*The crime rate was increasing/*increase despite the growing police force*). 20 participants were right-handed and had no left-handed close family member, 20 were right-handed and had a left-handed close blood relative, and 20 were left-handed. The first group exhibited only a P600 in response to morphosyntactic violations, with low interindividual variability, whereas both the left-handed and right-handed with left-handed blood relatives groups showed a biphasic N400-P600 pattern in the grand average. Variability in these groups was high, with roughly half of the participants showing a P600 only and the other half an N400 only. The authors conclude that left-handedness is associated with increased reliance on lexical/semantic mechanisms instead of combinatorial morphosyntactic ones. However, [Wampler \(2017\)](#) did not find any relationship between handedness or sinistrality of close relatives and the dominance of the response.

The exact factors determining the direction of ERP responses to morphosyntactic violations in native speakers are still to be determined. Working memory seems to play a role but not in all cases, and although familial sinistrality looks promising, replications of the findings by [Grey et al. \(2017\)](#) are needed. Nevertheless, the RDI and RMI have also been used to investigate individual differences among L2 learners.

Maud Pélissier

4 Individual differences among L2 learners

4.1 Effect magnitude in the L2

Although little research has been conducted on what influences effect magnitude in native speakers, the same is not true for L2 learners. There is an abundant literature that has tried to correlate in particular the amplitude of the P600 effect with a variety of predictors.

Effect magnitude has thus been repeatedly found to correlate with proficiency. [Tanner et al. \(2009; 2013\)](#) investigated first-year and third-year English-speaking L2 learners of German while they read sentences containing subject-verb agreement violations. Participants also completed a GJT. A positive correlation between the magnitude of the P600 effect and the d' score was significant for first-year learners, neared significance for third-year learners, and was highly significant for all learners when combined. More proficient learners thus exhibited larger P600 effects. There was also a small but significant negative correlation between the amplitude of the N400 and the performance on the GJT – less proficient participants showed larger N400 effects. [Batterink & Neville \(2013\)](#) also found a positive correlation between P600 amplitude and proficiency among native English speakers after just one hour of training in miniature French. [White et al. \(2012\)](#) obtained a similar correlation with Korean and Chinese late L2 learners of English after a 9-week intensive English course, when participants processed violations of regular past-tense, a structure that either did not exist in their L1 (Chinese participants) or worked differently (Korean speakers). A few studies have specifically used the RMI to look at the increase in overall response magnitude rather than at the amplitude of one or the other component. [Tanner et al. \(2012\)](#) and [Tanner et al. \(2014\)](#) found that a larger RMI was associated with higher proficiency, after controlling for age of acquisition, length of residence, frequency of L2 use, and motivation to speak like a native. However, the complete model was not significant. Their results are particularly interesting as they did not reveal an individual correlation between P600 amplitude or N400 amplitude and proficiency – the effect of proficiency was best captured by the overall response magnitude rather than by individual correlations ([Tanner et al. 2014](#)). [Fromont et al. \(2012\)](#) also observed that the RMI grew with proficiency (both N400 and P600 amplitudes increased with competence) among English-speaking learners of French.

Although proficiency is the most studied explanatory factor for effect magnitude, a few other predictors have been identified. [McLaughlin et al. \(2004\)](#) found that the amplitude of the N400 effect to pseudowords in the L2 was highly correlated with the number of hours of instruction received. However, [Tanner et](#)

3 Comparing ERPs between native speakers and second language learners

al. (2013) did not find an effect of hours of exposure on P600 amplitude during subject-verb agreement processing. This factor even removed predictive power in a model including the d' score as a predictive variable. Meulman et al. (2015) used general additive modelling to examine the effects of age of acquisition (AoA) on ERP responses to grammatical gender and non-finite verb violations among Slavic advanced learners of German. They found that AoA impacted the RDI only for one of the two types of violations. The verb tense violations – marked similarly in both languages and considered easy to acquire – elicited a P600 for all learners, independently of AoA. On the contrary, gender agreement violations, an L2-specific structure, were followed by a P600 for earlier learners of German but by an N400 for later learners. The authors conclude that late learners resort to less efficient and less computational strategies to process an L2-specific structure only. Finally, Faretta-Stutenberg & Morgan-Short (2018) found that memory capacity, specifically working memory capacities and procedural learning abilities, accounted for 62% of the variance in the change in RMI following a six-month study-abroad experience. For learners who stayed at home during the same period, declarative memory positively correlated with the magnitude of the response to phrase structure violations.

4.2 Response dominance in the L2

Variability in response dominance between learners has long been interpreted as reflecting differences in proficiency. There is a large literature supporting a qualitative evolution of ERPs elicited by (morpho)syntactic violations, from an N400 at the beginner level to a P600 or a biphasic LAN-P600 pattern at more advanced stages (e.g., Osterhout et al. 2006; Rossi et al. 2006; Kotz 2009; McLaughlin et al. 2010; see also Steinhauer 2014 for a review). Steinhauer et al. (2009)'s model thus postulates that beginners exhibit an N400 in response to syntactic violations because they use more lexico-semantic processes in real time. With increasing proficiency, structures are grammaticalised, which means that learners rely more on computational mechanisms to process them, as indexed by the P600. The P600 is first small and delayed (Tokowicz & MacWhinney 2005; Rossi et al. 2006) but can eventually grow into a nativelike one. Osterhout et al.'s (2006) and McLaughlin et al.'s (2010) longitudinal studies support this convergence hypothesis.

Gender can also influence the RDI: Wampler (2017) found that women were more likely to exhibit a P600 than men in response to L2 French violations, which she interprets as consistent with the idea that women learn L2s more quickly and achieve higher final proficiency.

Maud Pélissier

Response dominance can be affected by learning conditions. [Faretta-Stutenberg & Morgan-Short \(2018\)](#) compared the effect of stay-at-home instruction and a semester abroad on the processing of phrase structure violations in L2 Spanish. There were no ERP effects at the pre-test. At the post-test, they found that some participants in the stay-at-home group exhibited an N400 at the end of the semester while others showed a P600 effect, which suggests that learners developed different language-processing strategies. In the study-abroad group, the RDI shifted to a more negativity-dominant pattern as a group-level N400 appeared at the end of the semester. However, a subset of learners in this group exhibited a P600 effect. The authors note that the N400 effect here is similar to what was found by [Morgan-Short et al. \(2010\)](#) among implicitly-trained participants at an equivalent level of proficiency (75% accuracy on the GJT) – even though highly proficient participants in that study exhibited a biphasic LAN-P600 pattern at the end of training. As a study-abroad experience favours the use of meaning-based communicative strategies ([Tokowicz et al. 2004](#)), these results are consistent with the idea that the RDI depends on processing strategies that can evolve with proficiency and learning conditions.

Finally, [Tanner et al. \(2012\)](#) and [Tanner et al. \(2014\)](#) found that age of arrival and motivation to speak like a native speaker significantly predicted response dominance, in a model including the age of arrival in an L2-speaking country, the length of residence in that country, the frequency of L2 use, proficiency scores and motivation to speak like a native, which as a whole explained 61% and 54% of the variance, in [Tanner et al. \(2012\)](#) and [Tanner et al. \(2014\)](#), respectively. Earlier arrival and a higher motivation to speak like a native were highly correlated with a stronger positivity-dominant response, and these two predictors alone explained 48% of the variance in [Tanner et al. \(2014\)](#).

Although proficiency is generally considered the main predictor for both effect magnitude and response dominance in the L2, it is not the only relevant factor to account for interindividual differences. The role of several predictors has been investigated in the L1 but not yet in the L2, such as the impact of familial sinistrality on the RDI, which could very well play a role in the strategies recruited to process an L2. To our knowledge, only one study has directly compared language users' RDI in their L1 and their L2. [Wampler et al. \(2014\)](#) recorded EEG data from English-speaking second-year learners of French while they read grammatical and ungrammatical sentences in their L1 and L2. They found that their English (L1) RDI was unrelated to their French (L2) RDI – an individual's response dominance in their native language thus does not necessarily predict dominance in the L2. More data are needed to see if this relationship might change with proficiency and, specifically, if the RDI of a highly proficient, native-like L2 learner

3 Comparing ERPs between native speakers and second language learners

would be the same in their L2 and L1 or if they would remain different as learning conditions differ.

5 Comparing learners and native speakers with these measures: An example of application

In this last section, I present an example of application of the RDI and RMI measures to compare learners and native speakers, to see if previous results can be extended to less proficient foreign-language learners, and to a structure other than subject-verb agreement.

5.1 Description of the experiment

EEG data were recorded from 32 intermediate French learners of English (B1-B2 level) and 16 native speakers of English⁴ while they judged the semantic acceptability of stimuli – they were asked if the sentence they had just heard made sense to them. At the end of the experiment, they also completed a separate GJT on similar sentences. The target structure was past tense morphology with auxiliaries. In polar questions, auxiliaries ‘Did’ and ‘Had’ were followed either by a past participle or the base form of the verb, with half of 192 questions being grammatically unacceptable (*Did Mary finish/*finished her dinner?*; *Had Mary finished/*finish her dinner?*). 120 fillers, half of which contained number agreement violations (*Did John govern that/*those country for years?*; *Did John govern those/*that countries for years?*), as well as 120 sentences containing a semantic violation (*Had Mary fired what happened?*) were also included, yielding a total of 432 sentences per participant. Two lists containing the same number of stimuli were created so that each participant only heard one version of each sentence.

For the analysis of individual differences, following [Tanner et al. \(2014\)](#), the P600 effect was quantified as the mean amplitude of the difference between incorrect and correct conditions between 500 and 900 ms after the violation, while the N400 effect was the difference between correct and incorrect conditions⁵ in a 200–400 ms window.⁶ The region of interest was a large centro-parietal area including electrodes C3, Cz, C4, CP1, CP2, P3, Pz and P4.

⁴There were twice as many learners because they were later divided into two training groups. Results of the analyses are reported for illustration purposes but must be interpreted with caution as that is a small number of data points to look at continuous differences.

⁵Note that the difference here goes in the opposite direction from the P600 effect because the N400 effect is a negativity.

⁶This timing is slightly earlier than the one chosen in previous studies because stimuli were

Maud Pélissier

5.2 Grand mean analyses

Grand mean analyses were conducted with linear mixed effect models in R (Team 2019) with package lme4 (Bates et al. 2015). A model with Condition (Congruent/Incongruent), Region (Anterior/Central/Posterior), Hemisphere (Left/Right) and Group (Native speakers/Learners) as fixed effects and with the maximal random structure that would converge (an intercept by participant as well as a slope for Condition and Hemisphere) was fitted to the data. The highest order significant interaction was Condition:Region:Groupe ($F(2.3678) = 7.87, p < 0.001$).⁷ Post-hoc analyses conducted with the package emmeans (Lenth 2019) revealed a significant positive difference between the Incongruent and Congruent conditions in the central ($M_{I-C} = 0.53 \mu V, SE = 0.25, t(147) = 2.17, p = 0.03$) and posterior region ($M_{I-C} = 0.65 \mu V, SE = 0.21, t(77) = 3.11, p = 0.003$) for the native speakers only. Only this group thus exhibited a P600.

5.3 Individual differences: Magnitude

The first step was to examine the correlation between the N400 effect and the P600 effect in learners and native speakers, in order to assess whether participants exhibited one or the other effect instead of the expected biphasic pattern. There was indeed a significant negative correlation between the presence of a P600 and an N400 effect among learners ($r = -0.41, t(30) = -2.49, p < 0.05$) and native speakers ($r = -0.68, t(14) = -3.50, p < 0.01$), which is illustrated in Figure 1, where the blue line shows the best linear approximation for the correlation with a 95% confidence interval. This shows that, consistent with previous studies, most participants exhibited either an N400 (participants to the left/above the dashed line, which represents equivalent N400 and P600 effects) or a P600 (participants to the right/below the dashed line) but not both. This can also be seen in Figure 2, which shows ERP waveforms for P600-dominant and N400-dominant native speakers and learners at Pz, a midline parietal electrode. Note that for P600-dominant learners, there appears to be a separate early positivity in the time window of the N400 before the P600, which suggests the engagement of attention-related mechanisms. The pattern for the native speakers is unusual in that the waveform in the correct condition contains a long-lasting negativity

presented auditorily in this experiment instead of visually as in previous research. The synchronisation point was the beginning of the *-ed* ending on the main verb instead of the beginning of the critical word, therefore reducing the elapsed time between the critical point and the beginning of the cerebral response.

⁷A type III anova was run on the model with a Satterthwaite estimation of the degrees of freedom.

3 Comparing ERPs between native speakers and second language learners

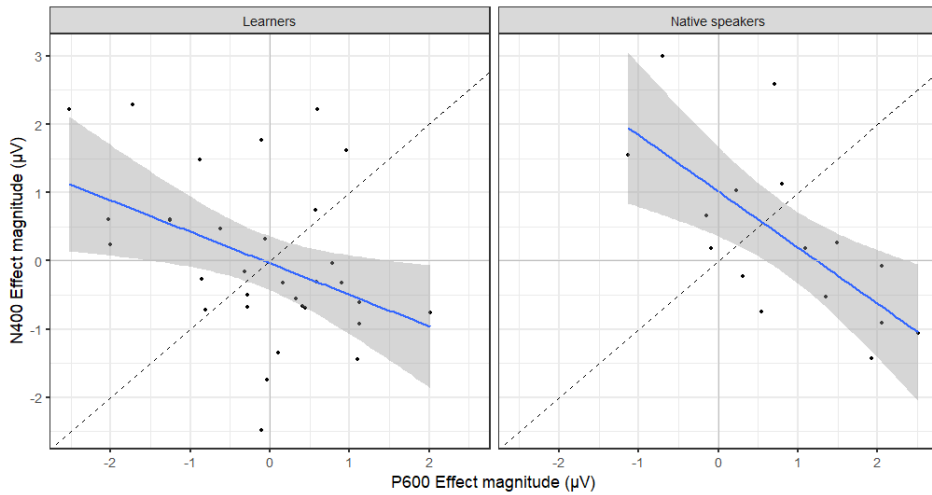


Figure 1: Correlation between N400 and P600 effect magnitudes for learners and native speakers

starting from around 400 ms, which could reflect the cost of maintaining the critical word in memory to judge whether the sentence was acceptable. It is also worth noting that the N400 effect in the N400-dominant group seems to start right before the critical morpheme. This is hard to explain as this means that the difference started before the critical violation. A possible explanation is that there were some slight acoustic differences in the pronunciation of the verbs with and without the morpheme which these participants picked up on and which helped them anticipate the correctness of the word.

The second step was to evaluate the effect of the most studied predictor on GJT response magnitude – proficiency. A sensitivity index (d' score) was computed for performance on the critical sentences – it is therefore a measure of structure-specific proficiency. There was a significant d' difference between the two groups ($t(46) = -8.25, p < 0.001$): Learners were less proficient ($M = 0.80, SD = 1.11$) than native speakers ($M = 3.30, SD = 0.70$). There was no significant correlation between the amplitude of the P600 effect and proficiency for all participants combined ($r = 0.18, t(46) = 1.21, p > 0.2$), nor when learners and native speakers were examined separately (Learners: $r = -0.19, t(30) = -1.04, p > 0.3$; Natives: $r = -0.34, t(14) = -1.33, p > 0.2$). However, there was a general positive correlation between the amplitude of the N400 effect and the d' score ($r = 0.38, t(46) = 2.80, p < 0.01$, see Figure 3). Participants who were more adept at detecting critical violations were thus more likely to exhibit an N400

Maud Pélissier

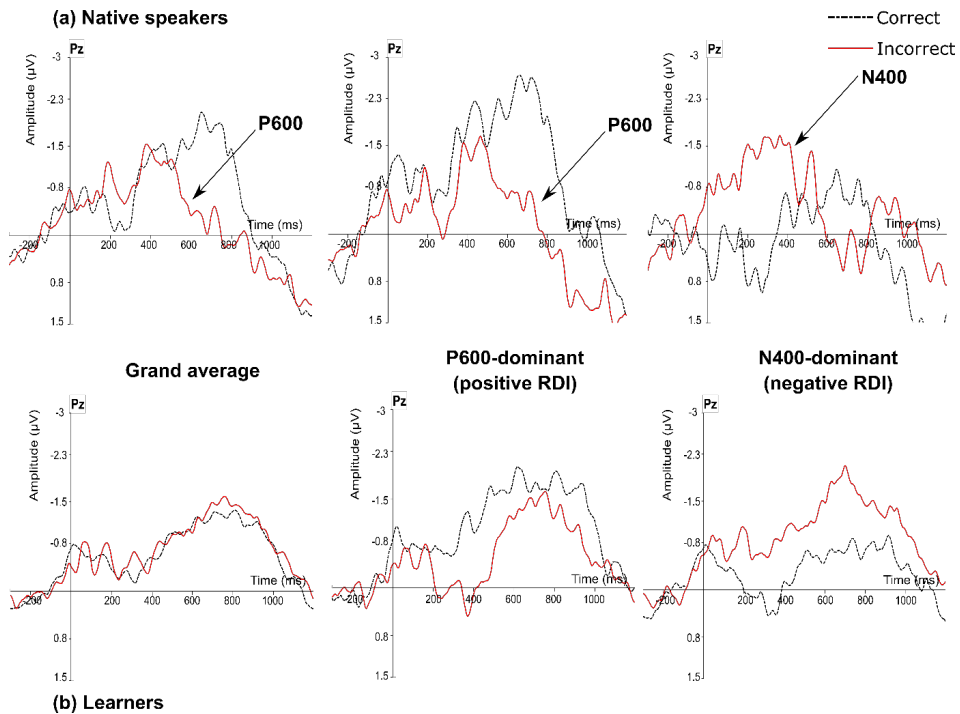


Figure 2: ERP waveforms to Correct (black dashed line) and Incorrect (red solid line) stimuli per group (Native speakers vs. Learners) for all participants and both RDI subgroups (P600-dominant and N400-dominant) at electrode Pz (midline parietal electrode)

than a P600. This goes in the opposite direction from what we normally expect, which is that more proficient participants (especially as evaluated on a task that targets explicit knowledge like the GJT does) will show a P600 following syntactic violations. A separate correlation test for grammatical items revealed a similar positive correlation ($r = 0.36$, $t(46) = 2.58$, $p < 0.05$).

Participants who accepted more correct items exhibited a larger N400, while the correlation with ungrammatical items neared significance ($r = 0.27$, $t(46) = 1.91$, $p = 0.06$), which is even more unexpected. When groups were examined separately, the correlation between the d' score and the amplitude of the N400 effect was significant for learners ($r = 0.46$, $t(30) = 2.82$, $p < 0.01$) but not for native speakers. It is interesting to note that there was no significant difference in the amplitude of the N400 between the two groups ($t(46) < 1$) but a difference in the amplitude of the P600 effect ($t(46) = -2.89$, $p < 0.01$): Native speakers had a much larger P600 effect ($M = 0.81$, $SD = 1.05$) than learners, who did not

3 Comparing ERPs between native speakers and second language learners

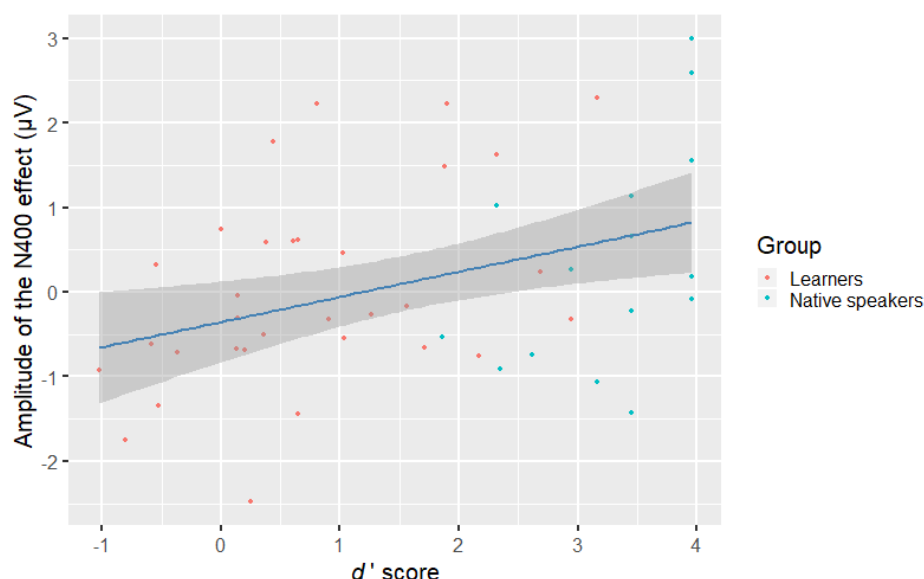


Figure 3: Amplitude of the N400 effect as a function of the d' score

exhibit a reliable P600 in response to violations ($M = -0.12$, $SD = 1.06$). Native speakers were more proficient and showed a significant P600 as a group, but among learners, more proficient participants tended to exhibit a larger N400.

The RMI was also computed. However, the correlation between RMI and d' score did not reach significance ($r = 0.24$, $t(46) = 1.71$, $p = 0.09$). There was no significant difference in RMI between the two groups ($t(46) < 1$; $M_{\text{Natives}} = 1.61 \mu\text{V}$, $SD_{\text{Natives}} = 0.87 \mu\text{V}$; $M_{\text{Learners}} = 1.37 \mu\text{V}$, $SD_{\text{Learners}} = 0.76 \mu\text{V}$), despite the difference in proficiency (as reflected by the d' score). In our case, the results were thus best explained by a simple relationship between the amplitude of the N400 and proficiency, rather than by a link between proficiency and the magnitude of the response in general.

These findings are surprising, as proficiency has previously been associated with a larger P600 amplitude or a more positive RMI in general. Native speakers' performance was at ceiling, with a mean accuracy of 92.81% ($SD = 14.60\%$) on grammatical items and 95.31% ($SD = 10.24\%$) on ungrammatical items – with a median at 100% for both. However, there was much more variability among learners: They did relatively well on grammatical items ($M = 79.06\%$, $SD = 15.37\%$, median = 85%, range = 50–100%) but were much less accurate on ungrammatical sentences ($M = 45\%$, $SD = 27.85\%$, median = 35%, range = 5–100%). For them, better proficiency was associated with a more negative-going response. This may

Maud Pélissier

be due to their overall proficiency in English, which was lower-intermediate. At this proficiency level, it is not uncommon for learners to exhibit an N400 after syntactic violations. They may have only reached the second stage of [Steinhauer et al.](#)'s (2009) model: After showing no response at all to syntactic violations, relatively more proficient learners show an N400 effect, which will evolve into a P600 with proficiency, like the one native speakers exhibit as a group.

5.4 Individual differences: Response dominance

The next step was to look at the Response Dominance Index. There was no significant difference in RDI between the two groups ($t(46) < 1$). The RDI was also not correlated with the d' score when participants were grouped together ($t(46) < 1$). However, for learners, the RDI correlated with the d' score ($r = -0.39$, $t(30) = -2.33$, $p < 0.05$), which is consistent with the relationship that was found between the N400 effect and proficiency: More accurate participants were more likely to exhibit a negative-going effect rather than a P600. This correlation was driven by the performance on grammatical items, which was itself correlated with the RDI ($r = -0.36$, $t(30) = -2.14$, $p = 0.04$). The processing of grammatical items is thought to engage implicit knowledge ([Roehr-Brackin 2015](#)), and it is worth noting that participants trained implicitly on artificial languages in studies by [Morgan-Short et al. \(2010; 2012\)](#) also exhibit an N400 at an intermediate proficiency level.

For native speakers however, the RDI- d' correlation did not reach significance ($r = -0.44$, $t(14) = -1.84$, $p = 0.09$). To understand this unexpected finding, one must keep in mind that during the EEG recording, participants did not complete the GJT but a semantic acceptability judgment task. I noticed that native speakers had difficulties with that task, specifically with ignoring grammatical incongruities in semantically acceptable sentences. Although their performance on the GJT was at ceiling, it is possible that their performance on the EEG task influenced the type of processing strategies they engaged in. To test this hypothesis, I computed a semantic dv from the semantic acceptability task,⁸ reflecting how well native speakers managed to focus on the semantic aspect of sentences. The semantic d' does not provide a measure of structure-specific proficiency, which is why the d' of the GJT was used in the original analyses. Participants with a high semantic d' score accepted sentences containing a grammatical violation but no semantic incongruity, while participants with a low semantic d'

⁸In the semantic d' , hits were sentences correctly identified as semantically correct, which could contain a syntactic violation. Sentences containing a semantic violation were fillers, and always syntactically correct.

3 Comparing ERPs between native speakers and second language learners

score tended to reject ungrammatical items as semantically unacceptable. There was a strong correlation between the RDI and the semantic d' score ($r = -0.56$, $t(30) = -3.74$, $p < 0.001$): Participants who had a lower semantic d' and thus focused more on the grammatical aspects of the stimuli had a more positive RDI and therefore exhibited a P600. Following this, I divided all participants (native speakers and learners) into two groups corresponding to a high or low semantic d' .⁹ A t -test comparison between the two groups revealed that participants who had a lower semantic d' also showed a larger (and positive) RDI ($t(157) = 3.64$, $p < 0.001$; $M_{\text{LowSemD}'} = 0.37V$, $SD_{\text{LowSemD}'} = 1.21 \mu V$; $M_{\text{HighSemD}'} = -0.38 \mu V$, $SD_{\text{HighSemD}'} = 1.39 \mu V$). Rather than group membership (learners vs. native speakers), what seems to have influenced the type of electrophysiological response the most is participants' attitude on the task and the strategy they chose to adopt. This is in line with the hypothesis that even native speakers do not all use the same mechanisms to process language. Participants who had difficulties ignoring the grammatical incongruities present in the input exhibited a P600 in response to the violations, because their attention was attracted to them and because they used combinatorial mechanisms even when processing language for meaning. On the contrary, participants who had a high semantic d' successfully ignored grammatical violations to only reject semantically unacceptable sentences. In the case of learners, this success might simply be a correlate of the fact that they had great difficulties detecting violations, as their performance on the GJT suggests. However, there was no correlation between the semantic d' and the capacity to detect violations (as measured by the performance on ungrammatical items; $t(126) < 1$), so learners who obtained a high semantic d' did not reach it just because they could not perceive the ungrammaticalities. There is no doubt that native speakers perceived the violations; those who performed well on the semantic acceptability task did so because they focused more on the lexico-semantic aspects of language, which is consistent with the fact that they exhibited much larger N400 effects.

The task completed by participants during EEG data acquisition may well influence the RDI. Commonly used GJTs focus attention on form and may increase the likelihood of observing a P600, especially when stimuli are presented with the traditional and yet very artificial method of the rapid serial visual presentation. [Tanner \(2019\)](#) found similar results to those previously reported with his more ecological self-paced reading presentation – but with a simultaneous GJT. In my experiment, I had participants process stimuli for meaning instead of form, which

⁹The chosen splitting point was the median, as Hartigan's dip test for unimodality did not reveal a multimodal distribution of the data ($D = 0.03$, $p > 0.1$).

Maud Pélissier

let them use slightly more natural processing strategies. Not all language users grant the same importance to grammar in their native language, and this is evident from the results of our individual differences analyses. When given a choice, some people have no difficulties ignoring ungrammatical sentences, because they use other cues – lexico-semantic cues, as it appears – to interpret meaning, while others cannot do without combinatorial syntactic processes. Unfortunately, I do not have data concerning the number of left-handed close relatives of our native speakers, but this parameter may explain the differences in processing strategies (Grey et al. 2017). Using a less explicit task than a GJT proved to be of interest for studying individual differences, as it brought forth differences in strategies used to process meaning and not just form.

6 Conclusion

Comparing the electrophysiological correlates of language processing between learners and native speakers is proving difficult due to a high degree of individual variability even among native speakers. The traditional biphasic pattern of the LAN (or N400) followed by a P600 seems not to be representative of most individuals' responses to morphosyntactic violations – our data extend findings obtained with agreement and phrase structure violations to tense morphology incongruities. The RMI did not prove a valuable measure for our data: Proficiency was associated with a larger amplitude of the N400 effect specifically. More research is needed to determine why in some cases proficiency is associated with the amplitude of a specific component (e.g. Tanner et al. 2009; 2013; White et al. 2012), whereas at other times it is reflected by the general amplitude of the response. The RDI is a useful way of qualifying the type of response elicited by the violations, which reflects the strategy recruited by language users. Response dominance has long been indirectly associated with learners' proficiency, with models proposing an evolution from no response to an N400 and finally a P600 (Steinhauer et al. 2009). In our data, the RDI was directly associated with proficiency – among our group of intermediate learners, which can be hypothesized to be at the intermediate stage, more accurate learners exhibited more negative responses. But the most significant predictor in our case was participants' strategy to complete the task, as measured by the performance on the semantic acceptability judgment task.

Individual differences among native speakers question the traditional syntax-first model (e.g. Friederici 2002) – there is not one single processing route that is nativelike, even when processing language for meaning and not to monitor

3 *Comparing ERPs between native speakers and second language learners*

grammatical incongruities. An important next step will therefore be to understand why this is the case and where the variability comes from: Is it random, or linked to genetic or environmental factors? Using artificial languages might be profitable to that end: Is individual variability as prevalent when all participants have learned the language in the exact same context and used it for the exact same purpose? A related open issue is how stable this individual variability is, over time (over several repeated sessions) but also across structures. [Tanner & van Hell \(2014\)](#) found a correlation between the RDIs following two types of violations (subject-verb agreement and verb tense, i.e., missing or superfluous *-ing* ending on the main verb), but more studies directly comparing RDIs across different but comparable structures are needed. Variable learner data cannot be fully interpreted without a good understanding of what drives the variability among native speakers.

Another important issue will be to isolate the actual impact of the task on individual variability. Do we observe different processing strategies because of different task-solving strategies, or do native speakers resort to different processing mechanisms in everyday language use? Experiments comparing ERPs to the same structure while completing different tasks, such as acceptability judgments but also priming studies or comprehension questions, should be run to investigate this question. The development of existing technologies also offers new research perspectives – there are now smaller and cheaper EEGs that can be used outside of the lab, to study language in interaction for example. Even though it will be a challenge to obtain data that are controlled enough to do ERP analyses, these new devices will make it possible to study language processing in more ecological settings, which in turn may shed some light on the origins of individual variability in a less task-dependent way. In the meantime, when interpreting learner data, one must keep in mind the possible influence of the specific task on the observed results.

The absence of a clear native-speaker norm means that we cannot clearly interpret data as nativelike or not. Identifying the sources of individual variability among native speakers may allow us to compare more similar populations across native speakers and non-native speakers (e.g., right-handed speakers with no left-handed blood relatives), but that is quite restrictive and we need to go beyond what is eminently nativelike to question what makes processing strategies different at high proficiency. If we cannot clearly determine whether proficient language learners use the same mechanisms as native speakers, we might still be able to investigate whether they use the same range of mechanisms, and whether the same factors affect which processes are recruited and when.

Maud Pélissier

Acknowledgments

The experiment was funded by an Institut Universitaire de France grant awarded to Dr Emmanuel Ferragne.

References

- Alemán Bañón, José, Robert Fiorentino & Alison Gabriele. 2014. Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research* 30(3). 275–306.
- Alemán Bañón, José, David Miller & Jason Rothman. 2017. Morphological variability in second language learners: An examination of electrophysiological and production data. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(10). 1509–1536.
- Astésano, Corine, Mireille Besson & Kai Alter. 2004. Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research* 18(2). 172–184.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Batterink, Laura & Helen Neville. 2013. Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *Journal of Cognitive Neuroscience* 25(6). 936–951.
- Birdsong, David. 2006. Age and second language acquisition and processing: A selective overview. *Language Learning* 56(1). 9–49.
- Bond, Kristi, Alison Gabriele, Robert Fiorentino & José Alemán Bañón. 2011. Individual differences and the role of the L1 in L2 processing: An ERP investigation. In Darren Tanner & Julia Herschensohn (eds.), *Proceedings of the 11th generative approaches to second language acquisition conference (gasla 2011)*, 17–29. Somerville, MA: Cascadia Proceedings Project.
- Bowden, Harriet Wood, Karsten Steinhauer, Cristina Sanz & Michael T. Ullman. 2013. Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia* 51(13). 2492–2511.
- Caffarra, Sendy, Nicola Molinaro, Doug Davidson & Manuel Carreiras. 2015. Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience and Biobehavioral Reviews* 51. 31–47.
- Chen, Lang, Hua Shu, Youyi Liu, Jingjing Zhao & Ping Li. 2007. ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition* 10(2). 161–174.

3 Comparing ERPs between native speakers and second language learners

- Clahsen, Harald & Claudia Felser. 2006. Grammatical processing in language learners. *Applied Psycholinguistics* 27(01). 3–42. DOI: [10 . 1017 / S0142716406060024](https://doi.org/10.1017/S0142716406060024).
- Clahsen, Harald & Claudia Felser. 2018. *Some notes on the Shallow Structure Hypothesis*.
- DeLong, Katherine A., Laura Quante & Marta Kutas. 2014. Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia* 61. 150–162. DOI: [10.1016/j.neuropsychologia.2014.06.016](https://doi.org/10.1016/j.neuropsychologia.2014.06.016).
- DeLong, Katherine A., Thomas P. Urbach & Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8). 1117–1121.
- Fabiani, Monica, Gabriele Gratton & Kara D. Federmeier. 2007. Event-related brain potentials: Methods, theory, and applications. In John Cacioppo, Louis G. Tassinary & Gary G. Berntson (eds.), *Handbook of psychophysiology*, 85–119. Cambridge: Cambridge University Press.
- Faretta-Stutenberg, Mandy & Kara Morgan-Short. 2018. The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research* 34(1). 67–101.
- Federmeier, Kara D. 2007. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44(4). 491–505.
- Federmeier, Kara D. & Marta Kutas. 1999. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language* 41(4). 469–495.
- Foucart, Alice & Cheryl Frenck-Mestre. 2010. Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition* 14(3). 379–399.
- Foucart, Alice & Cheryl Frenck-Mestre. 2012. Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language* 66(1). 226–248.
- Foucart, Alice, Clara D. Martin, Eva M. Moreno & Albert Costa. 2014. Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40(5). 1461–1469.
- Friederici, Angela D. 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6(2). 78–84.
- Friederici, Angela D., Erdmut Pfeifer & Anja Hahne. 1993. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research* 1(3). 183–192.

Maud Pélissier

- Fromont, Lauren A., Phaedra Royle & Karsten Steinhauer. 2012. *On-line processing of French syntactic word categories in native speakers and late L2 learners: An ERP study*. Poster presented at the CUNY Conference on Human Sentence Processing (CUNY 2017). Cambridge, MA.
- Gillon-Dowens, Margaret, Taomei Guo, Jingjing Guo, Horacio Barber & Manuel Carreiras. 2011. Gender and number processing in Chinese learners of Spanish: Evidence from event-related potentials. *Neuropsychologia* 49(7). 1651–1659.
- Gillon-Dowens, Margaret, Marta Vergara, Horacio Barber & Manuel Carreiras. 2010. Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience* 22(8). 1870–1887.
- Grey, Sarah, Darren Tanner & Janet G. van Hell. 2017. How right is left? Hand-edness modulates neural responses during morphosyntactic processing. *Brain Research* 1669. 27–43.
- Gunter, Thomas C., Angela D. Friederici & Herbert Schriefers. 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience* 12(4). 556–568.
- Guo, Jingjing, Taomei Guo, Yan Yan, Nan Jiang & Danling Peng. 2009. ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics* 22(2). 123–134.
- Hahne, Anja, Jutta L. Mueller & Harald Clahsen. 2006. Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience* 18(1). 121–134.
- Hakuta, Kenji, Ellen Bialystok & Edward Wiley. 2003. Critical evidence. *Psychological Science* 14(1). 31–38.
- Isel, Frédéric. 2017. Apport de l'électro-encéphalographie à l'étude des atypies du traitement du langage. In Caroline Bogliotti, Frédéric Isel & Anne Lacheret-Dujour (eds.), *Atypies langagières de l'enfance à l'âge adulte: Apport de la psycholinguistique et des neurosciences cognitives*, 149–192. Paris: De Boeck Supérieur.
- Isel, Frédéric, Anja Hahne, Burkhard Maess & Angela D. Friederici. 2007. Neurodynamics of sentence interpretation: ERP evidence from French. *Biological Psychology* 74(3). 337–346.
- Isel, Frédéric & Michèle Kail. 2018. Morphosyntactic integration in French sentence processing: Event-related brain potentials evidence. *Journal of Neurolinguistics* 46. 23–36.
- Kaan, Edith. 2007. Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass* 1(6). 571–591.

3 Comparing ERPs between native speakers and second language learners

- Kim, Albert E., Leif Oines & Akira Miyake. 2018. Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44(3). 406–420.
- Kotz, Sonja A. 2009. A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language* 109(2–3). 68–74.
- Kotz, Sonja A., Phillip J. Holcomb & Lee Osterhout. 2008. ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologica* 128(3). 514–527.
- Kutas, Marta, Katherine A. DeLong & Nathaniel J. Smith. 2011. A look around at what lies ahead: Prediction and predictability in language processing. In Moshe Bar (ed.), *Predictions in the brain: Using our past to generate a future*, 190–207. Oxford: Oxford University Press.
- Kutas, Marta & Steven A. Hillyard. 1980. Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology* 11(2). 99–116.
- Lenth, Russel. 2019. *emmeans: Estimated marginal means, aka least-square means*. R package version 1.4.1.
- Luck, Steven J. 2014. *An introduction to the event-related potential technique*. 2nd edition. Cambridge, MA: MIT Press.
- Martín-Loeches, Manuel, Francisco Muñoz, Pilar Casado, A. Melcón & C. Fernández-Frías. 2005. Are the anterior negativities to grammatical violations indexing working memory? *Psychophysiology* 42(5). 508–519. DOI: [10.1111/j.1469-8986.2005.00308.x](https://doi.org/10.1111/j.1469-8986.2005.00308.x).
- McLaughlin, Judith, Lee Osterhout & Albert Kim. 2004. Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience* 7. 703–704.
- McLaughlin, Judith, Darren Tanner, Ilona Pitkänen, Cheryl Frenck-Mestre, Kayo Inoue, Geoffrey Valentine & Lee Osterhout. 2010. Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning* 60. 123–150.
- Mehravari, Alison S., Karen Emmorey, Chantel S. Prat, Lindsay Klarman & Lee Osterhout. 2017. Brain-based individual difference measures of reading skill in deaf and hearing adults. *Neuropsychologia* 101. 153–168.
- Meulman, Nienke, Martijn Wieling, Simone A. Sprenger, Laurie A. Stowe & Monika S. Schmid. 2015. Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them. *PloS One* 10(12).
- Molinaro, Nicola, Horacio A. Barber & Manuel Carreiras. 2011. Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47(8). 908–930. DOI: [10.1016/j.cortex.2011.02.019](https://doi.org/10.1016/j.cortex.2011.02.019).

Maud Pélissier

- Molinaro, Nicola, Francesco Vespignani & Remo Job. 2008. A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research* 1228. 161–176.
- Morgan-Short, Kara, Mandy Faretta-Stutenberg & Laura Bartlett-Hsu. 2015. Contributions of event-related potential research to issues in explicit and implicit second language acquisition. In Patrick Rebuschat (ed.), *Implicit and explicit learning of languages*, 349–384. Amsterdam: John Benjamins.
- Morgan-Short, Kara, Cristina Sanz, Karsten Steinhauer & Michael T. Ullman. 2010. Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning* 60(1). 154–193.
- Morgan-Short, Kara, Karsten Steinhauer, Cristina Sanz & Michael T. Ullman. 2012. Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience* 24(4). 933–947.
- Mueller, Jutta L., Masako Hirotsu & Angela D. Friederici. 2007. *ERP evidence for different strategies in the processing of case markers in native speakers and non-native learners*. BMC Neuroscience.
- Nakano, Hiroko, Clifford Saron & Tamara Y. Swaab. 2010. Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience* 22(12). 2886–2898.
- Newman, Aaron J., Antoine Tremblay, Emily S. Nichols, Helen J. Neville & Michael T. Ullman. 2012. The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience* 24(5). 1205–1223.
- Newman, Aaron J., Michael T. Ullman, Roumyana Pancheva, Diane L. Waligura & Helen J. Neville. 2007. An ERP study of regular and irregular English past tense inflection. *NeuroImage* 34(1). 435–445.
- Ojima, Shiro, Hiroki Nakata & Ryusuke Kakigi. 2005. An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience* 17(8). 1212–1228.
- Osterhout, Lee. 1997. On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language* 59(3). 494–522.
- Osterhout, Lee & Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31(6). 785–806.

3 Comparing ERPs between native speakers and second language learners

- Osterhout, Lee, Judith McLaughlin, Albert Kim, Ralf Greenwald & Kayo Inoue. 2004. Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In Manuel Carreiras & Charles Clifton Jr. (eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*, 271–308. New York, NY: Psychology Press.
- Osterhout, Lee, Judy McLaughlin, Ilona Pitkänen, Cheryl Frenck-Mestre & Nicola Molinaro. 2006. Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning* 56. 199–230.
- Pakulak, Eric & Helen J. Neville. 2010. Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience* 22(12). 2728–2744.
- Paradis, Michel. 2009. *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.
- Qi, Zhenghan, Sara D. Beach, Amy S. Finn, Jennifer Minas, Calvin Goetz, Brian Chan & John D. E. Gabrieli. 2017. Native-language N400 and P600 predict dissociable language-learning abilities in adults. *Neuropsychologia* 98. 177–191.
- Roehr-Brackin, Karen. 2015. Explicit knowledge about language in L2 learning: A usage-based perspective. In Patrick Rebuschat (ed.), *Implicit and explicit learning of languages*, 117–138. Amsterdam: John Benjamins.
- Rossi, Sonja, Manfred F. Gugler, Angela D. Friederici & Anja Hahne. 2006. The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience* 18(12). 2030–2048.
- Sassenhagen, Jona, Matthias Schlesewsky & Ina Bornkessel-Schlesewsky. 2014. The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language* 137. 29–39.
- Schneider, Julie M., Alyson D. Abel, Diane A. Ogiela, Anna E. Middleton & Mandy J. Maguire. 2016. Developmental differences in beta and theta power during sentence processing. *Developmental Cognitive Neuroscience* 19. 19–30.
- Steinhauer, Karsten. 2014. Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics* 35(4). 393–417.
- Steinhauer, Karsten, Erin J. White & John E. Drury. 2009. Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research* 25(1). 13–41.
- Tanner, Darren. 2019. Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex* 111. 210–237.

Maud Pélissier

- Tanner, Darren, Kayo Inoue & Lee Osterhout. 2012. *Brain-based individual differences in on-line L2 sentence comprehension*. Manuscript.
- Tanner, Darren, Kayo Inoue & Lee Osterhout. 2014. Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition* 17(2). 277–293.
- Tanner, Darren, Judith Mclaughlin, Julia Herschensohn & Lee Osterhout. 2013. Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition* 16(2). 367–382.
- Tanner, Darren, Lee Osterhout & Julia Herschensohn. 2009. Snapshots of grammaticalization: Differential electrophysiological responses to grammatical anomalies with increasing L2 exposure. In Jane Chandlee, Michelle Franchini, Sandy Lord & Gudrun-Marion Rheiner (eds.), *BUCLD 33: Proceedings of the 33rd Annual Boston University Conference on Language Development*, 528–539. Boston, MA: Cascadilla Press.
- Tanner, Darren & Janet G. van Hell. 2014. ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* 56(1). 289–301.
- Team, R Core. 2019. *R: A language and environment for statistical computing*. (Version 3.6.1.) Vienna.
- Tokowicz, Natasha & Brian MacWhinney. 2005. Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition* 27(2). 173–204.
- Tokowicz, Natasha, Erica B. Michael & Judith F. Kroll. 2004. The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition* 7(3). 255–272.
- van de Meerendonk, Nan, Herman H. J. Kolk, Dorothee J. Chwilla & Constance Th. W. M. Vissers. 2009. Monitoring in language perception. *Language and Linguistics Compass* 3(5). 1211–1224.
- van Hell, Janet G. & Natasha Tokowicz. 2010. Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research* 26(1). 43–74.
- Wampler, Emma K. 2017. *Individual differences in grammatical error processing*. Seattle, WA: University of Washington. (Doctoral dissertation).
- Wampler, Emma K., Judith Mclaughlin & Lee Osterhout. 2014. *How gender, handedness, and L1 processing strategy influence L2 grammatical processing*. Poster presented at the Sixth Annual Society for the Neurobiology of Language Conference (SNL 2014). Amsterdam.
- Weber-Fox, C. & Helen Neville. 1996. Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience* 8. 231–256.

3 Comparing ERPs between native speakers and second language learners:
Dealing with individual variability

- Weiss, Sabine, Horst M. Mueller, Baerbel Schack, Jonathan W. King, Martha Kutas & Peter Rappelsberger. 2005. Increased neuronal communication accompanying sentence comprehension. *International Journal of Psychophysiology* 57(2). 129–141.
- White, Erin Jacquelyn, Fred Genesee & Karsten Steinhauer. 2012. 7(12). PloS One.
- Xue, Jin, Jie Yang, Jie Zhang, Zhenhai Qi, Chen Bai & Yinchun Qiu. 2013. An ERP study on Chinese natives' second language syntactic grammaticalization. *Neuroscience Letters* 534. 258–263.
- Zawiszewski, Adam, Eva Gutiérrez, Beatriz Fernández & Itziar Laka. 2011. Language distance and non-native syntactic processing: Evidence from event-related potentials. *Bilingualism: Language and Cognition* 14(03). 400–411.

Chapter 4

Replication: Measuring the influence of typologically diverse target language properties on input processing at the initial stages of acquisition

Marzena Watorek

Université Paris 8 & UMR-SFL, CNRS

Rebekah Rast

American University of Paris & UMR-SFL, CNRS

Xinyue Cécilia Yu

Inalco - CNRS - EHESS, CRLAO

Pascale Trévisiol

Université Paris 3 & DILTEC EA 2288

Hedi Majdoub

Université Paris 8 & UMR-SFL, CNRS

Qianwen Guan

City University of Hong Kong

Xiaoliang Huang

Beijing Foreign Studies University



Watorek *et al.*

This study applies a “first exposure” approach to second language acquisition, based on data collected from learners’ very first contact with the target language. The VILLA project (*Varieties of Initial Learners in Language Acquisition: Controlled classroom input and elementary forms of linguistic organisation*) (Dimroth *et al.* 2013) has made a significant contribution to the development of methodological tools used to observe initial input processing in the first 14 hours of exposure to a target language (Polish) by native speakers of five different languages (Italian, French, English, German and Dutch). The VILLA project dataset allows for a new type of analysis, which compares, under the same controlled input conditions, the performance of learners with different native languages exposed to the same target language.

With a view to expanding and strengthening the cross-linguistic dimension of second language acquisition research, replications of the VILLA methodology with new source-target language combinations are in the planning stages. This chapter presents the design of three replications in which three separate groups of French native speakers will be exposed, in an instructional setting, to Modern Standard Arabic, Mandarin Chinese or Japanese, all typologically different from Polish. The design of these pilot projects draws on the same organizational principles of the VILLA Polish language course, in that learners will be exposed to an unfamiliar language and their performance in the new target language will be tested at various intervals by means of tasks adapted from the VILLA database.

Two specific challenges have arisen while designing replications of the VILLA project that involve different target languages. The first concerns the target language input learners will receive in that the choice of linguistic paradigms to be presented in the input must allow for comparability across VILLA and its replication studies. The other concerns tasks in Arabic, Chinese and Japanese that must be designed based on the Polish model, while also allowing for comparability across studies. This chapter focuses on reflections about these challenges and decisions made about a variety of methodological issues regarding replications when source-target language combinations differ from the initial study.

1 Introduction

Replications in the field of applied linguistics are gaining support not only because they provide insights into the overall validity of results, but also because they allow us to generalise (or limit) results across populations. While Marsden *et al.* (2018) point out that little is known about replication in second language research, their useful guidelines and recommendations are intended to move the field forward in its practice of replicating, through increased collaboration and

4 Influence of diverse target language properties on input processing

transparency of materials and data. They identify a variety of replication categories based on a review of the literature, from a broad to narrow understanding of what studies might entail or self-report. In sum, they recognize three categories, direct, partial and conceptual:

Direct replications make no intentional change to the initial study and seek to confirm methods, data, and analysis; *partial replications* introduce one principled change to a key variable in the initial study to test generalizability in a clearly pre-defined way; and *conceptual replications* introduce more than one change to one or more significant variables. In all cases, ensure that potential heterogeneity and contextual details are documented as fully as possible (366-367).

This chapter introduces three future studies that will replicate an initial study, the French component of the VILLA project (*Varieties of Initial Learners in Language Acquisition: Controlled classroom input and elementary forms of linguistic organisation*), a first exposure study conducted within a functional framework of second language acquisition (Perdue 1993; Watorek 2004; Dimroth 2013). Native speakers of Dutch, English, French, German and Italian received instruction in Polish, a language unfamiliar to all participants. The project contributed significantly to the development of methodological tools used to observe the initial processing of a target language by native speakers of different source languages.

The three replication studies discussed here will change one variable, that of the target language. Following Marsden et al. (2018), these studies could be considered “partial” replications in that one principled change to a key variable in the initial study is introduced with a view to testing the generalizability of findings of the French learners of Polish in the VILLA project. However, when the variable being changed is the target language, this logically triggers changes to other significant variables, such as the language features under investigation. Given this reality, the studies discussed here may be more accurately categorized as “conceptual” replications in that not all features and variables of the initial study can be maintained in cross-linguistic replications like these.

The objective of the current chapter is to describe the unfolding conceptual replications of the VILLA project’s methodology with a view to making cross-linguistic comparisons with other target languages that present different types of acquisitional problems, namely in the acquisition of nominal morphology.¹ Given that the primary reason for these replications is to further our knowledge of the influence of target language properties on input processing at the initial

¹Reflections on replications of the VILLA project were first presented by Rast et al. (2017) as part of the EuroSLA panel “Consolidating and sustaining a principled replication effort in SLA research”.

Watorek *et al.*

stages of acquisition, we have selected three target languages that differ typologically from the VILLA project's target language Polish and from first language (L1) French, particularly with respect to nominal morphology: Modern Standard Arabic (henceforth Arabic), Mandarin Chinese (henceforth Chinese) and Japanese. For each replication, a separate group of French native speakers will be exposed to one of these target languages. The input script used in the Polish instruction of the VILLA project will need to be replicated in the new target languages, as will the tasks designed to measure learners' proficiency level, performance and language development.

Replicating studies within the VILLA project has revealed two particular challenges, one related to the target language input learners will receive, and the other related to the language tasks designed to measure learner performance and development over time. We present these challenges in the form of questions:

- What linguistic paradigms examined in the target language Polish of the VILLA project can be considered “equivalent” in Arabic, Chinese and Japanese, and how will these be presented in the classroom input?
- What Polish tasks designed for the VILLA project can be adapted for Arabic, Chinese and Japanese and how?

This chapter will begin with a brief overview of first exposure studies and the VILLA project, and will be followed by reflection on methodological issues regarding the replications of an instructed language experiment in different target languages, especially when the languages differ typologically in the properties to be investigated.

2 First exposure studies and the acquisition of inflectional morphology

Research concerned with the role of input in the processing and appropriation of a second language (L2) has gained interest due, in part, to studies conducted within the “usage-based” framework (Tomasello 2003; Ellis 2008). This approach claims a strong influence of the statistical distribution of target language input properties for language acquisition. In addition, scholars such as Flege (2009) and others cited in Piske & Young-Scholten's (2009) *Input Matters* highlight the important role of input in second language acquisition and encourage further investigation, even if controlling input is a complex endeavour. Research that focuses on initial exposure has often relied on artificial languages (e.g. Reber

4 *Influence of diverse target language properties on input processing*

1967; Hulstijn & DeKeyser 1997; Williams 2005) or has been limited to the analysis of participants' performance after only a few minutes of exposure to the input (e.g. Gullberg et al. 2012). A study conducted by Rast (2008), which reports on the first 8 hours of exposure to a new target language, contributed significantly to the development of methodology adopted for the study of input at first contact with a novel language and in the minutes and hours that follow. The VILLA project, designed within the same theoretical and methodological framework, can be viewed as emerging directly from this study (see also Rast 2017 for a follow-up to the 2008 study).

With Polish as the target language, a main focus of the VILLA project was inflectional morphology, primarily nominal morphology. A plethora of research has confirmed the difficulties faced by L2 learners in acquiring inflectional morphology (Bardovi-Harlig 2000; Larsen-Freeman 2010). According to Meisel (1987); Bardovi-Harlig (1992); Klein & Perdue (1997) and Starren (2001), adult learners, for instance, code temporal concepts with lexical items (a semantic domain typically expressed by morphology in richly inflected languages) before acquiring inflectional markers.

Fairly recent studies address the challenge of acquiring inflectional morphology in the early stages of L2 acquisition, in a variety of target languages (e.g. Carroll & Widjaja 2013; Han & Liu 2013; Hinz et al. 2013; Rast et al. 2014). However, in spite of the difficulty in acquiring a new target morphological system, it appears that learners develop very early on – after only a few hours of exposure to the input – a sensitivity to target language morphological forms. Studies of target language Polish (in the VILLA project and its precursors), for instance, have shown that learners manage to judge Polish nominal morphology correctly and produce simple utterances in context using case marking after very limited instruction. Even though the type of task was shown to have an effect on learners' processing (Watorek et al. 2016), these results still show some level of early form-meaning mapping.

Research conducted within the VILLA project has contributed to the debate concerning the relative importance of inflectional morphology in initial learner varieties.² Polish, a highly inflected language with a rich case system, has provided an excellent testing ground to observe learners' processing and acquisition of nominal morphology in particular. The three target languages discussed in this chapter differ from Polish in a variety of ways, and allow for a new examination of the acquisition of other target languages that are highly inflected (e.g. Arabic) and of languages that show little inflection (e.g. Chinese and Japanese).

²For detailed information about the VILLA project, see Dimroth et al. 2013; Rast 2017, and Saturno 2017.

3 The VILLA Project

The aim of the VILLA project is to investigate the absolute first stages of the acquisition of a foreign language by observing what learners do when exposed to language instruction, Polish in this case. The project developed the methodological means to do the following:

- observe the acquisitional process from the moment of first contact with the target language through 14 hours of exposure;
- study learners' processing of the target language input in the domains of perception, comprehension, grammatical analysis and production;
- examine the role of typological features of the learners' L1 on their acquisition of Polish (based on cross-linguistic comparisons), as well as the role of universal principles that are specific to language and communication.

The VILLA database offers a complete documentation of the Polish lessons, language development, and learners' individual profiles (Durand 2019). It follows that the database offers the possibility to examine with precision the instructional sequences (and hence the input "content") relative to learner performance, interactions between learners, and interactions between learners and the instructor.

Ten groups of learners from five European countries – France, Italy, Germany, England and the Netherlands (two groups per country) – attended a beginning level Polish course taught by a native speaker of Polish. A communication-based teaching approach was used in the classroom, with linguistic content introduced relative to the situational context of the lesson through simple dialogues and question/answer sequences in Polish only. The instructor used no other languages in the classroom. The 14-hour course was held over two weeks (9 days of 90-minute sessions, with a final session of 15 minutes on the 10th day before final testing). Central themes of the Polish instruction included introductions, professions, nationalities, languages, cities, countries, tastes and preferences, as well as ordering food and giving directions. In all project countries, the Polish lessons were filmed with two cameras (one focused on the instructor and the other on the learners) and audio-recorded with high quality multi-channel recording equipment (Macbook HD recording with RME Fireface and Presonus preamps), table microphones (Audix) for each learner and a wireless microphone (Dpa) for the instructor. As such, the database includes video and audio recordings of the instructor, as well as interactions and oral productions of individual learners during the lessons. The input content was carefully planned in advance, in particular

4 *Influence of diverse target language properties on input processing*

with respect to the choice of lexical and grammatical items to be taught, and the frequency and transparency of these items.

With respect to frequency, research in both first and second language acquisition has shown its important role in input processing (Slobin 1985; Braine et al. 1990; Rott 1999; Ellis 2002; Gullberg et al. 2010). The effect of frequency, however, is not necessarily immediate (Slobin 1985; Rast 2008). One objective of the VILLA project, which controls the input from the first moments of contact with the new language, is to identify when frequency begins to have a substantial influence on acquisition. Regular tasks administered throughout the data collection period made it possible to test the effect of frequency on learners' processing of Polish in a variety of linguistic domains. To do this, frequency categories were established prior to instruction, and the Polish instructor was asked to use certain words frequently and regularly and to avoid using others. Based on frequency analyses conducted by Goldschneider & DeKeyser (2001), the VILLA project methodology established the category of "frequent" as more than 20 occurrences of a word in the input at the time of testing. Words categorized as "absent" never appear in the classroom input.

Concerning transparency, words that were frequent in the input fit into two categories: transparent or opaque. This classification was based on the results of a transparency test taken by native speakers of the five L1s of the project, who knew no Polish or other Slavic language. They were asked to listen to Polish words and translate them as best they could into their native language. Words that were correctly translated by more than 50% of the participants in each language group were classified as "transparent". Those translated by no members of the group were considered "opaque". These criteria for frequency and transparency provided the basis for the word list established before the Polish course began, with words classified in one of four categories: frequent and transparent; frequent and opaque; absent and transparent; absent and opaque.

A series of Polish tasks were administered to learners before the Polish course began to test their ability to perceive aspects of the new language at absolute first exposure and to serve as a benchmark for language development that took place over the 14-hour period of instruction. Polish tasks were administered regularly throughout the data collection period to gather information about learners' abilities in a variety of language activities and the influence of frequency and transparency, if any. The VILLA project database, thus, includes not only documentation of the input during the lessons, but also documentation of learner performance on language tasks, enabling analyses that compare learners' performance with the input they encountered.

Watorek et al.

The goal of the replication studies discussed in this chapter is to investigate whether the findings of the VILLA project with respect to learners' ability to acquire and make use of the nominal morphology system of Polish can be generalised to other new target languages, namely Arabic, Chinese and Japanese. In the VILLA project, the Polish instructor introduced nominal morphology via the themes of nationalities and professions. These themes will remain consistent in the input of the three replication studies. Following exposure to the novel language, learners will be administered replications of carefully selected VILLA tasks: *Grammaticality Judgment*, *Oral Question-Answer*, and *Picture Verification*. The processes and challenges of replication will be discussed in the following section.

4 Replicating the VILLA project

To extend the cross-linguistic dimension of the VILLA project, the studies presented here propose replications of the VILLA methodology using native speakers of the same L1 (French in this case) learning three typologically different target languages: Arabic, Chinese and Japanese.

As mentioned above, this proposal faces two methodological challenges. The first is related to the input script of the language course, which will need to be designed in Arabic, Chinese and Japanese in the same way that it was created in Polish for the VILLA project. For first exposure replication studies, it will be important to select linguistic paradigms that can be introduced in beginning level language courses, to choose comparable linguistic paradigms to those taught in the VILLA project Polish course, and to maintain the variables of frequency and transparency of the lexical items taught. The objective is not only to control the input over a time period that extends beyond several minutes, but also to provide input (in the form of language courses) that is comparable across target languages with very different features.

The second methodological challenge involves the design of the language tasks in the new target languages relative to the linguistic paradigms taught in the language courses. Ideally, the language tasks would be direct replications of the VILLA tasks. However, this is not possible because the linguistic paradigms needed to communicate the same information in Polish and the other target languages differ. For instance, in Polish, when referring to nationalities and professions, the grammatical subject constrains nominal morphology and the predicate requires certain case marking depending on the context. The predicate requires the nominative case if the subject of the copula 'to be' is the demonstrative pronoun *to* ('this') – this form is generally used to introduce a person (e.g., *to jest*

4 Influence of diverse target language properties on input processing

student, ‘this is a student’). The predicate requires the instrumental case if the subject of the copula is the personal pronoun *on/ona* (‘he/she’) or a proper or common noun – this form is generally used to describe a person who has already been introduced (e.g., *Luc jest studentem*, ‘Luc is a student’). As in English, French nouns do not show case. Hence, this distinction in Polish for native speakers of languages like French presents a difficulty for the acquisition of the new morphological system. The VILLA project assured that evidence for this distinction appeared in the Polish input to learners and tested learners’ ability to make this distinction in comprehension and production. Cross-linguistic replication studies need to find ways to replicate this methodology in other source-target language combinations that do not necessarily have the same or even similar morphological systems. In Arabic, for example, there is no instrumental case, so one possibility for replication is to adapt the input and tasks to use the distinction between the nominative form (e.g., *huwa faransiyy-un*, ‘he is French’) and the genitive form (*al=sayyarat-u li=l-faransiyy-I*, ‘the car belongs to the French man’). In Japanese and Chinese, there is no distinction through case marking. The methodological challenge is to identify what linguistic features of Japanese and Chinese can appear in the input and be tested in such a way as to have access to the processes implicated in learning a new morphological system that would be comparable with the task French speakers face when confronted with Polish morphology.

The methodological challenges mentioned above, not only selecting the linguistic paradigms and creating the input that includes these, but also creating language tasks in Arabic, Chinese and Japanese that replicate the VILLA tasks, directly affect the lesson plan design, which will need to respect the principles of teaching methodology adopted in the VILLA project (a communication-based approach) and the progression of the VILLA classes. Given the cross-linguistic differences between Polish and the three languages of the replication studies, organizing pedagogical sequences in such a way that they can be compared with pedagogical sequences in VILLA also involves a careful choice of linguistic paradigms specific to each language.

4.1 Methodological challenge I: Replicating the language course input

The general organisation of the language courses in the replication studies is the first major challenge. The replications consist of organising beginning-level courses in each of the three languages following a similar schedule to the VILLA project Polish course. When possible, the courses will address the same themes as the VILLA project (e.g. nationalities and professions). Keeping themes in line

Watorek et al.

with the VILLA project guarantees that certain lexical items and linguistic properties studied in the VILLA project will be present in the input and tasks of the replication studies, hence facilitating comparability. For each language, 20 learners will be selected, all French native speakers with no prior knowledge of the target language and with similar profiles to those of the VILLA learners (university students aged 20-25, studying disciplines other than languages, psychology and linguistics).

4.1.1 The choice of linguistic paradigms and cross-linguistic differences

In order to compare results of the replication studies with VILLA project results, it is important to carefully select the linguistic paradigms in Chinese and Japanese, on one hand (languages with little to no inflection), and Arabic on the other (a language with nominal morphology that nonetheless attests important differences from nominal morphology in Polish). Table 1 provides a brief overview of the major relevant differences between these languages.

This cross-linguistic comparison highlights the different features that need to be taken into account when selecting linguistic paradigms for the replication studies. Arabic and Polish, despite their differences, offer the possibility of comparing a similar morpho-syntactic paradigm through the investigation of learners' processing of inflectional marking relative to word order. Chinese and Japanese, despite their differences, have similar characteristics in contrast to Arabic and Polish. For replication purposes, given the absence of case marking in Chinese and Japanese, this feature, studied in detail in the VILLA project, needs to be replaced by other productive and teachable phenomena in the beginning level classes in Chinese and Japanese.

The French speakers of the VILLA project were exposed to Polish, which differs from French in morpho-syntactic features. Polish, a member of the western group of Slavic languages, attests rich nominal morphology (seven cases). Agreement is marked not only between subjects and verbs (systematic marking of person, number, and in the past tense, gender), but also between nouns and adjectives and certain numerals (gender, number and case). The rich morphology is associated with the pragmatic organisation of constituents and with the null subject feature. Polish nominal morphology marks not only gender and number, but also syntactic function, through case marking. Polish is a non-configurational language, that is, it is characterised by a relatively free word order. It is the case markers that signal relations between the different constituents of a clause.

The French speakers of the replication studies will be exposed to one of the three target languages, Arabic, Chinese or Japanese, all of which differ from Pol-

4 Influence of diverse target language properties on input processing

Table 1: Cross-linguistic comparison of the languages of the VILLA project and replication studies

	Target languages				Native lang.
	Polish	Arabic	Chinese	Japanese	French
Language family/-group	Slavic	Semitic	Sino-Tibetan	Japonic	Romance
Verbal morphology	Yes	Yes	No	Yes	Yes
Nominal morphology (case, gender)	7 cases 3 genders	3 cases 2 genders	No	No (but case particles)	No
Word order	Relatively Free	Relatively Free	SVO	(S)OV	SVO
Morphological compound-ing	Yes	Yes	Yes, very rich	Yes, very rich	Yes
Nominal classifiers	No	No	Yes, very rich	Yes, very rich	No
Alphabet	Latin	Arabic	Characters	Characters + Kanas	Latin

ish with respect to inflectional systems. These differences, however, are not of the same nature. Chinese and Japanese have little to no inflectional morphology, while Arabic has systems of nominal and verbal morphology that differ quite radically from Polish.

Designing language courses for the replications based on the VILLA project protocol will likely be easier in Arabic, a Semitic language of the Afro-Asiatic language family. It is an inflectional language that is both agglutinating and fusional, and its case system allows for flexible word order. Nouns are generally marked for gender, number and case by means of suffixes. The plural can also be formed with infixes. Even though Arabic tends towards VSO word order, other configurations are possible (Ryding 2005). During the Arabic instruction, markers of nominal morphology (suffixes) will be integrated immediately into the input to sensitise learners to the markers of gender and number, which are needed to distinguish referents in Arabic. These markers will be necessary for the themes,

Watorek et al.

namely nationalities and professions (as per the VILLA project), to be used in the replications. For example, the suffixes *-at*, *-ūna* and *-āt* can be added to the lexeme *mu'allim* to indicate gender and number: *mu'allim* (teacher-M.SG), *mu'allim-at* (teacher-F.SG), *mu'allim-ūna* (teacher-M.PL), *mu'allim-āt* (teacher-F.PL). The paradigm to be taught and tested in Arabic contrasts the nominative and accusative feminine and masculine forms in relation to three word-orders: VSO, SVO and VOS. As will be seen in section 4.2, adapting VILLA project language tasks to Arabic should not be too difficult given its system of inflectional morphology.

In contrast to the replication in Arabic, preparing the courses in Chinese and Japanese will be a greater challenge. Regarding morpho-syntax, Mandarin Chinese, which belongs to the Sino-Tibetan family, does not resort to inflectional morphology; verbs only combine with a few aspectual markers. There is neither S-V agreement nor tense or case marking. The most significant morphological phenomenon in Chinese is compounding. A compound can be defined as a combination of two or more lexemes, such as 'blackboard' ('black' + 'board') in English and its equivalent in Chinese 黑板 *heiban* (黑 *hei* 'black' + 板 *ban* 'board'). The canonical word order is SVO in Chinese, and in the absence of inflectional morphology, Chinese has a relatively strict word order. Under certain pragmatic conditions, however, such as new/old information or a topic/comment distribution of information, a surface word order such as OSV or SOV is legitimated. Verbal arguments such as subject or object, when implicit in the context, can even be omitted in the surface structure. Classifiers are used in Chinese when expressing quantification.

Japanese, a Japonic language, can be viewed as an inflected language when referring to certain word classes, in particular verbs, adjectives and auxiliaries that carry aspectual-temporal marking. As for Japanese nouns, they are non-inflecting, have no gender or grammaticalised number and take no articles. Japanese word formation involves various types of suffixes, and it also has productive mechanisms of compound formation with native words, Sino-Japanese or a combination of words of different origin (Shibatani 1990). Free and bound morphemes are attested in both derivation and compounding processes. Another feature of morphology is the use of nominal classifiers to express quantified objects. Japanese is classified as a language with SOV canonical word order. Apart from the constraint of the predicate in final position (verbal, adjectival or nominal), Japanese word order is relatively free. The subject and direct object do not have fixed positions in an utterance, and topicalized objects can precede the subject (OSV) when they refer to old or known information. Furthermore, none of the arguments of the predicate is obligatory in a strictly syntactic sense, even the subject,

4 *Influence of diverse target language properties on input processing*

which is the same case in Chinese.

It is worth noting that even though some features are widely attested cross-linguistically, they do not necessarily fit with the VILLA project. This is because the manifestation of a given feature can vary from one language to another so considerably that the term ‘feature’ can only be understood in a functional sense, meaning that structures assuming the same function can be entirely different in languages. This is the case of possessive constructions, which are realized morphologically in languages with declensions like Polish, but fall into the realm of syntax in Chinese. The possessive construction in Chinese takes a ‘A-*de*-B’ form, where *de* is a functional word which links A to B and the ‘possessor-possessed’ is just one of the numerous relations, all syntactic in nature, that can possibly hold between A and B. Due to this fact, it is difficult to establish a correspondence between the possessive construction in Chinese and the construction assuming the same function in Polish, namely the genitive case.

This is a different situation than the case of numeral classifiers in Chinese, for which a correspondence can be defined when compared to French. Both Chinese and French categorize nouns via semantic features, which are manifested by the selection of classifiers in the former and by gender marking in the latter. This idea has been advanced on different grounds, for example by Aikhenvald (2000) from a typological perspective and by Picallo (2008) (cited in Rouveret 2016) from the perspective of formal syntax. At the same time, classifiers in Chinese involve number features (see Cheng & Sybesma 1998 and Li 1999 among others), which are also present in French. All in all, it is fair to say that in both Chinese and French the same sets of features come into play syntactically in the nominal domain.

Following this line of thought, two linguistic paradigms in Chinese and Japanese were identified as the focus of the replication studies: morphological compounding and nominal classifiers. Similar to the VILLA project’s examination of nominal morphology, the Chinese and Japanese replications will examine how learners go about analysing the internal structure of compound words composed of different morphemes in Chinese or Japanese. As mentioned above, language acquisition research has shown L2 inflectional morphology to be particularly difficult to acquire. Replications in Chinese and Japanese address the question of whether morphological features in non-inflected languages pose the same type of acquisitional challenge.

Due to the (quasi) absence of inflectional morphology in target languages Chinese and Japanese, unlike in Polish of the VILLA project, the focus of these studies will be the morphological awareness of compounding. Compounding consists

Watorek et al.

of combining two or more morphemes to produce a new lexical unit that functions as one word. In Chinese and Japanese, most compounds have the internal structure ‘modifier + modified’. The modified morpheme is the head of the compound word, and the modifier semantically modifies it.

As for morphological awareness, it refers to the ability to reflect upon and manipulate morphemes and the morphological structure of words (Carlisle 2003). These morphological structures include inflection and derivational morphology, as well as compounding. Whereas sensitivity to (or awareness of) inflectional morphology was the focus of the VILLA project, the focus in the Chinese and Japanese replications will be awareness of morphological compounding. Hence, an important question for the Chinese and Japanese replications is whether morphological awareness has an impact on vocabulary knowledge and vice versa. Given that many previous studies (Ku & Anderson 2003; Zhang & Koda 2014; Ichikawa 2014; Zhang et al. 2016, among others) have shown a positive correlation between morphological awareness and vocabulary knowledge, we would also like to know whether this correlation starts already at the very beginning stages of the acquisition of a novel language.

Both Chinese and Japanese differ from French and Polish in that they are “classifier languages” (see details in 4.2.5.2.2). Classifiers can be used to quantify both countable and uncountable objects. Previous studies on the acquisition of classifiers in L1 and/or L2 Chinese (Liang 2008; Gong 2010; Kong 2012, among others) reveal different processes observed in L1 children and L2 adults with respect to the acquisition of classifiers due to cognitive knowledge difference and cross-linguistic influence. However, to our best knowledge, none of these studies have taken input into consideration (to explain overgeneralisation, for example). The question, therefore, is whether the processing of morphological compounding and nominal classifiers represents the same cognitive cost and degree of difficulty as nominal morphology does when processing languages like Polish.³

4.1.2 The variables frequency and transparency in the replication studies

In an effort to keep true to the VILLA project methodology, the frequency and transparency of lexical items in the input will need to be established before the organisation of language courses and be taken into account by the instructor during the lessons. This type of replication should be able to contribute to our understanding of whether frequency effects can be established independent of

³It should be noted that certain proposed tasks will be programmed using E-prime software to record reaction time, which is one way to measure the cognitive cost involved when performing such tasks.

4 *Influence of diverse target language properties on input processing*

target language and to what extent transparency plays a role in the processing of different target languages. Typological differences between target languages of the study, however, lead to new challenges, especially in the case of transparency. The semantic transparency of an item as it was operationalised in the VILLA project depends entirely on the relation between prior known languages and the target language, based on results of a transparency test, as described in section 3 above.

With this observation in mind, the research teams of the three future replication studies discussed in this chapter proceeded to conduct transparency tests following the VILLA project methodology. As such, the same type of transparency test was administered to three distinct groups of native speakers of French who were unfamiliar with Arabic, Chinese and Japanese. The participants listened to words in the respective target language, 76 in total, and were asked to translate them as best they could into French. As in the VILLA project, words were chosen relative to the themes of instruction (nationalities, professions, etc.). An item was considered “transparent” when it was translated correctly by at least 50% of the participants. Note that this methodological challenge, which seeks transparency across speakers of the same L1 learning different target languages, differs from that of the VILLA project, in which transparency was sought across speakers of different L1s learning the same target language (see Section 3). In both cases, the objective is to establish a list of transparent items across language learner groups, with various source-target language combinations for the purpose of comparability.

The results of the transparency test demonstrate the difficulty in establishing a list of common transparent items across the target languages of the replication studies. Only 3 items were transparent across the three target languages for native speakers of French (the words ‘Italian’, ‘Marseille’ and ‘Mauritania’). Analyses of the data for each target language show 27 transparent items between French and Arabic, 18 between French and Japanese, and 6 between French and Chinese. Of these items, 7 are shared between Arabic and Japanese, 3 between Japanese and Chinese and only one item between Arabic and Chinese.

These results are not surprising. Chinese has few items borrowed from Indo-European languages, and those that are borrowed are adapted to the phonological system of Chinese, which makes them difficult for French speakers to perceive. It is also clear from these results that transparency indeed varies according to the source and target languages in question. Given these results, maintaining transparency as a variable across all three replication studies reported here is not possible. Transparency will be maintained, however, for Japanese and Arabic, although different transparent and opaque items will be selected for the two

Watorek et al.

languages given the lack of overlap in the results of the transparency test.

4.2 Methodological challenge II: Replicating the language tasks

Designing comparable tasks to those of the initial study is also a challenge. This section describes the tasks that have been adapted from the VILLA project to examine the processing of the selected linguistic paradigms (described above) during the observation period. The studies described here intend to replicate two or three of the VILLA tasks: *Grammaticality Judgement*, *Picture Verification*, and *Oral Question-Answer*. The first two receptive tasks test the processing of a given linguistic paradigm, whereas the third task, a focused production exercise, tests whether a learner can make use of the linguistic paradigm in oral production.

4.2.1 Task: Grammaticality Judgement

In the VILLA project, the acquisition of different properties of Polish inflectional morphology was investigated by means of a series of experiments that were repeated at various time intervals. As mentioned above, one primary focus was the acquisition of case marking. The Nominative vs. Instrumental contrast was captured by a reaction-timed grammaticality judgement task. In this experiment, participants heard two types of Polish copula constructions involving either a correct construction with one noun marked for nominative and the other for instrumental, or an incorrect construction with a double nominative in which the noun in the predicate was used in the incorrect context. The learners were asked to indicate whether they thought the sentence was correct or not. Examples in (1) and (2) are taken from the VILLA grammaticality judgement task. Target items were evenly distributed with respect to transparency (transparent/opaque) and frequency (frequent/absent) in the input.

- (1) Correct sentences
 - a. Transparent items:
 - i. Albert jest fotograf-**em**
Albert is photographer.M-INST
 - ii. Helena jest studentk-**a**
Helena is student.F-INST
 - b. Opaque items:
 - i. Patryk jest lekarz-**em**
Patryk is doctor.M-INST

4 *Influence of diverse target language properties on input processing*

- ii. Eliza jest Włoszk-a
Eliza is Italian.F-INST

(2) Incorrect sentences

- a. Transparent items:
 - i. *Tomasz jest fotograf-Ø
Tomasz is photographer.M-NOM
 - ii. *Krystyna jest studentk-a
Krystyna is student.F-NOM
- b. Opaque items:
 - i. *Dawid jest lekarz-Ø
David is doctor.M-NOM
 - ii. *Sandra jest Włoszk-a
Sandra is Italian.F-NOM

The task was administered after 4.5 hours and again after 10.5 hours of Polish instruction, such that data from the judgement and production (oral question-answer) experiments could be compared.

A grammaticality judgement task is generally used to test learners' intuitions about the grammatical acceptability of a decontextualised sentence, be it oral or written. This task is difficult to replicate in Chinese and Japanese with the two paradigms chosen – morphological compounding and nominal classifiers – because both require the involvement of a semantic component. In these paradigms, grammaticality alone cannot be judged. Sentences that contain morphological components or classifiers must be presented in context or linked to images so that participants can comment on the acceptability of the sentence, which would be a “semantic”, not “grammatical” acceptability. For this reason, the grammaticality judgement task will be replicated in Arabic only.

4.2.2 Task: Picture Verification

The picture verification task was designed primarily to discover precisely what type of grammar rules learners developed for the assignment and interpretation of argument roles (subject vs. direct object) with respect to the Polish input. The experiment comprised transitive SVO and OVS sentences in order to tease apart the relative role of word order and morphological case marking (nominative vs. accusative). In this task, participants listened to pre-recorded Polish transitive sentences in either SVO, OVS or OSV word order (e.g. ‘The brother calls the sister’). The following examples illustrate the sentence types:

Watorek et al.

- (3) a. Brat-Ø woła siostrę (SVO)
 brother.SG-NOM calls.PRS.3SG sister.SG.ACC
 ‘Brother calls sister.’
- b. siostrę woła brat-Ø (OVS)
 sister.SG.ACC calls.PRS.3SG brother.SG-NOM
 ‘Brother calls sister.’
- c. siostrę brat-Ø woła (OSV)
 sister.SG.ACC brother.SG-NOM calls.PRS.3SG
 ‘Brother calls sister.’

Each sentence was accompanied by two pictures depicting the two protagonists involved in the action. One picture showed the event with the agent and patient roles as stated in the sentence they heard; in the other picture the agent and patient roles were switched. The task was designed in this way in order to tap learners’ preferred interpretations of the Polish sentences, in particular to observe whether they relied mainly on word order, or rather on morphological case marking when trying to figure out the meaning of a sentence. This task was run after 9 hours and again after 13.5 hours of Polish instruction.

In the Arabic replication, this task will be administered as per the VILLA project protocol. In Chinese and Japanese, picture verification tasks in each language will be adapted for the paradigm to be tested, morphological compounding in one task and nominal classifiers in the other. The objective is to observe the influence of length of input exposure and the frequency of items in the input on learners’ capacity to process complex compounding structures. With respect to transparency, only in Japanese will the influence of transparent items on this processing be examined. As noted in section 4.1.2, the previously conducted transparency experiments reveal that we cannot maintain this variable in the construction of lessons and tasks across all three target languages.

4.2.3 Task: Oral Question-Answer

In an oral question-answer task, learners saw a picture of a man or a woman on a screen and heard a pre-recorded copula question in Polish asking for the person’s profession or nationality. The questions came in one of two formats, requiring a noun phrase with either the nominative or the instrumental case in the answer. After the question, a picture symbolizing a profession or a nationality appeared on the screen. The learners’ task was to answer the question with a simple affirmative copula sentence, stating the person’s profession or nationality. Examples taken from the VILLA oral question-answer task are provided in (4):

4 *Influence of diverse target language properties on input processing*

- (4) a. Question: kto to jest?
who is this?
- b. Expected answer: to jest student-Ø
this is student.M-NOM
- c. Question: kto to jest?
who is this?
- d. Expected answer: to jest studentk-a
this is student.F-NOM
- e. Question: kim on jest?
what is he/she?
- f. Expected answer: on jest student-em
he is student.M-INST
- g. Question: kim on jest?
what is he/she?
- h. Expected answer: ona jest studentk-a
she is student.F-INST

Items elicited in the oral question-answer task were evenly distributed for transparency and frequency in the input. This task was administered after 4.5 hours and again after 10.5 hours of Polish instruction, such that data from the grammaticality judgement and production experiments concerning the same target language properties could be directly compared.

In Arabic, the oral question-answer task will be administered as per the VILLA project protocol. In Chinese and Japanese, two versions of this task will be designed, to test both morphological compounding and nominal classifiers.

The following sections will provide descriptions of the task designs in progress:

- Grammaticality judgement in Arabic (one task);
- Picture verification in Arabic (one task), in Chinese (two tasks, one for morphological compounding and one for nominal classifiers), in Japanese (also two tasks, as in Chinese);
- Oral question-answer in Arabic (one task), in Chinese (two tasks, one for morphological compounding and one for nominal classifiers), and in Japanese (also two tasks, as in Chinese).

Watorek et al.

4.2.4 Grammaticality Judgement task in Arabic

As a reminder, the Polish grammaticality judgement task of the VILLA project focuses on the Nominative/Instrumental opposition in two types of utterance structures used in introducing or identifying a person and conveying information about profession or nationality. This paradigm can be easily used to design appropriate pedagogical units for beginner foreign language courses, and also lends itself to transparent items. Nouns that designate professions and nationalities in Polish, for example, tended to be transparent with the L1s of the VILLA project.

In the Arabic replication, in order to maintain comparability with the transparency variable in the VILLA project, a similar structure in Arabic was selected, one that uses the nominative case in the speech act of introducing someone. The target lexical items, however, correspond to nationalities only and not to professions because the latter are not transparent between Arabic and French. Also, contrary to Polish, in Arabic, a person cannot be introduced by means of a structure that requires a different case. For this reason, the task will test the opposition between the nominative case and some other case, which will appear in the incorrect sentences only. The genitive case, used to express belonging, was chosen to contrast with sentences comprising a nominative form, and will be used in the oral question-answer production task in Arabic as well. Both tasks will be designed in such a way to be able to compare learners' oral productions and judgements of similar structures.

More specifically, the linguistic paradigm tested in Arabic will be the opposition between the masculine and feminine singular forms of the nominative and genitive cases. The target items will use the four categories of the VILLA project with respect to frequency and transparency, namely: frequent and transparent; frequent and opaque; absent and transparent; absent and opaque. The correct sentences of the grammaticality judgement task will contain lexical items that refer to nationality and require the nominative case.

The examples in (5a) and (5b) illustrate this type of sentence, which, in Arabic, serves as a copula without the verb "to be". The nominative case marker corresponds to a morpheme *-un* (*-u* marks nominative case and *-n* marks the indefinite) and does not indicate gender. Gender is marked in the adjectives of nationality by the addition of *-iyy-* (masculine) or *-iyyat-* (feminine) to the root. Contrary to Polish, gender is not integrated into Arabic case markers (Kouloughli 2007).

- (5) a. Transparent items:

4 *Influence of diverse target language properties on input processing*

- i. Jean faransiyy-**un**
Jean French.M-**NOM**
'Jean is French.'
- ii. Marie norwiyyat-**un**
Marie Norwegian.F-**NOM**
'Marie is Norwegian.'
- b. Opaque items:
 - i. Charles yunaniyy-**un**
Charles Greek.M-**NOM**
'Charles is Greek.'
 - ii. Jessie nesmawiyyat-**un**
Jessie Austrian.F-**NOM**
'Jessie is Austrian.'

Incorrect sentences of the grammaticality judgement task, as in (6a) and (6b), will be modelled after the correct sentences, but they will contain lexical items with the genitive marker *-in* (*-i* marks genitive case and *-n* marks the indefinite) in place of the correct nominative marker. Both the genitive and nominative markers will appear in the input to learners.

- (6) a. Transparent items:
- i. *Jean faransiyy-**in**
Jean French.M-**GEN**
'Jean is French.'
 - ii. *Marie norwiyyat-**in**
Marie Norwegian.F-**GEN**
'Marie is Norwegian.'
- b. Opaque items:
- i. *Charles yunaniyy-**in**
Charles Greek.M-**GEN**
'Charles is Greek.'
 - ii. *Jessie nesmawiyyat-**in**
Jessie Austrian.F-**GEN**
'Jessie is Austrian.'

Watorek et al.

Following the VILLA project task construction model, including the same number of stimuli, this task in Arabic will comprise 64 test sentences, 32 of which will be correct (nominative) and 32 incorrect (genitive), as well as 32 distracter sentences.

4.2.5 Picture Verification task in Arabic, Chinese and Japanese

The Arabic replication of the picture verification task coincides well with the task created for the VILLA project. Given the morpho-syntactic similarities between Arabic and Polish, the task can be used to test learners' ability to comprehend a sentence by relying on nominal morphology in both Polish and Arabic. As discussed above, this is not the case with Chinese and Japanese given the absence of nominal morphology in these two target languages. For these replication studies, the tasks will be adapted to test the comprehension of compound nouns and nominal classifiers.

4.2.5.1 Arabic

Methodologically speaking, this new picture verification task in Arabic will be identical to that of the VILLA project in that learners see two pictures of two people involved in some sort of action. In one of the pictures, person A is the agent of the action and person B is the patient. In the other, the roles are reversed. The learners hear a sentence describing one of the pictures and are asked to identify which picture corresponds to the sentence they heard. Two verbs that were used frequently in the VILLA Polish instruction, 'kiss' and 'teach', will be maintained for Arabic. The three sentences in (7), for example, show the three possible word orders in Arabic. If learners rely on their native French canonical word order, SVO, when processing Arabic, they will need to learn the relevant morphological markers to be able to accurately identify who is doing the action. In the following examples, for instance, they will have to understand whether the Italians or the French are doing the teaching. Learners' ability to perceive and comprehend the nominative and accusative markers (i.e. their 'morphological awareness') will thus be tested in the three different word order conditions.

- (7) a. al=itāliyy-ūna yu'allimūna al=faransiyy-īna (SVO)
 DEF=Italian.PL-NOM teach.PRS.3PL DEF=French.PL-ACC
 'Italians teach French.'
- b. yu'allimu al=itāliyy-ūna al=faransiyy-īna (VSO)
 teach.PRS.3SG DEF=Italian.PL-NOM DEF=French.PL-ACC
 'Italians teach French.'

4 Influence of diverse target language properties on input processing

- c. *yu'allimu al=faransiyy-īna al=itāliyy-ūna* (VOS)⁴
 teach.PRS.3SG DEF=French.PL-acc DEF=Italian.PL-NOM
 'Italians teach French.'

The task, programmed in OpenSesame to capture accuracy and reaction time, will be administered following the same schedule as the VILLA project. It will be interesting to observe input properties, such as frequency and transparency, as well as the role played by L1 (French) in the learning of a different target language (Arabic), after such limited input, and compare these findings with those of the VILLA project.

4.2.5.2 Chinese and Japanese

The picture verification task will be used to test the acquisition process of two linguistic paradigms in Chinese and Japanese, morphological compounding and nominal classifiers, both of which will be introduced to learners during the Chinese and Japanese language instruction.

4.2.5.2.1 Morphological compounding

Morphemes in Chinese can be roughly divided into four categories according to whether they are free or bound, lexical or functional (Packard 2000):

- function word (+free, +functional): *le* 了 (aspectual marker)
- root word (+free, +lexical): *shu* 书 (book)
- bound root (+bound, +lexical): *guo* 国 (country)
- affix (+bound, +functional): *-zhe* 者 (someone who)

It is important to note that a root word can either appear independently as a word or be combined with another morpheme(s) to form a compound word. On the contrary, a bound root always appears in a compound word.

As with the Polish instruction in the VILLA project and the Arabic instruction in the replication described above, two themes for the Chinese and Japanese

⁴Note that the canonical VSO word order of Arabic has an impact on subject-verb agreement. When the verb precedes the subject, the latter only agrees in gender, not in number. When the subject precedes the verb, it agrees in gender and number. This should not influence the results of this task because the focus here is on nominal morphology, namely the distinction between nominative and accusative forms, not on verbal morphology.

Watorek et al.

instruction will be nationalities and professions. Nouns with the internal structure ‘modifier + modified’ will be taught from the very beginning of instruction. The modified morpheme is the head of the compound word, and the modifier semantically modifies it. For example, in Chinese, the compound noun *faguoren* (French people, as in nationality) is composed of the modifier *faguo* (France), which modifies the head morpheme *-ren* (person).

(8) Examples in Chinese:

- a. *faguo-ren*
France-person
‘French’ (nationality)
- b. *xibanya-ren*
Spain-person
‘Spanish’ (nationality)

In Japanese, in a similar manner, a free morpheme, such as *furansu* (France) serves as a determiner to the bound morpheme *-jin* (person) by semantically modifying it to refer to someone’s nationality.

(9) Examples in Japanese:

- a. *furansu-jin*
France-person
‘French’ (nationality)
- b. *supein-jin*
Spain-person
‘Spanish’ (nationality)

This compounding applies regardless of the lexical origin of the modifier morpheme. In this way, the suffixes *-go* (language) and *-jin* (person) can link to lexical morphemes of Indo-European origin, as in *furansu* (France) and *supein* (Spain), as well as to morphemes of Sino-Japanese origin like *chuugoku* (China) or *kankoku* (South Korea). In both Chinese and Japanese, there is no modification of the morpheme itself, regardless of person, gender and/or number of the referent.

For the Chinese language instruction, five head morphemes of root words and five head morphemes of bound roots will be chosen. The two sub-categories of head morphemes (root words and bound roots) will occur differently in the input; the root words will appear both as independent words and in compound nouns, whereas bound roots will always be ‘bound’ to other morphemes (free or bound)

4 Influence of diverse target language properties on input processing

to function as a word. In other words, the bound root will always appear in a compound noun.

In the following example, *ren* appears as an independent word:

- (10) Jiaoshi li you ji ge ren?
 classroom in have how.many CL person
 ‘How many people are there in the classroom?’

The root words selected for the input in Chinese are hyperonyms to the compounds containing them. In the compound noun *faguoren* ‘French (person)’, the modifier *faguo*, the literal meaning of which is ‘France’, modifies the head morpheme *ren*, which means ‘person’. Thus, *faguoren* (French (person)) is a kind of *ren* (person).

As for Japanese language instruction, it is not possible to proceed exactly as in Chinese because we cannot count on a systematic alternation between “root words” and “bound roots”.

An example in the input for Japanese will be: 花屋 hana-ya (= flower-shop/florist)

In the compound used to express the place *hanaya* (flower shop) and by extension, the profession (florist), the head morpheme *ya* 屋 (roof, house, shop) will change to the word *mise* 店 when used to refer to a ‘shop’ in a free morpheme as in the example:

- (11) この店は花屋です
 kono **mise** wa hanaya desu
 this shop TOP flower.shop AUX
 ‘This shop is a flower shop.’

This is why in the Japanese replication, only words with “bound roots” will be tested because the corresponding morphemes attest different forms when they are “free”.

A picture verification task will be administered twice during the period of exposure, as per the VILLA project protocol and in line with the Arabic replication. The methodology of this task will differ somewhat from the Polish and Arabic task design (in which learners see two pictures, hear a sentence and are asked to select the picture that corresponds to the sentence they heard). In the Chinese and Japanese versions of the task, learners will see a picture and hear a word, and they will be asked to verify whether the item they hear corresponds to the picture by responding ‘yes’ or ‘no’. If learners manage to judge words correctly

Watorek et al.

that have been taught (frequent words in the classroom input), but not those that are not taught (absent from the input), this would suggest that they are basing their knowledge on vocabulary. On the other hand, if they correctly judge new words, those that are absent from the input, in the same way they judge words that are frequent in the input, this would suggest that they base their knowledge on both vocabulary and morphology.

The experimental items will contain the ‘modifier-modified’ structure. By manipulating the frequency and transparency, the four conditions of the VILLA project are obtained for Japanese: frequent and transparent; absent and transparent; frequent and opaque; absent and opaque. In Chinese, however, only frequency (frequent or absent) will be maintained for reasons explained above (cf. 4.1.2). The target items will be composed of head morphemes in Japanese: *-jin* (person/nationality), *-ka* (expert), *-ya* (house/store) and *-go* (language). In Chinese, the categories of head morphemes will be included in the experiment: bound roots like *-yu* (language), *-jia* (expert), *-guan* (establishment), *-ji* (machine) and root words like *ren* (person/nationality), *-dian* (shop), *che* (vehicle), and *piao* (ticket).

To illustrate, let us take a look at some sample stimuli in both Chinese and Japanese.

In Chinese, the learners will see a picture of a singer and hear a word. They may hear *gechang-jia* (sing-expert = singer). The modifier *gechang* (sing) will have been frequent in the input and has a semantically relevant modified head. In this case, the correct response is ‘yes’. Or they may hear **gechang-ren* ((sing-person)). The modifier is frequent in the input and has a semantically irrelevant modified head, because *-ren* (which means ‘person’ and signifies ‘nationality’ in a compound) cannot link to the lexical item *gechang* (sing). In this case, the correct response is ‘no’.

In a similar fashion, in Japanese, the learners will see a picture of a person speaking French and hear a word, such as *furansu-go* (France-language = French). The modifier is transparent and frequent in the input. The modified head is semantically relevant. In this case, the correct response is ‘yes’. Or they will hear **furansu-ka* (France-expert). The modifier is both transparent and frequent in the input but has a semantically irrelevant modified head, because *-ka* (which signifies an expert) cannot link to the lexical item *furansu* (France). In this case, the correct response is ‘no’.

In the VILLA project, the challenge of the picture verification task was to perceive and comprehend the nominal morphology in order to understand which person was performing the action. The challenge in this task for French learners will be to perceive and comprehend the morphemes, mapping form to meaning, as well as to understand the modifier-modified relation, which is in the reverse

4 Influence of diverse target language properties on input processing

order compared to French (e.g. *langue française* (language French) in French vs. *France-langue* (France-language) in Japanese and Chinese.

4.2.5.2.2 Nominal classifiers

During the first stages of acquisition of target language Chinese or Japanese, a challenge for the French learner is expressing quantified objects. The potential difficulty of this resides in the fact that Chinese and Japanese are “classifier languages”, both having non-individual classifiers and individual classifiers.

Non-individual classifiers are independent nouns in Chinese (12a) and Japanese (12b), as in French, and are used to count mass objects.

- (12) a. Example in Chinese:
 san gongjin pingguo
 three kilo apple
 ‘three kilos of apples’
 b. Example in Japanese:
 ringo san kiro
 apple three kilo
 ‘three kilos of apples’

An important difference with respect to French is the use of so-called “individual classifiers” (CL), namely “measure words” that combine with countable nouns. These nouns are normally preceded by a numeral in languages like French and English. In both Chinese and Japanese, individual classifiers combine with countable nouns according to a semantic feature matching (shape, animacy, function among others) (Nishio 2000; Zhang 2007).

For example, in Chinese, *tiao* marks a long shape and flexible feature as in:

- (13) san tiao shengzi
 three CL rope
 ‘three ropes’

In another Chinese example, the element *zhang* marks a flat surface and thin feature as in:

- (14) san zhang zhaopian
 three CL pictures
 ‘three pictures’

Watorek et al.

Japanese classifiers are quite similar to those in Chinese except for the grammatical nature of the individual classifier, which is an affix, and the syntax: “Noun-Numeral-Classifier” in Japanese; “Numeral-Classifier-Noun” in Chinese.

The following are examples in Japanese:

- (15) a. bin san-**bon**
 bottle three-CL.LONGOBJECT
 ‘three bottles’
 b. shashin san-**mai**
 picture three-CL.FLATOBJECT
 ‘three pictures’

In both Chinese and Japanese, the use of a numeral requires an individual classifier in quantification by counting. Hence, the absence of the individual classifier will lead to ungrammaticality in Chinese just as it does in Japanese.

With respect to the acquisition of classifiers in Chinese and Japanese, in our input provided to learners, we have strictly selected vocabulary controlled for frequency with two types of classifiers: non-individual and individual. Keeping with the themes proposed in the VILLA project, in addition to nationalities and professions, this study will use the theme of talking about and ordering food to introduce quantified food and drinks in a social event such as a picnic or party. Noun phrases instantiated in the Numeral-Classifier-Noun structure in Chinese and in the Noun-Numeral-Classifier in Japanese will be presented during instruction in sentences illustrated with pictures to facilitate comprehension.

Thus, another picture verification task, measuring accuracy and reaction time, will be administered twice as well, following the VILLA protocol. In this task, learners will be asked to judge whether the acoustic stimulus corresponds to the picture. More specifically, learners will see a picture and hear a noun phrase and be asked to verify if the noun phrase corresponds to the picture. By manipulating the frequency of nouns (frequent or absent in the input), syntactic grammaticality (classifier) and semantic relevance (classifier type), six conditions are obtained: frequent noun and syntactically ungrammatical classifier; absent noun and syntactically ungrammatical classifier; frequent noun and semantically relevant classifier; absent noun and semantically relevant classifier; frequent noun and semantically irrelevant classifier; absent noun and semantically irrelevant classifier.

To illustrate the methodology, six experimental conditions are provided. In the first three conditions with a frequent item, either in Chinese or Japanese, learners will see a picture of three pictures (flat objects) and hear one of the structures

4 Influence of diverse target language properties on input processing

below. Only one of the structures is correct (16b). Of the two incorrect structures, one has no classifier, and the other has the wrong classifier. The meaning in examples (16a-c) is ‘three pictures’.

- (16) a. Numeral + Ø (syntactically ungrammatical) + frequent noun =>
 **san-Ø-zhaopian* (Chinese) / frequent noun + Numeral + Ø =>
 **shashin san-Ø* (Japanese)
- b. Numeral + semantically relevant CL + frequent noun =>
 san-zhang-zhaopian (three-CL-picture) (Chinese) / frequent noun +
 Numeral + CL => *shashin san-mai* (picture three-CL) (Japanese)
- c. Numeral + semantically irrelevant CL + frequent noun =>
 **san-tiao-zhaopian* (Chinese) / frequent noun + Numeral + CL =>
 **shashin san-bon* (Japanese)

In the other three conditions with an absent item, in either Chinese or Japanese, learners will see a picture of three pancakes (also flat objects). Again, only one of the structures is correct (17b) and the meaning in examples (17a-c) is ‘three pancakes’.

- (17) a. Numeral + Ø (syntactically ungrammatical) + absent noun =>
 **san-Ø-jianbing* (Chinese) / absent noun + Numeral + Ø => **pankeeki*
 san-Ø (Japanese)
- b. Numeral + semantically relevant CL + absent noun =>
 san-zhang-jianbing (three-CL-pancake) (Chinese) / absent noun +
 Numeral + CL => *pankeeki san-mai* (pancake-three-CL) (Japanese)
- c. Numeral + semantically irrelevant CL + absent noun =>
 **san-tiao-jianbing* (Chinese) / absent noun + Numeral + CL =>
 **pankeeki san-bon* (Japanese)

The challenge of making form-meaning connections required in this picture verification task is potentially similar to that faced by L1 French speakers learning Polish or Arabic in that learners have to adapt to a new system which is very different from their L1 in terms of how they express relations between verb arguments, on the one hand, and notions like quantification, on the other. In this latter case, in Japanese, as in Chinese, the French learner has to create a new category (the classifier) while paying attention to semantic features.

Watorek et al.

4.2.6 Oral Question-Answer task in Arabic, Chinese and Japanese

4.2.6.1 Arabic

As in the VILLA project, the replications of the grammaticality judgement and the oral question-answer tasks in Arabic will test the Nominative/Genitive opposition. Using the same testing paradigm, data from judgements and productions can be compared.

The oral question-answer task in Arabic will elicit utterances comprising either the nominative or genitive forms, keeping the same four frequency and transparency categories as in the grammaticality judgement task, namely frequent and transparent, frequent and opaque, absent and transparent, absent and opaque. The target nouns will correspond to nationalities to remain consistent with target items in the grammaticality judgement task. Two selected contexts are expected to elicit these two inflections. On one hand, question (18a) is meant to elicit an utterance with the nominative form: “who is he? / *man huwa?*” or “who is she? / *man hiya?*”. On the other hand, question (18b) “to whom does this object belong? / *li=man hāza al=chay?*” should elicit a response with the genitive form.

Learners will see a series of 32 images, half of which will contain two icons, one referring to a specific gender and the other to a nationality. For half of the images they will hear question (18a), to which the expected response will contain the nominative form.

- (18) a. i. Question:
man huwa?
Who he
‘Who is he?’
ii. Expected answer:
huwa faransiyy-un
he French.3SG.M-NOM
‘He is French.’
b. i. Question:
min hiya?
who she
‘Who is she?’
ii. Expected answer:

4 *Influence of diverse target language properties on input processing*

hiya faransiyyat-**un**
 she French.3SG.F-**NOM**
 ‘She is French.’

For the other half of the images, participants will hear question (19), eliciting the genitive form. These images show, for example, a car (the object of belonging) and an icon that represents the gender (male or female) of the car owner and other symbols that reveal the car owner’s nationality.

- (19) a. i. Question:
 li=man as=sayyarat-u?
 PREP=INT DEF=car-NOM
 ‘Whose is the car?’
 ii. Expected answer:
 al=sayyarat-u li=l-faransiyyat-i
 DEF=car-NOM PREP=DEF-French.F-GEN
 ‘The car belongs to the French woman.’
 b. i. Question:
 li=man as=sayyarat-u?
 PREP=INT DEF=car-NOM
 ‘Whose is the car?’
 ii. Expected answer:
 al=sayyarat-u li=l-faransiyy-i
 DEF=car-NOM PREP=DEF-French.M-GEN
 ‘The car belongs to the French man.’

In sum, the 8 target items will be balanced with respect to the two independent variables, frequency and transparency, resulting in four distributions: frequent and transparent; frequent and opaque; absent and transparent; absent and opaque. The items will also be classified according to case (nominative or genitive) and gender (masculine or feminine).

4.2.6.2 Chinese and Japanese

Whereas the picture verification task will test comprehension, the oral question-answer task will elicit oral responses that test effective production and use of morphemes found in the compounds and nominal classifiers.

With respect to compounding, the same target items as those in the picture verification task will be used, eliciting focused productions of several types of

Watorek et al.

head morphemes preceded by a modifier: in Chinese, *-ren*/person (nationality), *-yu*/language, *-jia*/expert, *-dian*/store, etc.; in Japanese, *-jin*/person (nationality), *-go*/language, *-ka*/expert, *-ya*/store, varying the conditions of frequency in Chinese, and frequency and transparency in Japanese. Distractors will also be included in the tasks.

The learners will be given simple instructions: “Describe what you see in the picture”. By introducing images that illustrate the new items that are absent from the input, we will be able to observe to what extent learners are capable of generalising their use of paradigms presented in the language course by making use of their morphological knowledge of the target language.

With respect to classifiers, we will test the production of items that correspond to a certain number of objects, again in the form of images that alternate individual and non-individual classifiers, while also varying the conditions relative to frequency in the Chinese input, and frequency and transparency in the Japanese input, as well as syntactic grammaticality/agrammaticality and semantic relevance/irrelevance.

In Chinese, the following classifiers will be targeted:

- non individual: *-jin* ‘half kilo’, *-he* ‘box’, *-ping* ‘bottle’, *-wan* ‘bowl’
- individual: *-tiao* ‘long and flexible object’, *-zhang* ‘flat object’, *-liang* ‘vehicle’, *-ben* ‘book’

In Japanese, we will elicit productions of the following classifiers:

- non individual: *-kilo* ‘kilo’, *-hako* ‘box’, *-bin* ‘bottle’, *-hai* ‘bowl’
- individual: *-hon* ‘long object’, *-mai* ‘flat object’, *-dai* ‘vehicle’, *-satsu* ‘book’.

Instructions are also important when testing classifiers. As with the compound-ing task, the learners will be given simple instructions: “Describe what you see in the picture”. Simple instructions like this are particularly important in this context in order to avoid leading questions like “how many/much?”, which, in both Japanese and Chinese, signal that a classifier is required for countable objects.

The replication of the oral question-answer task allows us to test learners’ procedural knowledge of nominal morphology in the three target languages: the Nominative/Genitive opposition in Arabic, and the use of compounds and classifiers in Chinese and Japanese. In order to accomplish this task, learners must take into consideration linguistic constraints with respect to morpho-syntax and semantics in their use of nominal morphemes.

4 *Influence of diverse target language properties on input processing*

5 Conclusion

The three replications of the VILLA project described in this chapter and summarised in Table 2 are designed for the purpose of comparing the initial processing and acquisition of typologically different languages by native speakers of French. More specifically, comparisons will be made between the acquisition of Polish, the target language of the VILLA project, and each of the three target languages of the replication studies, Arabic, Chinese and Japanese.

Following Marsden et al. (2018), one principled change to a key variable of the initial study to test generalisability, the target language in this case, might designate a “partial” replication. While designing these replication studies, however, it became clear that including target languages that are typologically different from the target language of the initial study and/or from each other implies changes to other variables, such as the linguistic features under investigation. For this reason, rather than “partial” replications, this chapter describes three “conceptual” replications, each of which introduces more than one change relative to the initial study. We focus here on the challenges inherent in conducting this type of replication study, to which we attempted to respond by posing the following questions:

- What linguistic paradigms examined in the target language Polish of the VILLA project can be considered “equivalent” in Arabic, Chinese and Japanese, and how will these be presented in the classroom input?
- What Polish tasks designed for the VILLA project can be adapted for Arabic, Chinese and Japanese?

With respect to the first question, linguistic paradigms that allow for an investigation into similar acquisition processes as those examined in the VILLA project were identified in the replication languages. Firstly, we limited the replications to the study of nominal morphology. In this way, processes observed in the acquisition of Polish and Arabic, languages that attest rich nominal morphology, could be compared. For Chinese and Japanese, the acquisition of morphological compounding and nominal classifiers were selected as linguistic paradigms that might require similar processing on the part of the learners to that of Polish and Arabic nominal morphology. Results of the VILLA project reveal a degree of learner sensitivity to morphological markers for all L1s of the project. In a similar manner, we predict that learners of Chinese and Japanese will show signs of morphological awareness when processing and producing nominal morphemes

Table 2: Tasks of the VILLA project and plans for replication in Arabic, Chinese and Japanese

Task name	Task type	Task focus	Measures used	Time intervals ^a	Target languages		
					Polish ^b	Arabic	Chinese Japanese
Grammaticality judgement	receptive/ judgement	case marking	accuracy reaction time	Time 1: after 4.5 hrs. Time 2: after 10.5 hrs.	Yes	Yes	No No
Picture verification	receptive/ choice or judgement	case marking, word order, compounding, nominal classifiers	accuracy	Time 1: after 9 hrs. Time 2: after 13.5 hrs.	Yes	Yes	Yes Yes
Oral question and answer	production	case marking, compounding, nominal classifiers	accuracy	Time 1: after 4.5 hrs. Time 2: after 10.5 hrs.	Yes	Yes	Yes Yes

^aWith respect to hours of input
^b(VILLA)

4 *Influence of diverse target language properties on input processing*

in these target languages. Secondly, certain input properties of the initial study were maintained, namely the frequency and transparency of the lexical items to be taught during language instruction. Frequency poses few problems for replications; the VILLA project carefully defined criteria for frequency along with a clear protocol for controlling and documenting the target language input, which all replication studies can follow. Transparency, on the other hand, proved to be particularly challenging. Although not surprising, transparency tests conducted in preparation for the three replication studies revealed how sensitive transparency is to typological difference. Thirdly, given that the Polish lessons of the VILLA project used a communication-based approach to language teaching, it was important to choose similar themes that fit this model in order to preserve comparability across the studies. Indeed, the replications followed the Polish protocol and included lexical items within the realm of professions, nationalities, and food. Within the functional framework adopted in the VILLA project and its replications, the acquisition of specific linguistic paradigms is always studied within a communicative context. To this end, properties in the input are presented as tools of communication in comparable situations.

The second major challenge of these replication studies involves the selection and design of target language tasks. Taking into account the points mentioned above about the challenges of replicating the study of the acquisition of nominal morphology in different target languages, we have designed tasks relative to the linguistic paradigms selected. In Arabic, as per the VILLA project, we will test sensitivity to morphological marking. In Chinese and Japanese, the tasks are designed to test morphological awareness when learners are exposed to morphological compounding and nominal classifiers in the input. The variables of frequency and transparency will be incorporated into the language tasks when possible, and the same themes used in the Polish instruction and tasks, such as professions and nationalities, will be used as content in the replication tasks as well.

Although ecological “live” input studies cannot be replicated with exact precision, they can be replicated in a variety of ways, as we have shown here. These replications are essential for the future of input processing research in that replications, even if partial or conceptual, help refine hypotheses and tighten methodology. Despite the many challenges identified in this chapter, this description and analysis of the methodology of cross-linguistic replication studies reveals that properties of the input within and across studies, such as frequency, transparency in some cases, and certain linguistic paradigms, can be closely replicated. Most importantly, replications require particularly careful planning in the pre-data collection phase, and when this occurs, the field of applied linguistics will,

Watorek *et al.*

without a doubt, benefit from such studies in the future.

Acknowledgments

We sincerely thank Amanda Edmonds, Pascale Leclercq and Aarnes Gudmestad for their support, encouragement and helpful suggestions throughout the process of writing this chapter. We would also like to thank our reviewers, whose comments contributed significantly to improving this work.

The VILLA project was supported by a grant from ORA (Open Research Area in Europe for the Social Sciences) across three granting agencies: ANR in France, DFG in Germany, and NWO in the Netherlands. The British Academy and a PRIN grant supported the English and Italian teams. Additional funding was received from the French lab *Structures Formelles du Langage* (UMR 7023 – CNRS) and from the University of Paris 8.

References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Bardovi-Harlig, Kathleen. 1992. The relationship of form and meaning: A cross-sectional study of tense and aspect in the interlanguage of learners of English as a second language. *Applied Psycholinguistics* 13. 253–278.
- Bardovi-Harlig, Kathleen. 2000. *Tense and aspect in second language acquisition: Form, meaning and use*. Oxford: Blackwell.
- Braine, Martin D. S., Ruth E. Brody, Patricia J. Brooks, Vicky Sudhalter, Julie A. Ross, Lisa Catalano & Shalom M. Fisch. 1990. Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language* 29(5). 591–610. DOI: [10.1016/0749-596X\(90\)90054-4](https://doi.org/10.1016/0749-596X(90)90054-4).
- Carlisle, Joanne F. 2003. *Reading Psychology* 24(3–4). 291–322. DOI: [10.1080/02702710390227369](https://doi.org/10.1080/02702710390227369).
- Carroll, Susanne & Elizabeth Widjaja. 2013. Learning exponents of number on first exposure to an L2. *Second Language Research* 29(2). 201–229.
- Cheng, Lisa & Rint Sybesma. 1998. Yi-wan tang, yi-ge Tang: Classifiers and massifiers. *Tsing-Hua Journal of Chinese Studies New Series* 28. 385–412.
- Dimroth, Christine. 2013. Learner varieties. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*, 3256–3263. Malden, MA: Blackwell.

4 Influence of diverse target language properties on input processing

- Dimroth, Christine, Rebekah Rast, Marianne Starren & Marzena Watorek. 2013. Methods for studying a new language under controlled input conditions: The VILLA project. *EuroSLA Yearbook* 13. 109–138.
- Durand, Marie. 2019. *La découverte et la compréhension des profils d'apprenants classification semi-supervisée et acquisition d'une langue seconde*. Saint-Denis, Université Paris 8. (Doctoral dissertation).
- Ellis, Nick C. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24. 143–188.
- Ellis, Nick. C. 2008. Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 end state. In Peter Robinson & Nick Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition*, 372–405. New York NY: Routledge.
- Flege, James E. 2009. Give input a chance! In Thorsten Piske & Martha Young-Scholten (eds.), *Input matters in SLA*, 175–190. Bristol: Multilingual Matters.
- Goldschneider, Jennifer M. & Robert DeKeyser. 2001. Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning* 51. 1–50.
- Gong, Jiang Song. 2010. Graduate Student Theses, Dissertations & Professional Papers. Graduate School, University of Montana. Chinese classifier acquisition: Comparison of L1 child & L2 adult development.
- Gullberg, Marianne, Leah Roberts & Christine Dimroth. 2012. What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International Review of Applied Linguistics in Language Teaching* 50(4). 239–276.
- Gullberg, Marianne, Leah Roberts, Christine Dimroth, Kim Veroude & Peter Indefrey. 2010. *Adult language learning after minimal exposure to an unknown natural language*.
- Han, Zhao-Hong & Zehua Liu. 2013. Input processing of Chinese ab initio learners. *Second Language Research* 29(2). 145–164.
- Hinz, Johanna, Carina Krause, Rebekah Rast, Ellenor Shoemaker & Marzena Watorek. 2013. Initial processing of morphological marking in nonnative language acquisition: Evidence from French and German learners of Polish. *EUROSLA Yearbook* 13. 39–175.
- Hulstijn, Jan & Robert DeKeyser (eds.). 1997. *Testing SLA theory in the research laboratory*. Special issue.
- Ichikawa, Shingo. 2014. The role of morphological awareness in L2 vocabulary development in logographic languages. *Journal of Nagoya Gakuin University* 25(2). 25–37.

Watorek et al.

- Klein, Wolfgang & Clive Perdue. 1997. The basic variety (Or: Couldn't natural languages be much simpler?) *Second Language Research* 13(4). 301–347. DOI: [10.1191/026765897666879396](https://doi.org/10.1191/026765897666879396).
- Kong, Stano. 2012. Adult English speakers' acquisition of Chinese count-mass classifiers. *Journal of Chinese Linguistics* 40(2). 442–477.
- Kouloughli, Djamel Eddine. 2007. *L'arabe*. Paris: Presses Universitaires de France - PUF.
- Ku, Yu-Min & Richard. C. Anderson. 2003. Development of morphological awareness in Chinese and English. *Reading and Writing: An Interdisciplinary Journal* 16. 399–422.
- Larsen-Freeman, Diane. 2010. Not so fast: A discussion of L2 morpheme processing. *Language Learning* 60(1). 221–230.
- Li, Yen-hui Audrey. 1999. Plurality in a classifier language. *Journal of East Asian Linguistics* 8(1). 75–99. DOI: [10.1023/A:1008306431442](https://doi.org/10.1023/A:1008306431442).
- Liang, Neal Szu-Yen. 2008. The acquisition of Chinese shape classifiers by L2 adult learners. In Marjorie K. M. Chan & Hana Kang (eds.), *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, 309–326. Volume 1. Columbus, OH: The Ohio State University.
- Marsden, Emma, Kara Morgan-Short, Sophie Thompson & David Abugaber. 2018. Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning* 68(2). 321–391.
- Meisel, Jürgen M. 1987. Reference to past events and actions in the development of natural second language acquisition. In Wollman C. Pfaff (ed.), *First and second language acquisition processes*, 206–224. Rowley, MA: Newbury House.
- Nishio, Sumikazu. 2000. *Classificateurs numériques en japonais: constructions et catégories*. Lyon. (Doctoral dissertation).
- Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.
- Perdue, Clive (ed.). 1993. *Adult language acquisition: Cross-linguistic perspectives*. Volumes I and II. Cambridge: Cambridge University Press.
- Picallo, M. Carme. 2008. Gender and number in Romance. *Lingue e linguaggio* 7(1). 47–66.
- Piske, Thorsten & Martha Young-Scholten (eds.). 2009. *Input matters in SLA*. Bristol: Multilingual Matters.
- Rast, Rebekah. 2008. *Foreign language input: Initial processing*. Clevedon: Multilingual Matters.
- Rast, Rebekah. 2017. *L'apprentissage et l'enseignement d'une langue étrangère: de l'exposition initiale aux premières productions [foreign language learning and teaching: From first exposure to first productions]*. Grenoble.

4 Influence of diverse target language properties on input processing

- Rast, Rebekah, Christine Dimroth, Marianne Starren & Marzena Watorek. 2017. *Replicating input-based studies, contextual factors, and ecological validity*. Presented at the panel “Consolidating and sustaining a principled replication effort in SLA research” during the European Second Language Association Conference 25, Reading, UK, 31 August 2017.
- Rast, Rebekah, Marzena Watorek, Heather Hilton & Ellenor Shoemaker. 2014. Initial processing and use of inflectional markers: Evidence from French adult learners of Polish. In Zhao-Hong Han & Rebekah Rast (eds.), *First exposure to a second language: Learners’ initial input processing*, 64–106. Cambridge: Cambridge University Press.
- Reber, Arthur. S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 77. 317–327.
- Rott, Susanne. 1999. The effect of exposure frequency on intermediate language learners’ incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition* 21. 589–619.
- Rouveret, Alain. 2016. *Arguments minimalistes: une présentation du Programme Minimaliste de Noam Chomsky*. Lyon: ENS Éditions. DOI: [10.4000/books.enséditions.5588](https://doi.org/10.4000/books.enséditions.5588).
- Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Saturno, Jacopo. 2017. *Inflectional morphology and utterance structure in initial varieties of Polish L2*. Bergamo / Paris: Università degli studi di Bergamo & University of Paris 8. (Doctoral dissertation).
- Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
- Slobin, Dan I. 1985. Introduction. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition: The data*, vol. 1, 3–24. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Starren, Marianne. 2001. *The second time: the acquisition of temporality in Dutch and French as a second language*. Utrecht: LOT.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Watorek, Marzena. 2004. Construction du discours par des apprenants de langues, enfants et adultes. *Acquisition et Interaction en Langue Etrangère* 20. 129–171.
- Watorek, Marzena, Marie Durand & Katarzyna Starosciak. 2016. L’impact de l’input et du type de tâche sur la production de la morphologie nominale en polonais par des apprenants francophones débutants. *Discours* 18. DOI: [10.4000/discours.9163](https://doi.org/10.4000/discours.9163).

Marzena Watorek, Rebekah Rast, Xinyue Cécilia Yu, Pascale Trévisiol, Hedi Majdoub, Qianwen Guan & Xiaoliang Huang

- Williams, John N. 2005. Learning without awareness. *Studies in Second Language Acquisition* 27(2). 269–304.
- Zhang, Dongbo & Keiko Koda. 2014. Awareness of derivation and compounding in Chinese-English biliteracy acquisition. *International Journal of Bilingual Education and Bilingualism* 17(1). 55–73.
- Zhang, Dongbo, Keiko Koda & Che Kan Leong. 2016. Morphological awareness and bilingual word learning: A longitudinal structural equation modeling study. *Reading and Writing* 29(3). 383–407.
- Zhang, Hong. 2007. Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics* 16(1). 43–59.

Chapter 5

On the relationship between epistemology and methodology: A reanalysis of grammatical gender in additional-language Spanish

Aarnes Gudmestad

Virginia Polytechnic Institute and State University

In the current study I explore the relationship between epistemology and methodology through a reanalysis of production data on grammatical gender in additional-language Spanish that were analysed in [Gudmestad et al. \(2019\)](#). This reanalysis consists of a shift in the epistemology from the one adopted by [Gudmestad et al.](#), where gender marking, which occurs between nouns and both determiners and adjectives, is a unified linguistic phenomenon. In contrast, the assumption in the present investigation is that the acquisition of gender marking entails learning gender assignment and gender agreement, two different learning processes that are observable in language behaviour with determiners and adjectives, respectively. In order to reflect critically on the relationship between epistemology and methodology and specifically on its influence on the interpretation of learner data, I conduct a multi-step analysis that is guided by the differentiation between gender assignment, which can be observed on determiners, and gender agreement, which can be observed on adjectives. I also discuss how the interpretation of the findings can be impacted by the epistemology that guides the current study.

epistemology, methodology, grammatical gender, Spanish, SLA

1 Introduction

As attention has been increasingly paid to methodological reform in applied linguistics ([Byrnes 2013](#); [Phakiti et al. 2018](#)), there have been calls for change on



Aarnes Gudmestad

many fronts, such as open science (Marsden & Plonsky 2018), the reporting of quantitative results (Larsen-Hall & Plonsky 2015), and the need for replication (Porte & McManus 2018). And consequently, the methodological norms in the field are changing (e.g., Marsden et al. 2018). Improvement in quantitative methods is one of the specific areas that has received the most consideration (e.g., Plonsky 2015) and its import is clear: The veracity of the findings that emerge from statistical tests is contingent on the appropriate use of those tests. Another, perhaps more global, issue that is equally important but seems to have garnered less explicit attention is the connection between methodology and epistemology (Ortega 2005). This relationship pertains to the ways in which methodological practices are linked to epistemology or “what counts as knowledge ... and how this relates to truth, belief, and justification” (Young 2018: 40). In the current study, I aim to contribute to discussions about the connection between methodology and epistemology through a focus on grammatical gender. Specifically, I explore this relationship through a reanalysis of production data on grammatical gender in additional-language¹ Spanish, originally reported on in Gudmestad et al. (2019). This reanalysis follows from a change in the epistemology. Whereas Gudmestad et al. treated gender marking as a single phenomenon, in the current study, gender assignment and gender agreement are considered to be different learning processes that are observable in language behaviour with determiners and adjectives (see Section 2.2 on *Grammatical gender in additional languages* for details). I show how this change in epistemology can orient not only the data analysis but also the interpretation of the findings, thus fundamentally changing what counts as relevant knowledge in the field of second language acquisition (SLA).

2 Background

In this section, I first briefly describe the relationship between epistemology and methodology. I then discuss one specific assumption that exists in research on grammatical gender and that guides the current study. Lastly, I introduce grammatical gender in Spanish and I briefly describe Gudmestad et al. (2019), because I reanalyse the dataset from this previous study in the present investigation.

¹“Additional-language” is an inclusive term that refers to any language learned after the first language (cf. Group 2016).

5 *On the relationship between epistemology and methodology*

2.1 Methodology and epistemology

The methodological decisions that scholars make are linked to many facets of the research process. [Ortega \(2005\)](#), for example, highlights the relationship among methodology, epistemology, and ethics:

Research communities make decisions about best ways to approach the task of producing evidence (methodology) on the basis of agreed-upon notions of the nature of what can, or cannot, be captured and explicated as evidence (epistemology) and by drawing on agreed-upon valuations of what is, or is not, worth understanding and transforming (axiology). ([Ortega 2005](#): 317)

In brief, this connection among different components of scholarship means that when reflecting on methodological practices of interpreting data, it is also valuable to consider other aspects of the research process. While each of the three issues highlighted by [Ortega](#) is important, I focus the present investigation on the connection between epistemology and methodology.

[Creswell & Creswell \(2018: 5\)](#) note that epistemologies or ontologies are also called worldviews or paradigms by some scholars and that, regardless of the term, this dimension of research refers generally to the assumptions that researchers have about their discipline or the world that impact methodological decisions. An example of this link between epistemology and methodology is seen in recent calls for multivariate, quantitative analyses in learner corpus research ([Gries 2005](#)) and SLA ([Plonsky & Oswald 2017](#)). These researchers have argued for the need to move from univariate to multivariate analyses in quantitative scholarship because the latter better align with the complexities of the acquisitional process. In other words, since the epistemology is that there are numerous factors at play in the development of an additional language, then the methodological practices (in this case, the statistical analyses we conduct to examine language behaviour and acquisition) should align with this reality. To illustrate this relationship between epistemology and methodology, I now turn to grammatical gender in additional languages.

2.2 Grammatical gender in additional languages

One assumption that is made in some investigations on grammatical gender in SLA is that learners face two primary learnability issues, which are visible in the marking of gender on different sets of modifiers. The acquisitional challenges are learning the gender of the noun (gender assignment, a lexical property) and matching the gender of a modifier with the gender of the noun (gender agreement or gender concord, a morphosyntactic property). What is more

Aarnes Gudmestad

“learners ... need to acquire gender assignment for individual nouns in their internal grammars before being able to produce correct gender agreement in sentences” (Alarcón 2010: 268). Furthermore, some researchers (e.g., Ayoun 2007; Alarcón 2010; Kupisch et al. 2013) consider that gender marking on determiners reflects gender assignment (e.g., *la película* ‘the_{FEM} movie_{FEM}’), whereas gender marking on adjectives constitutes gender agreement (e.g., *duraznos amarillos* ‘yellow_{MASC} peaches_{MASC}’). With this distinction, data showing targetlike gender marking on determiners are interpreted to indicate that learners have acquired the appropriate gender of the noun, and data exhibiting targetlike gender marking between nouns and adjectives are understood to reflect learners’ ability to match the gender of modifiers with a noun’s gender. Investigations that subscribe to this epistemology have found lower rates of targetlike gender marking on adjectives compared to determiners, and this observation has been interpreted as an indication that the morphosyntactic marking of grammatical gender (i.e., gender agreement) is a more challenging learnability issue for learners than assigning a noun its appropriate gender (i.e., a lexical property). Thus, concerning the connection between epistemology and methodology, the assumption among Ayoun, Alarcón, and Kupisch et al. is that gender assignment and gender agreement are different learning processes that are observable in linguistic behaviour, as seen through gender marking on determiners and adjectives, respectively. In the present investigation, I adopt this epistemology, which I refer to as the assignment-agreement assumption. Under this epistemology, researchers can then make the methodological decision to conduct analyses that enable them to distinguish between gender marking on determiners and gender marking on adjectives. When differences are found between the two modifier types, they can be interpreted as evidence in support of this epistemology.

It is important to recognize, however, that the assumption that links gender assignment with determiners and gender agreement with adjectives is not held among all researchers who have investigated grammatical gender (see also Bruhn de Garavito & White 2002; Montrul et al. 2008: 510). As Gudmundson (2013) observes:

this difference is considered to be a theoretical one, difficult to apply in practice. The difference between assignment errors and agreement errors would be applicable to only a very small number of cases, produced several times by the same learner. This is very seldom the case, as agreement tokens frequently occur only once, and sometimes a correct form co-occurs with an incorrect form. It is thus impossible to judge whether an error is due to

5 On the relationship between epistemology and methodology

assignment or to agreement without running the risk of drawing incorrect conclusions. (Gudmundson 2013: 242)

In other words, linking the assignment-agreement assumption to analyses of targetlike behaviour according to modifier type is not without criticism, and Gudmundson highlights a methodological challenge to this epistemology: Researchers need to be able to observe a given learner's gender marking on multiple occurrences of the same noun (rather than just a single occurrence of a noun). For instance, if a participant uses the noun *libro* 'book' with a modifier only one time and the noun *mesa* 'table' with a modifier three times, then, according to Gudmundson, researchers can make observations about gender marking on *mesa* but not *libro*. The goal of the current study is not to take a position on whether or not the assignment-agreement assumption is valid but rather to reflect critically on the impact that it can have on methodological practices and the interpretation of data.

2.3 Grammatical gender in Spanish

In Spanish every noun has masculine or feminine gender and modifiers (i.e., determiners and adjectives) agree in gender with the noun they modify, as illustrated in (1). Gender is assigned according to biological sex for some nouns (*mujer* 'woman_{FEM}', *hombre* 'man_{MASC}'). For most nouns, however, the gender is assigned arbitrarily, such as those in (1). The canonical morpheme for nouns and modifiers is *o* for masculine and *a* for feminine, though there are exceptions (e.g., *mapa* 'map_{MASC}'). Furthermore, not all nouns and modifiers end in these vowels. Regarding nouns, there are other inflectional endings that are predictive of one gender (e.g., *tad* for feminine nouns as in *lealtad* 'loyalty' and *e* for masculine nouns, e.g., *bate* 'bat'), as well as endings that are not linked to a particular gender (e.g., *s*; *lunes* 'Monday_{MASC}' versus *oasis* 'oasis_{FEM}'; Teschner & Russell 1984). Concerning modifiers, not all determiners and adjectives are overtly marked for gender either (e.g., *tu* 'your' and *difícil* 'difficult').

- (1) a. El vestido roj-o
 the.M dress red-M
 'The_M red_M dress_M.'
- b. La bicicleta car-a
 the.F bike expensive-F
 'The_F expensive_F bike_F.'

Aarnes Gudmestad

Research on grammatical gender in additional-language Spanish spans various theoretical and analytical approaches (e.g., Universal Grammar, variationist SLA), has examined language processing and production (cf. Alarcón 2014), and includes investigations that subscribe to the aforementioned assignment-agreement assumption (e.g., Alarcón 2010; Kupisch et al. 2013) and others that do not (e.g., Montrul et al. 2008; Grüter et al. 2012; Gudmestad et al. 2019). I focus here on Gudmestad et al., which serves as a starting point for the reanalysis in the current study. In Gudmestad et al., we examined gender marking in language production using the longitudinal corpus LANGSNAP (<http://langsnap.soton.ac.uk>, e.g., Mitchell et al. 2017). Our epistemology was that in language production researchers can make observations about one acquisitional challenge pertaining to grammatical gender – the marking of gender on modifiers. Thus, we made no distinction between gender assignment and gender agreement and analysed each instance of the use of a noun with a modifier (determiner or adjective) that was overtly marked for gender. We adopted a variationist approach (Geeslin & Long 2014), which means that we sought to account for the variability in learners' marking of grammatical gender over time by explaining the linguistic and extra-linguistic factors that conditioned the participants' use of targetlike gender marking (see the *Method* section below for more information on the data, participants, variables, etc.). In general, we found that numerous factors worked together to condition learners' use of targetlike gender marking and that the factor of noun ending helped to explain changes in use along the developmental trajectory. It is worth pointing out that modifier type (determiners versus adjectives) was one of the factors we investigated. And, while we found that learners were more likely to be targetlike in their gender marking with determiners compared to adjectives, we did not interpret these findings in relation to the assumption that determiners reflect a lexical property and adjectives a morphosyntactic one. We interpreted the findings, instead, as evidence of the complex nature of variability in language use and development, such that modifier type was just one of several linguistic features that impacts how learners develop the ability to mark gender on modifiers in a targetlike way. In the present study, I reanalyse the dataset from Gudmestad et al. through the lens of the assignment-agreement assumption.

3 The current study

In order to consider how the assignment-agreement assumption may influence methodological decisions and the interpretation of data pertaining to the additional-language development of grammatical gender marking, I reanalyse the data from

5 *On the relationship between epistemology and methodology*

Gudmestad et al. (2019). The current study consists of a three-step data analysis in which I examine determiners and adjectives separately. I then interpret the findings in light of the assignment-agreement assumption and reflect on how new knowledge can emerge from this epistemology. In general, this type of re-analysis, in which assumptions are modified, has the potential to shed light on the link between epistemology and methodology highlighted by Ortega (2005). More specifically, I aim to concretely demonstrate how an epistemological shift leads to a particular methodological decision that, in turn, leads the researcher down a new interpretive path.

3.1 Method

3.1.1 Data

I examined data from the LANGSNAP corpus (<http://langsnap.soton.ac.uk/>, e.g., Mitchell et al. 2017). The corpus consists of production data collected over 21 months, which included an academic year abroad, from additional-language learners of Spanish. The data were collected six different times and at each point the participants completed three tasks: a written argumentative essay, an oral interview, and an oral narration.² For the essay, the participants were presented with a topic and asked to write a 200-word composition. The semi-guided interview consisted of opinion questions and questions about the participants' lives; it lasted about 20 minutes. The oral narration was a picture-based task. The participants looked over a set of images and then told the story in their own words. In the present investigation, I report on the data from all tasks that were collected at three of the data-collection periods (cf. Gudmestad et al. 2019). The first data-collection period, called pre-stay in the current study, was collected before the learners went abroad. The second data-collection point that I analysed was the third in-stay period in the LANGSNAP corpus (henceforth, in-stay); this data collection took place a year after the pre-stay and at the end of the academic year abroad. The final point was gathered 21 months after the pre-stay and was the second post-stay data collection in the LANGSNAP corpus (hereafter, post-stay).

3.1.2 Participants

I analysed data from 21 of the 27 learners of Spanish in the corpus.³ They were all pursuing an undergraduate degree in Spanish at a British university and had been

²I analyse the data from the three tasks together.

³All of the data were coded by hand. Due to how labour intensive this coding was, Gudmestad et al. (2019), and consequently the current study, analysed data from a subset of the participants

Aarnes Gudmestad

studying Spanish for an average of 5.4 years ($SD = 3.4$, range: 2-14 years). They ranged in age from 20 to 25 years ($M = 20.8$ years, $SD = 1.6$). Fifteen were women and six were men. Their first languages were Polish ($n = 1$), English ($n = 19$), and both English and Polish ($n = 1$). At the pre-stay, the participants completed a global proficiency measure – an elicited-imitation task. The group scored an average of 86.1 out of 120 points ($SD = 12.7$; range: 50-108). During the academic year abroad, they were teaching assistants ($n = 10$), exchange students ($n = 9$), and workplace interns ($n = 2$). Five participants were in Mexico and 16 were in Spain.

3.1.3 Coding and analysis

The coding started by identifying each occurrence of a determiner or an adjective that modified a referent ($K = 16,357$); only those modifiers that met two criteria were then analysed ($k = 11,832$). The first criterion was that the modifiers needed to exhibit overt gender marking. For example, an adjective like *bonito/a* ‘pretty’ was included in the analysis because it has an inflectional gender morpheme, but adjectives like *interesante* ‘interesting’ were excluded because the form is the same, regardless of whether it modifies a feminine or masculine noun. Second, the current dataset consists only of nouns; pronouns that were modified by adjectives were not analysed (e.g., *ella está contenta* ‘she is happy’). Following the assumption that gender marking on determiners and adjectives reflect different learning processes, I then separated the data by modifier type, determiners ($k = 9,107$) and adjectives ($k = 2,725$), in order to examine each modifier type separately. An example of the data is available in (2).

- (2) Tengo un-a amig-a español
 have.1S INDEF-F friend-F spanish.M
 ‘I have a_F friend_F Spanish_M.’ (Participant 165, post-stay, interview)

I analysed gender agreement (as seen on determiners) and gender assignment (as seen on adjectives) in three phases. The first two phases served to examine claims made in previous research about the differences between the two learning challenges mentioned in the literature review. The third step sought to further knowledge of the potential differences between these two processes by identifying factors that explain patterns in the data. The dependent variable for each step of the analysis was the targetlikeness of the gender marking: targetlike (the

and three of the six total data-collection points. The learners analysed in Gudmestad et al. and the current study were the first 21 participants in the corpus.

5 On the relationship between epistemology and methodology

gender of the modifier matched the gender of the noun) or nontargetlike (the gender of the modifier differed from that of the noun).

For the first phase of the analysis, I explored the assertion that gender assignment is acquired before gender agreement and that, under the assignment-agreement assumption, this claim leads to the expectation that targetlike use with determiners is higher than that with adjectives (cf. Alarcón 2010). In order to address this issue, I identified the mean rate of targetlike use for adjectives and determiners at each data-collection point. With a two-way ANOVA, I also assessed whether the rates of targetlike use at each data-collection point were similar or different between determiners and adjectives.

Next, some previous research that subscribes to the assignment-agreement assumption appears to consider the acquisition of gender assignment to be binary: Either learners have acquired a noun's gender or they have not (e.g., Alarcón 2010). In order to address this claim in the second phase of the analysis, I sought to determine whether gender assignment (a lexical property) and gender agreement (a morphosyntactic property) resulted in categorical behaviour of gender marking. I examined targetlike assignment and agreement with individual nouns that participants used more than once at pre-stay; this assessment shows how many unique nouns that participants produced more than once exhibited categorical targetlike use.⁴ It might be expected to find that, with determiners, learners exhibit either categorical targetlike or categorical nontargetlike use on individual nouns (i.e., rather than a mix of the two with a given noun, when a participant uses the noun more than once). However, the hypothesis for gender agreement may be different. Under the assumption that gender marking on adjectives reflects a morphosyntactic process, it may be reasonable to find that a noun, when used multiple times by a participant, shows targetlike agreement in some instances and nontargetlike agreement in others. This variability may be expected because the morphosyntactic features of an agreement relationship can differ each time a noun is used. For example, in one instance the adjective may be attributive, occurring in the noun phrase (*Tengo un gato blanco*. 'I have a white cat.') and in another case the adjective may be predicative, connected to the noun by means of a verbal phrase (*Mi gato es blanco*. 'My cat is white').

In light of the findings from the second part of the analysis (see the *Results and Discussion* section), which provided preliminary evidence of variability with both gender assignment and agreement in language production, I sought to explain

⁴While this analysis may be valuable for each data-collection point, I focus on the pre-stay data in order to offer an example of what this type of analysis may contribute to the understanding of grammatical gender marking.

Aarnes Gudmestad

this variability in targetlike gender marking through multivariate analyses. Thus, for the third phase in the analysis, I adopted a variationist approach (Geeslin & Long 2014) in order to investigate variable gender assignment and agreement in language production. This approach, which was also employed in Gudmestad et al. (2019), models variable language behaviour by examining a range of factors (i.e., independent variables, fixed effects) simultaneously. Through two separate multivariate analyses, I identified which factors significantly impacted gender assignment and those that predicted gender agreement. If these two processes are indeed distinct, it may be expected to find that there are some conditioning factors that differ between the two learning challenges.

In order to conduct multivariate analyses, I examined nine fixed effects for both determiners and adjectives, all factors that were motivated by previous research (see Gudmestad et al. 2019 for justification of these factors): noun gender, noun ending, noun class, noun number, task, time, noun frequency (individual), noun log-frequency (language), and initial proficiency. Noun gender distinguishes between feminine and masculine nouns. For noun ending there were four categories: canonical, deceptive, predictive, and other endings. Canonical endings refer to masculine nouns that end in *o* and feminine nouns that end in *a*, and deceptive endings are the opposite: Masculine nouns ending in ‘*a*’ and feminine nouns ending in *o*. Predictive endings are those that are strongly linked to one gender (e.g., ‘dad’ is linked with feminine gender, Teschner & Russell 1984) and other endings are those that are not strongly connected with one gender (e.g., *s* Teschner & Russell 1984). Noun class differentiates between nouns with biological and arbitrary gender. Noun number explores possible differences between singular and plural nouns. Task pertains to the oral interview, the oral narration, and the written argumentative essay. Time distinguishes between the pre-stay, in-stay, and post-stay data-collection periods. The four remaining factors that were investigated for both adjectives and determiners were continuous factors. Noun frequency (individual) refers to the number of times that each learner produced a noun with a gender-marked modifier in a specific task and data-collection point. The noun log-frequency (language) refers to how often a noun occurs per one million words in the *Corpus del español* (Davies 2016-). Initial proficiency considers the score that each participant received on the elicited-imitation task before going abroad (see Section 3.1.2). Furthermore, I coded for one factor that was unique to determiners and one that was unique to adjectives – two factors that were not examined in Gudmestad et al. since all determiners and adjectives were analysed together. Determiner type was investigated for determiners only. This factor was motivated by Bruhn de Garavito & White (2002) who found higher targetlike gender marking with definite articles compared to

5 On the relationship between epistemology and methodology

indefinite articles. In the current study, I examined a wider array of categories: definite article (*la* ‘the_{FEM}’), indefinite article (e.g., *un* ‘a_{MASC}’), demonstrative-this (e.g., *estos* ‘these_{MASC}’), demonstrative-that (e.g., *esa* ‘that_{FEM}’), indeterminate (e.g., *alguna* ‘some/any_{FEM}’), and possessive (*nuestra* ‘our_{FEM}’). The factor investigated for adjectives only was adjective position. The three categories were pre (the adjective came before the noun in the same noun phrase), post (the adjective was after the noun in the same noun phrase), and other (the adjective was in a different phrase than the noun). Prior studies have offered conflicting evidence as to whether adjective position plays a role in additional-language development (e.g., Bartning 2000; Dewaele & Véronique 2001). Finally, participant was examined as a random effect, in order to account for variability among the learners. In terms of the analysis, I fit two mixed-effects regression models – one for determiners and one for adjectives – using the statistical software R (Team 2017). Factors not found to be significant were removed from the statistical models. After the significant fixed effects were identified, I explored interactions between time and each of the remaining fixed effects in order to make observations about change over time. I also tested for correlations between independent variables to ensure that there were no strong correlations among the factors included in each regression model. Finally, I reported the McFadden’s R^2 (Smith & McKenna 2013) for each model, a metric that indicates whether each model does a good job of fitting the data. With this third phase in the analysis, I compared the determiner model with the adjective model in order to make observations about similarities and differences between gender assignment and gender agreement.

3.2 Results and discussion

In this section, I present the findings of each of the three steps of the data analysis. I also discuss the findings in relation to the assignment-agreement assumption. As a reminder, my objective is not to take a stance on the validity of the assignment-agreement assumption. Instead I am to reflect on the role it plays in methodological decisions and data interpretation.

3.2.1 Rates of targetlike use

Table 1 provides the average rate of targetlike use for the learners according to modifier type (adjective or determiner) and time (pre-stay, in-stay, and post-stay). I conducted a two-way ANOVA to examine the effect of modifier type and time on targetlike use. The interaction between modifier type and time was not significant, $F(2, 120) = 2.152$, $p = 0.121$. However, the main effect for modifier type

Aarnes Gudmestad

was significant ($F(1, 120) = 49.44, p < 0.001$), indicating that the participant group was more targetlike with determiners ($M = 96.69, SD = 2.52$) than adjectives ($M = 90.43, SD = 7.52$). This finding can be interpreted as evidence that is consistent with the epistemology that gender marking on determiners reflects gender assignment and gender marking on adjectives reflects gender agreement. Therefore, since learners need to acquire a noun’s gender (a lexical property) before being able to use targetlike gender agreement (a morphosyntactic property), the finding that targetlike gender marking was higher with determiners than adjectives was expected based on previous research (e.g., Alarcón, 2010). The main effect for time was also significant ($F(2, 120) = 15.705, p < 0.001$). The learners were more targetlike at in-stay ($M = 95.312, SD = 4.045$) and post-stay ($M = 95.340, SD = 4.260$) compared to pre-stay ($M = 90.029, SD = 8.456$),⁵ but there was no significant difference between in-stay and post-stay ($p = 1.000$). These findings suggest improvement in targetlikeness of gender marking as a whole during the academic year abroad that was maintained after returning to the United Kingdom.

Table 1: Rates of targetlike use (in percentages)

	Determiners			Adjectives		
	Pre-stay	In-stay	Post-stay	Pre-stay	In-stay	Post-stay
M	94.46	97.95	97.67	85.60	92.67	93.01
SD	2.69	1.40	1.60	9.91	4.12	4.82
Median	93.90	98.10	97.80	84.00	92.10	93.90
Range	89.0–98.7	94.8–100	93.4–100	64.3–100	85.0–100	79.3–98.1

3.2.2 Individual nouns at pre-stay

Next, I examined targetlikeness of gender marking for the individual nouns that each participant produced more than once with a modifier overtly marked for gender at pre-stay. This analysis focuses on nouns that learners used more than once at pre-stay, rather than those that were used just one time, given the argument by Gudmundson (2013: 242) that researchers need to examine multiple occurrences of a particular noun, in order to make observations about gender assignment and gender agreement.

⁵The p values for the pre-stay and in-stay comparison and the pre-stay and post-stay comparisons are both $p < 0.01$.

5 On the relationship between epistemology and methodology

The results for gender assignment (as seen on determiners at pre-stay) are presented in Table 2. For example, participant 150 used a total of 103 different nouns with a determiner overtly marked for gender at pre-stay, of which 50 were used more than once. With 48 of the nouns that she used more than once, the gender of the determiner was targetlike 100 percent of the time. In contrast, one noun (*rana* ‘frog’) exhibited variable targetlike use and one noun (*programa* ‘program’) was consistently used with a determiner that did not match the gender of the noun (i.e., categorical nontargetlike use). This participant also used 53 nouns just one time at pre-stay.

Continuing with Table 2, the results indicate that each participant assigned the targetlike gender to most nouns that they used at least twice. Only three participants used one unique noun multiple times and always used a determiner that differed in gender from the noun it modified (participants 150, 156, and 168). Moreover, each participant used between one and 10 nouns in which some instances exhibited targetlike use of gender on the determiner and others did not (i.e., use of both masculine and feminine determiners with the same noun). For example, participants 158 and 164 used both masculine and feminine determiners with four unique nouns. The nouns were *día* ‘day’, *objeto* ‘object’, *problema* ‘problem’, and *telenovela* ‘soap opera’ for participant 158 and *apartamento* ‘apartment’, *casa* ‘house’, *idea* ‘idea’, and *mujer* ‘woman’ for participant 164. Under the assumption that gender marking on determiners reflects gender assignment (i.e., the lexical property of noun gender), this observation may be surprising as it suggests evidence of variable knowledge or of varying degrees of the strength of the lexical representations between nouns and their gender (Halberstadt et al. 2018). Thus, it may be that a noun that exhibits variability in targetlike use on determiners has a weaker gender representation than a noun whose gender assignment is categorical. These findings contrast with previous research that has considered gender assignment to be a categorical property (cf. Alarcón 2010).

Turning to gender agreement at pre-stay, I assessed targetlike gender marking on adjectives for the unique nouns produced more than once by individual participants. These results are available in Table 3, which is organized like Table 2. Similar to gender assignment, learners exhibited targetlike gender agreement with most nouns. For example, among the 19 nouns that participant 150 used at least twice with an adjective, she exhibited targetlike gender agreement with 16 of them. Additionally, instances where learners used the same noun with an adjective multiple times but produced nontargetlike gender agreement categorically were uncommon. Participants 156, 157, 160, 166, and 167 each used one noun multiple times and were nontargetlike in their gender agreement every time they used that noun with an adjective. Moreover, there is variability in the marking

Aarnes Gudmestad

Table 2: Unique nouns and targetlikeness with determiners at pre-stay.
Note: Percentages may not add up to 100 due to rounding.

Part.	Nouns used more than once						Nouns used once		Total
	All targetlike		Variable		All nontargetlike		#	%	
	#	%	#	%	#	%			
150	48	46.60	1	0.97	1	0.97	53	51.46	103
151	29	37.18	5	6.41	0	0	44	56.41	78
152	29	33.33	1	1.15	0	0	57	65.52	87
155	32	39.02	1	1.22	0	0	49	59.76	82
156	19	24.36	2	2.56	1	1.28	56	71.79	78
157	16	25.81	6	9.68	0	0	40	64.52	62
158	29	31.87	4	4.40	0	0	58	63.74	91
160	28	35.00	1	1.25	0	0	51	63.75	80
161	20	33.90	2	3.39	0	0	37	62.71	59
162	25	29.76	2	2.38	0	0	57	67.86	84
163	18	22.78	3	3.80	0	0	58	73.42	79
164	25	37.31	4	5.97	0	0	38	56.72	67
165	22	33.3	3	4.55	0	0	41	62.12	66
166	41	33.6	3	2.46	0	0	78	63.93	122
167	26	28.89	6	6.67	0	0	58	64.44	90
168	20	32.25	2	3.23	1	1.61	39	62.90	62
169	20	25.32	10	12.66	0	0	49	62.03	79
170	23	31.08	1	1.35	0	0	50	67.57	74
171	22	25.29	1	1.15	0	0	64	73.56	87
172	34	33.01	3	2.91	0	0	66	64.08	103
173	24	32.43	3	4.05	0	0	47	63.51	74

5 On the relationship between epistemology and methodology

of gender on adjectives with some individual nouns. For instance, participants 150 and 163 exhibited variable gender marking on adjectives with three unique nouns. Participant 150 was variable with *chica* ‘girl’, *hombre* ‘man’, and *mujer* ‘woman’ and participant 163 was variable with *idea* ‘idea’, *identidad* ‘identity’, and *persona* ‘person’.

Under the assignment-agreement assumption, one might expect to see a higher proportion of nouns that exhibit variable targetlike use with adjectives than with determiners, given the fact that morphosyntactic properties can differ based on the linguistic context. Focusing exclusively on nouns that the participants used more than once and that exhibited variability, the average number of nouns that were connected to variable behaviour was higher for adjectives than determiners: 17.81 percent (26/146 nouns) of the group’s nouns that were used more than once by individual participants exhibited variable gender agreement, whereas 10.37 percent of their nouns (64/617 nouns) exhibited variable gender assignment.

3.2.3 Mixed-effects regression models

Findings from the second step of the analysis appeared to show variation in gender assignment and agreement. Specifically, the analysis in Section 3.2.2 pointed to the possibility that knowledge of the lexical property of gender assignment is not always categorical. It also indicated that targetlike gender agreement is not always categorical either. In light of these observations, it seems reasonable to look to research approaches in SLA that have implemented methodological tools for investigating variation in order to understand the factors that impact variability in gender assignment and agreement and to make comparisons between the two learning processes. In this vein, I adopt a variationist approach in order to investigate a range of factors that may condition variable gender marking on determiners and adjectives separately (see Section 3.1.3 for a general description of the type of multivariate analysis that is common in variationist SLA and for details on the factors I investigate).

I present the findings for two mixed-effects regression models (Tables 5 through 8). For the dependent variable and the nominal independent variables, both models compare a reference-point category of each variable to the other category (or categories) of the same variable. The reference point for the dependent variable is targetlike use and the reference points for all significant fixed effects are provided in the Tables 5 and 7 in brackets. The continuous fixed effects do not have reference points. The estimate listed with each category in the tables indicates whether there is a decrease (indicated by a negative estimate) or an increase (denoted by a positive estimate) in the log odds of targetlike use. The *p* value

Aarnes Gudmestad

Table 3: Unique nouns and targetlikeness with adjectives at pre-stay.
Note: Percentages may not add up to 100 due to rounding.

Part.	Nouns used more than once						Nouns used once		Total
	All targetlike		Variable		All nontargetlike		#	%	
	#	%	#	%	#	%			
150	16	40.00	3	7.50	0	0	21	52.50	40
151	6	28.57	1	4.76	0	0	14	66.67	21
152	5	21.74	0	0	0	0	18	78.26	23
155	4	16.67	1	4.17	0	0	19	79.17	24
156	3	13.04	1	4.35	1	4.35	18	78.26	23
157	2	12.5	1	6.25	1	6.25	12	75.00	16
158	8	26.67	0	0	0	0	22	73.33	30
160	1	4.76	2	9.52	1	4.76	17	80.95	21
161	4	19.05	2	9.52	0	0	15	71.43	21
162	8	25.00	0	0	0	0	24	75.00	32
163	8	25.00	3	9.38	0	0	21	65.63	32
164	0	0	1	4.17	0	0	23	95.83	24
165	9	24.32	1	2.70	0	0	27	72.97	37
166	12	34.29	0	0	1	2.86	22	62.86	35
167	3	10.00	2	6.67	1	3.33	24	80.00	30
168	3	18.75	2	12.5	0	0	11	68.75	16
169	7	26.92	2	7.69	0	0	17	65.38	26
170	1	5.00	1	5.00	0	0	18	90.00	20
171	2	7.41	1	3.70	0	0	24	88.89	27
172	5	16.13	1	3.23	0	0	25	80.65	31
173	8	36.36	1	4.55	0	0	13	59.09	22

5 *On the relationship between epistemology and methodology*

(alpha level of $p < 0.05$) reveals whether the estimate is significant. When nominal independent factors have more than two categories (as is the case with noun ending, task, time, determiner type, and adjective position), it is also possible to assess whether there are significant differences between non-reference point categories (e.g., in-stay versus post-stay for time). This can be done by examining the confidence intervals of the non-reference point categories. Overlap between the confidence intervals of categories indicates that the log odds of targetlike use is similar. When the confidence intervals of two categories do not overlap, the log odds of targetlike behaviour can be considered to be different.

Beginning with the mixed-effects regression for determiners, it was found that targetlike gender assignment in this dataset was influenced by noun ending, task, noun gender, noun frequency (individual), initial proficiency, determiner type, and time (Table 4). Noun log-frequency (language), noun class, and noun number were not significant, and I found no significant interactions between time and the other significant fixed effects. Furthermore, none of the fixed effects were strongly correlated. The results for the random effect for participant are available in Table 5. The McFadden's R^2 indicated a moderate fit for this model ($R^2_{\text{McFadden}} = 0.1339$).

For noun ending, the log odds of targetlike gender assignment were significantly lower with deceptive and other endings compared to canonical noun endings. Predictive and canonical endings were not statistically different. In the case of the non-reference point categories of noun ending, there was overlap between the other and predictive endings, which revealed that the log odds of targetlike gender assignment were similar between the two. However, the confidence intervals for deceptively marked nouns did not overlap with other and predictive endings and the values for the confidence intervals of deceptively marked nouns were lower than those for the other categories. This finding indicates that the log odds of targetlike use with deceptively marked nouns were lower than those of predictive and other endings. For gender assignment, these results suggest that noun ending played a role in whether learners assigned the targetlike gender to a noun. Specifically, deceptively marked nouns appeared to present learners with the greatest challenge.

For task, the log odds of targetlike gender assignment were lower with the oral narration and the oral interview compared to the written essay. The overlap in the confidence intervals for the two oral tasks also indicated that targetlike use was similar between the two. Thus, the findings demonstrated a difference between oral and written production. For gender assignment, these results were consistent with claims made by researchers who investigated explicit and implicit knowledge (e.g., Ellis 2006). Specifically, written tasks may enable learners

Table 4: Results for the fixed effects in the regression model for determiners. *Note:* The reference point for the dependent variable is target-like use. The reference points for the independent, nominal variables are in brackets in the table.

Effect	Estimate	SE	<i>z</i>	<i>p</i>	Lower	Upper
Intercept	3.54	0.73	4.84	1.30×10^{-6}	0.18	0.56
Noun ending [canonical]						
deceptive	-2.41	0.21	-11.48	$<2.00 \times 10^{-16}$	-2.82	-1.99
other	-0.69	0.15	-4.45	8.80×10^{-6}	-0.99	-0.38
predictive	-0.25	0.18	-1.35	0.18	-0.60	0.12
Task [written]						
narrative	-1.15	0.24	-4.76	1.95×10^{-6}	-1.63	-0.68
interview	-0.99	0.21	-4.69	2.77×10^{-6}	-1.42	-0.59
Noun gender [masculine]						
feminine	-1.33	0.14	-9.25	$<2.00 \times 10^{-16}$	-1.62	-1.05
Noun freq. ^a	0.13	0.03	3.84	0.0001	0.07	0.20
Initial prof.	1.69	0.80	2.11	0.03	0.04	3.38
Determiner type [definite]						
indefinite	-0.77	0.14	-5.62	1.94×10^{-8}	-1.03	-0.50
indeterminate	-1.60	0.31	-5.20	2.01×10^{-7}	-2.17	-0.96
possessive	-1.64	0.63	-2.60	0.01	-2.73	-0.17
demonstr. - this	-1.11	0.30	-3.69	0.0002	-1.67	-0.48
demonstr. - that	-1.57	0.46	-3.45	0.0006	-2.40	-0.57
Time [pre-stay]						
in-stay	1.01	0.14	7.10	1.22×10^{-12}	0.73	1.30
post-stay	0.85	0.17	4.95	7.39×10^{-7}	0.52	1.20

^a(individual)

5 *On the relationship between epistemology and methodology*

Table 5: Results for the random effect in the determiner regression model

Participant	Intercept
150	0.29
151	0.25
152	-0.14
155	0.03
156	-0.09
157	-0.13
158	0.32
160	0.18
161	0.23
162	0.50
163	-0.09
164	-0.01
165	-0.42
166	-0.08
167	-0.59
168	0.09
169	-0.14
170	0.18
171	-0.13
172	-0.33
173	-0.17

to tap into their explicit knowledge more than they do in oral production so it may be that this participant group has greater explicit knowledge, compared to implicit knowledge, of gender assignment.

The results for noun gender showed that the log odds of targetlike gender assignment were lower for feminine nouns compared to masculine nouns. Since previous research has demonstrated that the default gender for learners is masculine (e.g., [López Prego & Gabriele 2012](#)), this result may mean that the default facilitated gender assignment with masculine nouns.

The log odds of targetlike gender assignment were higher as the frequency with which learners used particular nouns increased and as their initial proficiency score increased. Additionally, the log odds of targetlike gender assign-

Aarnes Gudmestad

ment were greater at in-stay and post-stay, compared to pre-stay, and the confidence intervals revealed that targetlike use was similar between in-stay and post-stay. The findings for these three factors showed that as learners become more proficient in the language, as they used individual nouns more often, and after they completed an academic year abroad, their knowledge of gender assignment, as seen through language production, improved.

Finally, for determiner type, the learner group exhibited higher log odds of targetlike gender assignment with definite articles compared to all other determiner types. The confidence intervals for all of the non-reference point categories overlapped, indicating that their log odds of targetlike use were similar. This finding was similar to [Bruhn de Garavito & White \(2002\)](#) who found that learners were more targetlike with definite articles than indefinite articles. It also appears to suggest that the assumption that gender marking on determiners signifies whether learners have assigned the targetlike gender to nouns needs to be nuanced, because all determiner types are not identical when it comes to assigning gender in language production.

Thus, returning to the assignment-agreement assumption, the current study's results do not align with previous research that considers gender assignment to be categorical (i.e., learners have either learned the gender of a noun or not, [Alarcón 2010](#)). Instead, they appear to support the observation that learners can show variable knowledge of a noun's gender in language use and that this variability is conditioned by a range of factors. More generally, they suggest that making assessments about the acquisition of gender assignment in language use involves an analysis that goes beyond a univariate examination of targetlike use of determiner gender.

Continuing with the mixed-effects regression for adjectives, eight fixed effects were significant: Noun ending, task, noun gender, noun log-frequency (language), initial proficiency, noun number, adjective position, and time significantly impacted targetlike gender agreement. Noun class and noun frequency (individual) did not predict gender agreement. The interaction between time and initial proficiency was significant. However, because this interaction correlated with other main effects, I removed it from the model. The McFadden's R^2 indicated a moderate fit ($R^2_{\text{McFadden}} = 0.1563$). The results for the fixed effects are available in Table 6 and the random effect results are in Table 7.

For noun ending, the log odds of targetlike gender agreement were lower with deceptively marked nouns and other noun endings compared to nouns with canonical endings and there was no significant difference between nouns with predictive endings and those with canonical endings. The confidence intervals indicated similarities (i.e., overlap) between other and predictive endings. The

5 On the relationship between epistemology and methodology

Table 6: Results for the fixed effects in the regression model for adjectives. *Note:* The reference point for the dependent variable is targetlike use.

Effect	Estimate	SE	<i>z</i>	<i>p</i>	Lower	Upper
Intercept	−0.94	0.84	−1.123	0.26	−2.69	0.741
Noun ending [canonical]						
deceptive	−2.26	0.33	−6.82	9.21×10^{-12}	−2.90	−1.589
other	−0.54	0.19	−2.88	0.004	−0.90	−0.16
predictive	−0.37	0.22	−1.71	0.09	−0.78	−0.06
Task [written]						
narrative	−0.65	0.27	−2.39	0.02	−1.19	−0.11
Interview	−0.35	0.19	−1.82	0.07	−0.74	0.02
Noun gender [masculine]						
feminine	−1.68	0.18	−9.25	$<2 \times 10^{-16}$	−2.05	−1.33
Noun log-freq. ^a	0.086	0.03	3.43	0.0006	0.04	0.14
Initial prof.	3.36	0.93	3.63	0.0003	1.52	5.30
Noun number [plural]						
Singular	0.43	0.15	2.82	0.005	0.13	0.74
Adj. position [other]						
pre	1.58	0.21	7.55	4.45×10^{-14}	1.17	1.99
post	0.91	0.20	4.46	8.38×10^{-6}	0.51	1.31
Time [pre-stay]						
in-stay	0.76	0.17	4.55	5.41×10^{-6}	0.43	1.09
post-stay	0.87	0.19	4.46	8.29×10^{-6}	0.49	1.25

^a(language)

Aarnes Gudmestad

Table 7: Results for the random effect in the determiner regression model

Participant	Intercept
150	0.41
151	-0.15
152	0.11
155	0.08
156	-0.32
157	-0.17
158	0.38
160	-0.12
161	0.05
162	0.25
163	-0.41
164	0.11
165	0.23
166	-0.02
167	-0.72
168	0.35
169	0.25
170	-0.07
171	-0.05
172	-0.25
173	-0.17

confidence intervals also demonstrated that the participants were less likely to use targetlike gender agreement with deceptively marked nouns compared to nouns with other and predictive endings. These findings suggest that targetlike gender agreement was most challenging for these learners when the noun has a deceptively marked ending. These results were similar those for determiners, which indicated that nouns with deceptive endings posed challenges for gender assignment.

For task, the log odds of targetlike gender assignment were lower with the oral narration compared to the written essay, and there was no significant difference between the oral interview and the essay. The overlap in the confidence intervals for the two oral tasks indicated that targetlike gender agreement is similar

5 *On the relationship between epistemology and methodology*

between the two. Although task constrained both gender assignment and agreement for these participants, it may be worth noting a difference between the two. Unlike the findings for gender assignment, which pointed to a difference between the oral and written modes, the interview task was statistically similar to both the essay and the oral narration in gender agreement.

The results for noun gender demonstrate that participants were less likely to be targetlike in their gender agreement with feminine nouns compared to masculine nouns. Just as with gender assignment, learners exhibited greater challenges with gender agreement when the nouns were feminine, perhaps pointing again to the claim that the masculine gender is the default (López Prego & Gabriele 2012). Furthermore, participants were less likely to be targetlike in their gender agreement with plural nouns compared to singular nouns, which have also been considered to be a default for learners (López Prego & Gabriele 2012). The results for noun number constituted a difference between gender assignment and gender agreement, as this factor did not significantly predict targetlike use with determiners.

For the continuous factors, the log odds of targetlike gender agreement increased as noun log-frequency (language) increased; this factor was considered to be an indirect measure of input frequency (Gudmestad et al. 2019). The likelihood of targetlike gender agreement also increased as initial proficiency increases. In general, these findings demonstrated that experience with the language played a role in targetlike gender agreement. Moreover, while the results for initial proficiency were similar to those for gender assignment, the significant effects for frequency differed between determiners and adjectives. Noun frequency (individual) impacted gender assignment but noun log-frequency (language) constrained gender agreement.

Regarding adjective position, adjectives either before or after the noun in the same noun clause exhibited a higher log odds of targetlike gender agreement compared to adjectives that occurred outside of the noun clause, and there was overlap in the confidence intervals for the pre and post categories, indicating that targetlike use was similar between the two. In other words, the proximity between the noun and the adjective facilitated targetlike gender agreement.

Finally, the log odds of targetlike gender marking were higher at in-stay and post-stay compared to pre-stay and similar between in-stay and post-stay, indicating that learners' targetlike gender agreement improved during their academic year abroad and that this gain was maintained after returning home. This result was similar to the finding for gender assignment.

Thus, this multivariate analysis showed that noun ending, task, noun gender, noun log-frequency (language), initial proficiency, noun number, adjective posi-

Aarnes Gudmestad

tion, and time were the factors that influenced targetlike gender agreement for this group of additional-language learners of Spanish. Considering the assumption that gender marking on adjectives is taken to reflect gender agreement, the findings can be interpreted to indicate that learners rely on a complex array of linguistic and extra-linguistic information in order to use this morphosyntactic property (i.e., agreement) in a targetlike way in language production.

4 Conclusion

Although it may seem obvious to say that an epistemology has bearing on research findings, it does not appear to be common in SLA for researchers to try out different perspectives in order to see where they lead in terms of the interpretation of data or to make this type of work publicly available. This is precisely what I set out to do in this chapter. In this vein, the current study has offered a reflection on the relationship between epistemology and methodology through a reanalysis of production data on grammatical gender in additional-language Spanish. This reanalysis was shaped by a shift in epistemology. In my previous collaborative project (Gudmestad et al. 2019), our assumption was that gender marking, with no distinction between agreement and assignment, was the linguistic issue under investigation. In the current study, however, I adopted a different perspective, one in which gender assignment and gender agreement were different learning processes that were manifested through gender marking on determiners and adjectives, respectively (cf. Alarcón 2010; Kupisch et al. 2013). Through the reanalysis of the data in Gudmestad et al., I explored, in the current chapter, possible methodological decisions that an investigation of gender assignment and agreement in language production might entail.

Under the assumption that gender assignment and gender agreement are different processes with different surface manifestations, the results from the present investigation's analysis can be interpreted as follows. First, the higher rates of targetlike use for determiners compared to adjectives support the understanding that gender assignment is acquired before gender agreement (Alarcón 2010). Second, regarding the examination of targetlike use with individual nouns that participants use more than once, the result that some nouns exhibited variability in targetlike use with determiners may indicate that, in language production, learners show evidence of variable knowledge of gender assignment, which is counter to what some researchers have suggested (e.g., Alarcón 2010). Moreover, the evidence of variability with individual nouns in the examinations of determiners and adjectives suggested that pursuing regression analyses in order to uncover

5 *On the relationship between epistemology and methodology*

the variable patterns of use was warranted. It is worth making explicit, however, that my observations about variable use in the current analysis and the methodological decision to pursue multivariate statistical analyses were influenced by the variationist orientation of my research program more generally (cf. Young 2018). The separate mixed-effects models for determiners and adjectives have resulted in three additional observations. One was that a range of factors help to account for when learners were more likely to show evidence of targetlike gender assignment and agreement in language use. Another observation was that among the predictive factors, four impacted targetlike use on both determiners and adjectives: time, initial proficiency, noun gender, and noun ending. The epistemology that guided the present investigation may lead to expectations of finding some similarities between the two because gender assignment and agreement are related linguistic properties (i.e., they both deal with the gender of the noun). At the same time, though, the final observation that emerged from comparing the two mixed-effects models was that there were various differences in the factors impacting targetlike use between determiners and adjectives. In addition to finding that there were factors specific to each linguistic property that influenced use (determiner type and adjective position), the results also demonstrated that noun frequency (individual) only impacted gender assignment and that noun log-frequency (language) and noun number conditioned gender agreement only. Furthermore, although task was a significant constraint on both gender assignment and agreement, there were differences in the significant effects between the determiner and adjective models. These differences between the two mixed-effects models are expected, given the assumption that gender agreement and gender assignment are different learning properties that arguably have different acquisitional challenges. Thus, these multivariate analyses may be seen as bolstering to a degree the assignment-agreement assumption as they offered new details about how these learning properties differ in language use.

More generally, because the analysis in the present investigation uncovered numerous differences between assignment/determiners and agreement/adjectives, it also led to different conclusions from those drawn in Gudmestad et al. (2019), even though some of the findings are similar (e.g., the role of noun gender).⁶

⁶One example of a difference is that Gudmestad et al. (2019) found that targetlike use changed over time with regard to noun ending. However, in the current study neither the mixed-effects model for determiners nor the one for adjectives contained an interaction between time and another significant fixed effect. Another difference is that noun number impacted targetlike use with adjectives in the present investigation, but in Gudmestad et al. it was not a conditioning factor. An example of a similarity is that the two mixed-effects models in the current study and the one in Gudmestad et al. pertains to the factor of noun gender. Each regression analysis showed that the likelihood of targetlike use was higher with masculine nouns.

Aarnes Gudmestad

While it is not novel to say that analysing data differently may lead to different observations, explicit reflections on how epistemologies shape the research process are crucial. Ortega (2014: 194) explains, “by applying different theories, some findings appear to change only in the details and yet they seem to bring different “interlanguage truths” to the fore for consideration”. In a similar vein, Young (2018: 48) reflects on how applied linguists gain new knowledge and argues that “we know what we attend to and the habits of mind of researchers – their personal preferences as researchers and the early training they received – to a large extent determine the questions researchers ask, the design and implementation of research studies, and the way data are interpreted”. In sum, the current study has sought to contribute to methodological reflections in SLA by considering the important role that epistemology plays both in the analysis and interpretation of learner data and, as a consequence, in the advancement of new knowledge. Further scholarship on the connection between epistemology and methodology is important for SLA, because it demonstrates concretely the direct relationship between researchers’ (at times implicit) assumptions and the types of observations they make when interpreting learner data. There is value in making these assumptions more explicit in published research in order to illustrate concretely that knowledge is not absolute.

References

- Alarcón, Irma V. 2010. Gender assignment and agreement in L2 Spanish: The effects of morphological marking, animacy, and gender. *Studies in Hispanic and Lusophone Linguistics* 3(2). 267–299. DOI: [10.1515/shll-2010-1076](https://doi.org/10.1515/shll-2010-1076).
- Alarcón, Irma V. 2014. Grammatical gender in second language Spanish. In Kimberly L. Geeslin (ed.), *The handbook of Spanish second language acquisition*, 202–218. Malden, MA: Wiley Blackwell. DOI: [10.1002/84347.ch12](https://doi.org/10.1002/84347.ch12).
- Ayoun, Dalila. 2007. The second language acquisition of grammatical gender and agreement. In Dalila Ayoun (ed.), *French applied linguistics*, 130–170. Amsterdam: Benjamins.
- Bartning, Inge. 2000. Gender agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica* 54. 225–237.
- Bruhn de Garavito, Joyce & Lydia White. 2002. The second language acquisition of Spanish DPs: The status of grammatical features. In Ana Teresa Pérez-Leroux & Juana Muñoz Liceras (eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection*, 143–160. New York: Bilingual Press. DOI: [10.1007/978-94-010-0291-2_6](https://doi.org/10.1007/978-94-010-0291-2_6).

5 On the relationship between epistemology and methodology

- Byrnes, Heidi. 2013. Notes from the editor. *The Modern Language Journal* 97(4). 825–827.
- Creswell, John. W. & J. David Creswell. 2018. *Research design: qualitative, quantitative and mixed methods approaches*. 5th ed. Los Angeles, CA: London.
- Davies, Mark. 2016. *Corpus del Español: Two billion words, 21 countries*.
- Dewaele, Jean-Marc & Daniel Véronique. 2001. Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition* 4. 275–297.
- Ellis, Rod. 2006. Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics* 27. 431–463. DOI: [10.1093/applin/aml022](https://doi.org/10.1093/applin/aml022).
- Geeslin, Kimberly. L. & Avizia Yim Long. 2014. *Sociolinguistics and second language acquisition: learning to use language in context*. New York, NY: Routledge. DOI: [10.4324/9780203117835](https://doi.org/10.4324/9780203117835).
- Gries, Stefan T. 2005. Statistics for learner corpus research. In Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (eds.), *The Cambridge handbook of learner corpus research*, 159–181. Cambridge: Cambridge University Press.
- Group, The Douglas Fir. 2016. A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal* 100(Suppl.) 19–47. DOI: [10.1111/modl.12301](https://doi.org/10.1111/modl.12301).
- Grüter, Theres, Casey Lew-Williams & Anne Fernald. 2012. Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research* 28(2). 191–215.
- Gudmestad, Aarnes, Amanda Edmonds & Thomas Metzger. 2019. *Using variationism and learner corpus research to investigate grammatical gender marking in additional-language Spanish*. Language Learning.
- Gudmundson, Anna. 2013. Acquisition of grammatical gender and number agreement in Swedish learners of L2 Italian: Regularity and frequency effects. *Language, Interaction and Acquisition* 4(2). 232–255.
- Halberstadt, Lauren, Jorge R. Valdés Kroff & Paola E. Dussias. 2018. Grammatical gender processing in L2 speakers of Spanish: The role of cognate status and gender transparency. *Journal of Second Language Studies* 1(1). 5–30.
- Kupisch, Tanja, Deniz Akpınar & Antje Stöhr. 2013. Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism* 3(2). 150–179. DOI: [10.1075/lab.3.2.02kup](https://doi.org/10.1075/lab.3.2.02kup).
- Larsen-Hall, Jenifer & Luke Plonsky. 2015. Vol. 65. Reporting & interpreting quantitative research findings: What gets reported & recommendations for the field. 127–159.

Aarnes Gudmestad

- López Prego, Beatriz & Alison Gabriele. 2012. Examining the nature of variability in gender and number agreement in native and non-native Spanish. In Alia K. Biller, Esther Y. Chung & Amelia E. Kimball (eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development*, 315–327. Somerville, MA: Cascadilla Press.
- Marsden, Emma, Kara Morgan-Short, Pavel Trofimovich & Nick Ellis. 2018. Introducing registered reports at language learning: Promoting transparency, replication, and a synthetic ethic in the language sciences [editorial]. *Language Learning* 68. 309–320.
- Marsden, Emma & Luke Plonsky. 2018. Conclusion: Data, open science, and methodological reform in second language acquisition research. In Aarnes Gudmestad & Amanda Edmonds (eds.), *Critical reflections on data in second language acquisition*, 219–228. Amsterdam: John Benjamins. DOI: [10.1075/llt.51.10mar](https://doi.org/10.1075/llt.51.10mar).
- Mitchell, Rosamond, Nicole Tracy-Ventura & Kevin McManus. 2017. *Anglophone students abroad: Identity, social relationships, and language learning*. New York, NY: Routledge.
- Montrul, Silvina, Rebecca Foote & Silvia Perpiñán. 2008. Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning* 58. 503–553. DOI: [10.1111/j.1467-9922.2008.00449.x](https://doi.org/10.1111/j.1467-9922.2008.00449.x).
- Ortega, Lourdes. 2005. Methodology, epistemology, and ethics in instructed SLA research: An introduction. *The Modern Language Journal* 89. 317–327.
- Ortega, Lourdes. 2014. Trying out theories on interlanguage: Description and explanation over 40 years of L2 negation research. In ZhaoHong Han & Elaine Tarone (eds.), *Interlanguage: Forty years later*, 173–201. Amsterdam: John Benjamins.
- Applied linguistics research: Current issues, methods and trends. 2018. In Aek Phakiti, Peter De Costa, Luke Plonsky & Sue Starfield (eds.), *The palgrave handbook of applied linguistics research methodology*, 5–29. New York, NY: Palgrave.
- Plonsky, Luke (ed.). 2015. *Advancing quantitative methods in second language acquisition*. New York, NY: Routledge.
- Plonsky, Luke & Frederick L. Oswald. 2017. Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition* 39. 579–592.
- Porte, Graeme & Kevin McManus. 2018. *Doing replication research in applied linguistics*. New York, NY: Routledge.
- Smith, Thomas J. & Cornelius M. McKenna. 2013. A comparison of logistic regression pseudo r^2 indices. *Multiple Linear Regression Viewpoints* 39(2). 17–26.

5 *On the relationship between epistemology and methodology: A reanalysis of grammatical gender in additional-language Spanish*

- Team, R Core. 2017. *R: A language and environment for statistical computing*. Vienna.
- Teschner, V., Richard & William. M. Russell. 1984. The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics* 1. 115–132.
- Young, Richard F. 2018. Habits of mind: How do we know what we know? In Aek Phakiti, Peter De Costa, Luke Plonsky & Sue Starfield (eds.), *The Palgrave handbook of applied linguistics research methodology*, 31–53. New York: Palgrave Macmillan.

Chapter 6

Analysing interaction in primary school language classes: Multilevel annotation and analysis with EXMARaLDA

Heather E. Hilton

Université Lumière Lyon 2

John Osborne

Université Savoie Mont Blanc

Language classrooms provide a rich terrain for language acquisition research, and classroom observation has a long history (Passy 1885; Brebner 1898), including numerous studies of interaction within the foreign language class in the latter part of the 20th century (Véronique 1992). This interest has resulted in a considerable set of transcribing conventions and observation grids (Sinclair & Coulthard 1975; Moskowitz 1976; Allen et al. 1983; Ullmann & Geva 1984), but the analytical techniques have varied little since the initial conversation analyses of the 1970s and 1980s: transcription is often done without the aid of dedicated software and analyses are carried out by hand.

As part of an exploratory study of elementary school foreign-language learning, a French research team observed two classes (25 pupils in Year 1, 30 in Year 3) during their first year of beginning-level English lessons. Observation took place at three different times during the school year, with a total of three weeks of language lessons filmed in each classroom. This chapter presents the methodology adopted for transcribing and annotating the lessons using EXMARaLDA (Schmidt & Wörner 2014), a tool which facilitates objective comparison of the two learning environments, and multi-level analyses. Our initial analyses illustrate the ways in which well-designed transcription software can contribute to an understanding of methodological and interactional classroom variables, and how they may affect emergent language knowledge and skill in the classroom. Video-linked transcription and multi-tiered annotation in EXMARaLDA can enable automatic and semi-automatic analyses of various aspects of the classroom experience: talking time,



Heather E. Hilton & John Osborne

activity types, L1 and L2 use, teacher-learner and learner-learner interactions, metalinguistic input, visual input, manipulation of objects, and other actions. Our analyses will compare the two classrooms, and explore features of these young learners' initial contact with new words and their semantic-grammatical properties.

early language learning, classroom interaction, transcription methodology, teaching methodology

1 Introduction

Classroom observation has a long history in the context of language teaching methodology and teacher training (see, for example, [Passy 1885](#); [Brebner 1898](#)); and the language classroom became a valued research context for second language acquisition (SLA) and interaction research in the late 1960s ([Moskowitz 1976](#); [Jarvis 1968](#); [Wragg 1970](#); [Seliger & Long 1983](#); [Allwright 1984](#); [Véronique 1992](#)). Researchers were justifiably interested in observable factors that might influence the emergence of new language forms and structures, in learners of different ages and backgrounds. Since these early studies, interest in classroom interaction has remained steady, with particular attention paid to the interactions between learners as they work together in groups, or in computer-mediated “tandem” situations (for example, [Develotte et al. 2008](#)). The authors of this paper are newcomers to interaction research, having previously carried out work on native-speaker and learner corpora generated through monological or guided tasks. Our previous transcription experience had been with the CHILDES suite of software ([MacWhinney 2000](#)) and the associated PHON software ([Rose et al. 2006](#)); we are firmly committed to Brian MacWhinney's paradigm-changing stance on the need for shared data in language acquisition research ([MacWhinney 2010](#): 27-30).

With the objective, therefore, of taking a data-driven and data-sharing approach to the analysis of classroom interaction, this chapter will present our analyses of English as a Foreign Language (EFL) lessons filmed in two primary schools in France. The rationale for choosing the EXMARaLDA software ([Schmidt & Wörner 2014](#)) will be explained, as well as our transcription and annotation system. In the last two sections of the chapter, we will illustrate the types of analyses that can be carried out with a well-designed tool, and consider the potential of such analyses for research in second language acquisition and teaching.

2 Classroom interaction research (theoretical and methodological issues)

Early observations of classroom interaction (Flanders 1970; Brown 1975; Moskowitz 1976; Bowers 1980; Allen et al. 1983; Ullmann & Geva 1982; 1984) led to various means of representing what was going on in the classroom, often using tables or checklists completed by hand in real time. These observation methods were generally developed for pedagogical purposes such as teacher training rather than being part of a concerted research programme, and resulted in a spate of introductory texts for language teachers at the end of the 1980s (Allwright 1988; Chaudron 1988; van Lier 1988; Nunan 1989).

A notable exception, giving more attention to linguistic and pragmatic aspects of classroom exchanges, is the system developed by Sinclair & Coulthard (1975) for describing the structure of classroom discourse. In one of the most detailed early studies of interaction in the EFL classroom, Willis (1981) used a modified version of Sinclair & Coulthard's system to analyse a corpus of tape-recorded lessons. The recordings were made with a double-track machine, with one microphone for the teacher and one for the learners. Non-linguistic and inaudible features were hand-noted along with the time-counter position on the tape-recorder. These data were then transcribed by hand in a multi-column format, indicating the structure of exchanges and the type of act. This resulted in a total of 27 categories, based on Sinclair & Coulthard's initial inventory of: marker, starter, elicitation, check, directive, informative, prompt, clue, cue, bid, nomination, acknowledgement, reply, react, comment, accept, evaluate, metastatement, conclusion, loop and aside. The descriptive categories developed in this framework, either as originally defined by Sinclair & Coulthard or in a modified form, have subsequently been used in a large number of analyses, notably those deriving from the postgraduate programme in Teaching English as a Foreign Language at Birmingham University, but also by researchers elsewhere (Chaudron 1977; Grandcolas & Soulé-Susbielles 1986; Chapelle 1990). They have been partly adapted for the present study.

Classroom interaction studies have also drawn on the techniques used in Conversation Analysis (CA) for describing naturally occurring speech. These include accounting for such things as turn-taking organization (Sacks et al. 1974), repair (Schegloff et al. 1977; Schegloff 2000), the cooperative nature of "side sequences" (Jefferson 1972) or discourse as "interactional achievement" (Schegloff 1982), but also defining conventions for detailed transcription of interactions (Jefferson 2004), including those where participants have non-targetlike discourse characteristics, as in the case of children (Ochs 1979) or L2 speakers (Jefferson

Heather E. Hilton & John Osborne

1983; 1996). In Second Language Acquisition (SLA) research, interest in using CA techniques was stimulated partly by criticism that SLA reflected an imbalance in favour of individual cognition at the expense of interactional and sociolinguistic orientations to language (Firth & Wagner 1997). Whether or not one subscribes to Firth & Wagner's arguments (for responses see Kasper 1997; Poullisse 1997; Long 1997; Gass 1998) there is no doubt that they triggered interest in applying CA to various kinds of SLA data (Markee 2000; Seedhouse 2004; 2005) and in exploring the "intersection" between CA, SLA and language pedagogy (Mori 2007). A more recent development is the convergence between CA methodology and complexity theory to investigate the ways in which L2 classroom interaction displays characteristics of a complex adaptive system (Seedhouse 2010; 2015).

With the advent of video recording it became no longer necessary to use check-lists and annotations in real time since in principle all the data could be retrieved at leisure from the recording. However, as well as being more intrusive, filming necessarily imposes a frame on what is actually captured by the camera, and the subsequent transcription introduces a further filter depending on the transcription format and on what the transcriber chooses to pay attention to. Transcription is a selective process (Ochs 1979) and the transcript itself is an evolving flexible object (Mondada 2007). Researchers often continued to use a play-script format ill-adapted to the complexity of video data (see Erickson 2004). As Jones (2013: 17) notes, "[t]he problem with most early work using video was that technologies of transcription had not yet caught up with the technologies of recording." Dedicated software for multimodal transcription such as ELAN, EXMARaLDA or ANVIL (section 3.3 below) frees the transcriber from the constraints of a page format, facilitates the representation of overlaps, simultaneous events or non-linguistic features, and enables transcription segments to be time-linked to the digital recording. However, the raw data of the recording still have to go through a process of "entextualization" (Bauman & Briggs 1990) in order to be fully searchable, and it is the decisions made at this stage that will determine which aspects of the data can subsequently be retrieved for analysis.

Searchability is an important condition, both for quantifying chosen features of classroom interaction and for examining elements that may be dispersed throughout a lesson. Studies of interaction in SLA, inside and outside the classroom, have often followed a path suggested by Jefferson's (1972) notion of "side sequences" and have focused on instances of communicational problem solving and negotiation that are thought to have a potential for triggering acquisition, following work by de Pietro et al. (1989); Vasseur (1989) and Bange (1992). The methodology of these studies consists largely of micro-analyses of interactions (Pekarek Doehler 2000: 7) and although it is emphasised that behind these analyses lies an

6 Analysing interaction in primary school language classes

entire corpus (Arditty & Vasseur 2005: 3), the data presented for discussion consist essentially of selected extracts. This is fair enough within a given research perspective – interested specifically, say, in interactional shifts of focus between communication and the means of communication – but it is also possible to adopt a more ‘corpus-driven’ approach to classroom data, in which analysis is bottom-up and data-driven (Tognini-Bonelli 2001; Seedhouse 2005), with the aim of capturing patterns and associations that may not have been expected at the outset.

3 Methodology of the current study

3.1 The context

The *Seine & Marne Primary* project is an exploratory study by a multidisciplinary team of researchers, which was implemented between 2012 and 2015 in two public elementary schools in the *Seine & Marne département* west of Paris, France. During the 2012–2013 school year, two classrooms of beginning English were filmed at three intervals: early December, mid-February, and mid-May. In one classroom, 25 children in their first year of primary school (15 girls and 10 boys, all born in 2006) were six years old at the time of filming; in the other, 29 Year 3 learners (16 girls and 13 boys, all born in 2004), were eight years old at the time of the study¹. The two classrooms are in adjacent villages (two kilometres apart), part of a *regroupement scolaire*, or closely-linked network of rural schools, where the socio-economic composition of the classes is basically identical. Seven of the 54 children participating in the study are bilingual (users of a language other than French in their daily lives), according to language profile information provided by the parents (three children in Year 1, four in Year 3).

The institutional context of the *Seine & Marne Primary* study was a 2012 ruling by the French Ministry of Education to move obligatory foreign language tuition into the first year of elementary school, despite problems of in-service and even initial teacher training for this aspect of the elementary curriculum (Young & Mary 2010; Mary 2014). The French education system is highly centralized, with a national curriculum, so a common Communicative task-based methodology (of Europe 2001) is used in both classrooms: the syllabus is functionally-organized (greeting, asking and telling your name or age, expressing likes and dislikes, etc.), and classroom activities include small group work, interactional

¹The research team was able to take advantage of a change in the national curriculum for foreign languages, which lowered the starting point for L2 study in 2011, and enabled this comparison of children starting English at age six, and age eight in the same school system during the 2012–13 school year.

Heather E. Hilton & John Osborne

role plays, games and tasks (e.g. cooking, planting seeds). Both teachers use real objects and pictures to illustrate meaning (of new words, especially), and puppets to help trigger functional language. The Year 1 teacher created all of her support materials, and used storybooks at the end of each lesson; the Year 3 teacher based her lessons on a commercially-available textbook, with an increasing number of self-designed activities throughout the year but no use of storybooks. Both the first- and third-year groups had 80 to 90 minutes of English per week, in keeping with the national curriculum, although this total was distributed differently throughout the week, with more frequent, shorter sessions in Year 1, and two 45-minute sessions in Year 3. A one-hour interview with each teacher early in the year revealed a key institutional variable: both teachers are highly confident and competent professionals (displaying detailed knowledge of the curriculum, their learners and family attitudes towards language-learning, as well an advanced level of methodological analysis), but they possess very different levels of *linguistic* confidence, in relation to their obligatory English teaching. The Year 1 teacher majored in English at university, lived two years in the United States, and declared herself to be very comfortable (*très à l'aise*) with the foreign-language part of her curriculum. The Year 3 teacher majored in Economics, and volunteered to participate in the project precisely because she wanted help with her English lessons, feeling quite uncomfortable with her knowledge of the language (*tellement pas à l'aise*), in particular of English pronunciation and grammar. Section 4 below will discuss whether this difference in linguistic confidence may be reflected in the methodological or linguistic characteristics of each teacher's pedagogical approach and, as a consequence, in general classroom organization.

3.2 Overall study design and research questions

In the context of newly-imposed foreign language lessons in Year 1, the objectives of the *Seine & Marne Primary Project* were wide-ranging and exploratory, attempting to answer research questions as varied as: What sort of language-teaching methodology is used in primary English classrooms in France? Are there differences in the methodology used with six-year-olds and with eight-year olds, and if so, of what types? Do six- and eight-year-olds follow similar learning trajectories in the FL classroom, or are there fundamental differences? What role do individual variables (such as first-language knowledge, personality, cognitive capacity, motivation) play in the learning pathways observed?

In order to answer these questions, three data sets were collected: the video corpus of fourteen filmed lessons, the children's performance on a series of English tasks (measuring emergent knowledge and skill, and administered twice

6 *Analysing interaction in primary school language classes*

during the 2012-2013 school year), and their scores on a battery of psychometric measures of cognitive and social characteristics. The linguistic and psychometric data have been discussed in other studies (Hilton & Royer 2014; Hilton et al. 2016; Hilton 2017). In this chapter we will be presenting the methodology used to transcribe and annotate the filmed language lessons, as well as the types of classroom analyses that such an approach makes possible. The precise questions for this study are as follows:

To what extent can carefully transcribed and annotated language lessons shed light on classroom foreign-language teaching and learning processes?

1. Which automatic analyses (enabled by the transcription software) are revealing of classroom interactions?
2. Which additional analyses can add information on classroom interaction structures?
3. What types of conclusions for language-learning and teaching research can be drawn from such analyses?

To film the fourteen lessons comprising our video corpus, a single Canon Xf100 video camera was used (at times fixed to a tripod, at times roving and zooming in on the children); the teachers wore a cordless lapel microphone during the lessons, and a boom-held microphone (which could follow the sound around the room, for example during group work) was also used for an optimal sound feed. The corpus of filmed lessons was assembled in order to gather information on lesson content, and a concrete sample of the types of classroom activities used, for a more complete picture of the children's learning environment. The data obtained are very sparse (six to eight lessons out of an annual program), and the use of a single camera (focusing alternatively on the teacher or the learners) means that the footage obtained cannot be used for detailed observation of the classroom behaviour of each child. For this methodologically-oriented chapter, we will focus our analyses here on the two most similar lessons from our classroom corpus: both occurring in mid-February, and both devoted to the presentation and practice of new food vocabulary in a unit devoted to talking about food, preferences, and cooking.

3.3 Choice of transcription software

When choosing a transcription tool, the guiding criterion should be the fit between the objectives of the study and what the transcription software is designed

Heather E. Hilton & John Osborne

to do. To transcribe and analyse the lessons in the *Seine& Marne* project, our primary focus was classroom interaction and methodology: teacher talking time, learner talking time, language use (L2 English or L1 French), the linguistic content of teacher and learner productions, the different types of interactions between teacher and learner(s), the typology of classroom activities and teaching techniques. It therefore seemed logical to adopt software designed precisely for the annotation of interactive discourse.

Several freely available scientific tools can be used, most notably CLAN (MacWhinney 2000), ANVIL (Kipp 2014), ELAN (Wittenburg et al. 2006) and EXMARaLDA (Schmidt & Wörner 2014). Others, such as the *Digital Replay System* (Brundell et al. 2008), offer interesting features but are no longer supported. The most apparent difference between tools, from the transcriber's point of view, is whether they display discourse as a list of turns (as in CLAN) or in tiers of timelines (as in ELAN or EXMARaLDA). According to the purpose for which they have been designed – CLAN for morpho-syntactic and lexical analyses in child language, PHON for finer phonological transcriptions of emergent child speech, ELAN for easier coding of non-verbal phenomena in linguistic studies (Lausberg & Sloetjes 2009) – each tool has slightly different ways of setting up transcription tiers and incorporates different sub-programmes for segmentation and automatic pause recognition, concordancing, morpho-syntactic tagging, lexical analysis, etc. With the most widely-used tools, import and export from and to the other formats is possible, if not always lossless, so choice of a particular tool does not lock the user irrevocably into that environment. For an extensive overview of tools for multimodal annotation and interactional analysis see Cassidy & Schmidt (2017) and Glüer (2018).

The choice of EXMARaLDA, developed by Thomas Schmidt, Kai Wörner, Timm Lehmberg and Hanna Hedeland at the *Zentrum für Sprachkorpora* at Hamburg University (Schmidt & Wörner 2014), was determined by its tier-timeline format (more suitable for classroom interaction than a list of turns), by its ability to handle multi-level concordancing through the built-in EXAKT tool, and by what appeared, after initial tests with ELAN, to be a more flexible way of setting up transcription and annotation tiers for multiple speakers. However, we have no reason to suppose that the annotations and analyses discussed here could not also be carried out with ELAN, with a slightly different procedure for determining tier types. As with any multimedia software, ability to read video formats can be an issue. EXMARaLDA recommends mpeg4 for video files and wav for audio, and incorporates four different media players to choose from for the best compatibility.

3.4 *Seine & Marne* transcription architecture and annotation conventions

Because of the use of a single camera for filming the *Seine & Marne Primary* English lessons, we limited our transcription architecture to three primary speaker transcription tiers: Teacher, Learner-groups, and Individual learner. The “Learner-groups” line was used to transcribe any learner productions involving two or more children (with specifications concerning group size in a dependent tier); the “Individual learner” line was used to transcribe any production by an individual child (again, with individual learner characteristics (sex, project ID number) given in a dependent tier, whenever possible, our single-camera installation making it impossible to identify the source of every utterance. In order to facilitate automatic word-counts in EXMARaLDA and subsequent concordancing operations, we created two transcription lines for each speaker, one for productions in L2 English, and one for productions in L1 French. Three dependent tiers are attached to each speaker: one for the fixed coding of the interactional function of each transcribed segment, using the set of codes described in Table 1, an open tier for annotating any relevant or salient actions, and a tier for fixed coding of linguistic errors (selection from the list of codes). Additional independent tiers are used to annotate activity types, lesson plan structure, and support material used; a “comments” line enables the transcriber to note anything else of interest. A particularly transcriber-friendly aspect of EXMARaLDA is the possibility of formatting the transcription lines with colour-coding, for example, so that all tiers linked to the same speaker have the same background colour.

Classroom speech turns are lexically (and not phonologically) transcribed for each speaker, using a simplified version of basic CHAT transcription conventions (Codes for the Human Analysis of Transcripts, MacWhinney 2000), which the *Seine & Marne* research team had already used extensively. An example of the transcription output from EXMARaLDA is shown in Figure 1. The transcription format uses the following basic units:

An *interval* is a portion of the time-line in EXMARaLDA, and is typically the duration occupied by a single consecutive event (see below). Intervals are numbered consecutively from the beginning to the end of the recording, as shown in the top bar in Figure 1.

An *event* is a portion of the transcription, and can be either a speech event, containing speech by one of the speakers, or a classroom event, corresponding to an action with or without transcribed speech attached to it. Thus in Figure 1, there are five speech events (respectively “oh!”, “now”, “look and listen very carefully”, “okay?” and “okay”), and four classroom events. One of these (“gesturing...”) ac-

Heather E. Hilton & John Osborne

companies a speech event by the same participant, one (“changing file”) coincides with a speech event by other participants, and two are unaccompanied by any speech (“returning to seat” and “pointing to picture”).

An *utterance* can consist of a single word, a verbless phrase or a main clause with any of its dependent clauses. Typically, an utterance will correspond to a speech event, but utterances containing more than one interactional function (Table 1, below) are broken down according to these functions. For example, “Martin | sit down please” counts as two EXMARaLDA speech events, the first one nominative, the second directive.

A *segment chain* is an interrupted string of speech by one speaker (i.e. a speech turn), and can consist of one or more utterances. The “output” command in EXMARaLDA can be used to generate the transcription as a list of turns, with each segment chain on a separate line.

The EXMARaLDA Annotation Panel was used to simplify the coding of the interactional function of each segment, with a fixed set of codes based on Willis’ (1981) modified version of Sinclair & Coulthard’s (1975) interaction typology (see Section 2, above). We pared the system down further, to correspond to the particular types of interaction found in these beginning-level primary classrooms. Table 1 presents the 29 codes used to annotate interactional behaviours in our corpus, grouped in eleven interactional functions. The Annotation Panel enables the transcriber to insert the relevant code on the interaction tier for each speech event or interval, without typing it out each time.

Linguistic errors were coded according to a minimalistic version of the error codes used in the PAROLE corpus (Hilton 2008), which are based on error codes established for CHAT. An “error” is any divergence from expected forms in pronunciation, morphology, syntax or lexis (and no value judgment is placed on the use of this term, of course).

Figure 1 shows the partition-formatted *html* output for seven intervals in the Year 3 lesson: transcription tiers are indicated in black headings, and the dependent tiers in light grey; the partition illustrates the use of colours to link annotations to the relevant speaker.

This easily-obtained output format is useful for checking transcriptions (only tiers containing transcription or annotation are shown in each partition), and for subsequent qualitative analysis.

4 Analyses and preliminary findings

Once the transcriptions are finalised, it is possible to carry out a number of analyses automatically – with pre-programmed functions in EXMARaLDA (section

6 *Analysing interaction in primary school language classes*

Table 1: Coding system used to annotate classroom interactions

	interaction type, codes	Explanation
1	meta metaPRE meta MID metaPOST metaCHECK	comment on lesson structure, activity, language PRE: prior to activity or lesson (statement of objectives; beginning of new lesson phase) MID: explaining, commenting on something as it is happening (during activity) POST: summing up & analyzing; marking end of episode/ activity CHECK: checking comprehension, uptake
2	aside	elements of discourse not intended to elicit response or reaction (teacher “thinking out loud”, talking to self or film crew...)
3	directive	request nonlinguistic response; classroom management
4	presentation presLANG presCULT presMOD presINFO	declarative statement, presentation: of sound, word, structure (even question) of cultural information of model for student production/ repetition of general information (about something other than above)
5	elicitation elicitS elicitQ elicitNV elicitPROMPT elicit_REP elicit_CORR	request/elicit response with a statement with a question non-verbal elicitation reiterate stalled elicitation, give a clue, etc. elicit a repetition elicit a correction
6	nomination	call on a learner, give permission to respond
7	acknowledgement ack ackNEG	acknowledge a response, accept a bid positive acknowledgement, indicate appropriate response indicate that response or reaction is inappropriate/incorrect
8	response resp respANS respN respREP respRFORM respCORR	provide linguistic response to elicitation or prompt (any response not in the following categories) answer (in response to a question) naming (a picture or object, in response to non-verbal elicitation) repetition (repeating elicitation, with no change) reformulation (of preceding response; one or more changes; does not have to be by same speaker) provide a correction (does not have to be by same speaker)
9	production	spontaneous production (does not follow elicitation)
10	reaction	nonlinguistic response to directive or elicitation
11	bid	signal a desire to contribute

Heather E. Hilton & John Osborne

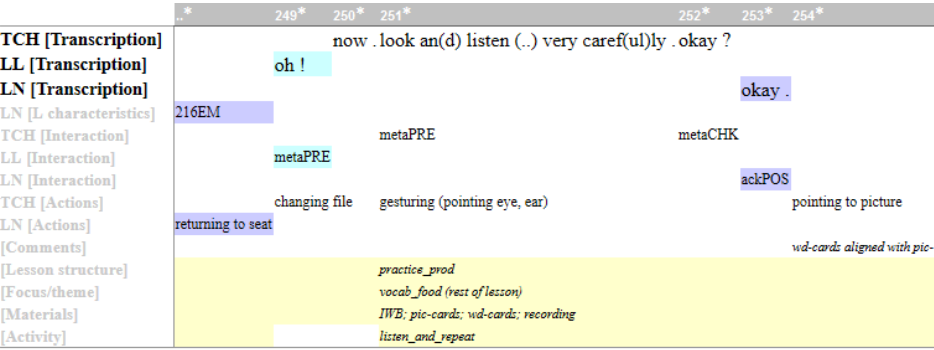


Figure 1: Partition-format output of finished transcription

4.1) – and semi-automatically – using the EXMARaLDA concordancing software, EXAKT (section 4.2). It is also easy to export EXMARaLDA files into a format enabling the use of the many powerful language-analysis programmes included in CLAN, but we will not have space to present these here.

4.1 Classroom comparisons through automatic analyses

Our first automatic tally concerns the number of transcribed ‘segments’: a segment (more specifically, in EXMARaLDA terminology, a ‘segment chain’) corresponds to a speech turn, or uninterrupted string of speech by one speaker, which may contain more than one utterance. For example, Figure 1 shows one segment by the teacher (“now. look an(d) listen (...) very caref(ul)ly. okay?”) covering intervals 250-252, bounded on either side by a learner segment (“oh” and “okay”, respectively). Speaker-specific segment counts are obtained with a single click in EXMARaLDA, and are presented in Table 2, below. These figures illustrate the intense interactional nature of the beginning language classroom, with around twenty speech turns per minute in both classrooms – that is, one every three seconds.

As with most audio- or video-linked transcription software, it is easy in EXMARaLDA to obtain a calculation of speaking time for each of the transcription tiers: in other words (for the transcription architecture presented here) total teacher speaking time, and total learner speaking time, subdivided into learner-group and individual-learner speaking time. Figure 2 illustrates the distribution of talking time in our two target lessons, with slightly more teacher talking time (the darker sectors on the right of each pie chart) in the Year 1 classroom, and both teachers (plus 0.4% pre-recorded sound files in Year 3) occupying about half of

6 *Analysing interaction in primary school language classes*

Table 2: Number of segments for each speaker (columns) per classroom (lines)

	teacher (+ record- ings)	learner- group	individual learner	Total	speaking turns per minute
Year 1 ^a	369	306	188	863	21
Year 3 ^b	404	223	239	866	20

^a41-minute lesson.

^b43-minute lesson. The Year 3 lesson includes ten pre-recorded one-word utterances.

the lesson time. The charts illustrate an interesting difference in learner participation in the two classrooms, with the Year 1 teacher eliciting more learner-group productions, and the Year 3 classroom characterised by more frequent individual learner productions.

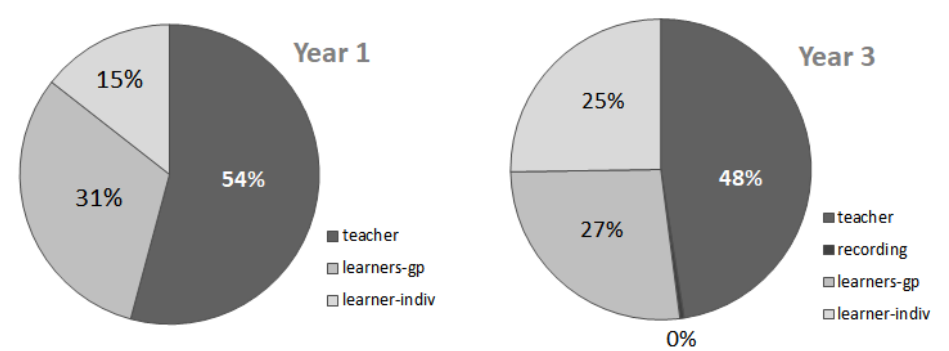


Figure 2: : Distribution of classroom talking time

Our separate transcription lines for speech turns in L2 English or L1 French enable an automatic breakdown of the numbers of segments and words produced in each of the classroom languages; Figure 3 gives a graphic presentation of the distribution of language use, based on the numbers of segments produced in each language. In both classrooms, at least 80% of the teacher’s output is in English, with 92% for the linguistically-confident Year 1 teacher; there is a higher percentage of English use overall in her classroom, particularly in the learners’ output. The unpruned data presented here includes asides and comments by the learners in L1 French; in both classrooms the inclusion of a food-tasting activity generated much excitement and a certain amount of L1 commentary. Both teachers

Heather E. Hilton & John Osborne

carried out a metalinguistic wrap-up in French at the end of the lesson, and the Year 3 learners also had a short cultural discussion in French early in the lesson.

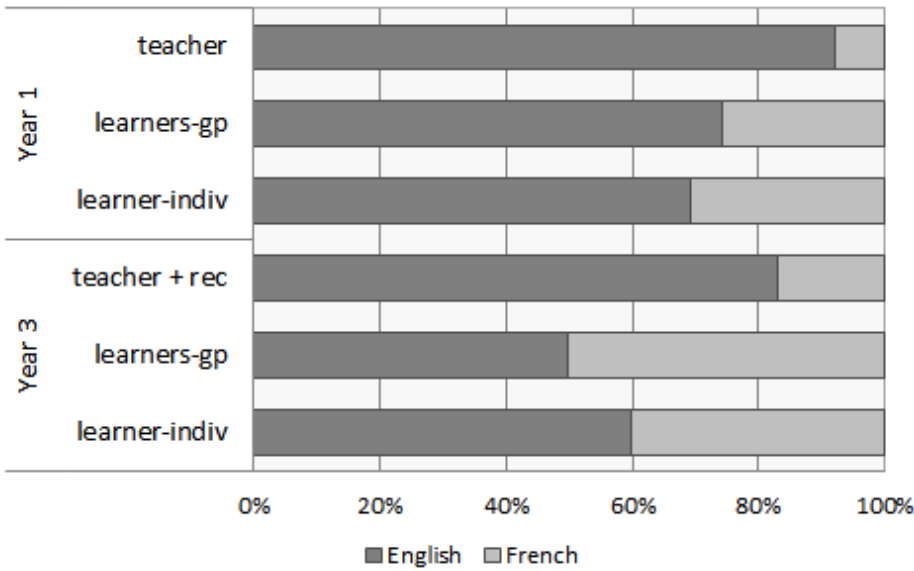


Figure 3: : Use of L2 English and L1 French by classroom and speaker

Before running the transcriptions through EXMARaLDA's EXAKT concordancing software, we can generate an unpruned word count, which enables a final automatic comparison between the two classrooms. Table 3 shows additional characteristics of the two classrooms, with longer utterances in L1 French (unsurprisingly) in both classrooms, but also some interesting linguistic differences between the two classrooms (highlighted in grey). Learners in the Year 1 classroom heard almost twice as many English words during their 41-minute lesson as the Year 3 learners; despite their younger age, their English input also consists of longer L2 utterances (over six words per segment on average). The younger learners' English output is also greater (in number of words produced), with slightly longer utterances in choral/group productions. The Year 3 learners produce more individual segments in L1 French than they do in English; many of these are comments on the new vocabulary words, or on the food-tasting activity

These analyses, derived from a one-click count of transcription segments and timeline intervals already point to linguistic and methodological differences between the Year 1 and Year 3 lessons. In the next section we will look more closely at the methodological, interactional, and linguistic characteristics of each class-

6 *Analysing interaction in primary school language classes*

Table 3: Numbers of words produced (unpruned tokens), by speaker and class

		Year 1		Year 3	
		tokens	wps ^a	tokens	wps
L2 English	teacher	2266	6.7	1206	3.7
	recording	–	–	10	1.0
	learner groups	458	2.0	159	1.4
	indiv learner	272	2.1	299	2.1
L1 French	teacher	401	13.8	703	10.3
	learner groups	129		150	
	indiv learner	157		462	

^awords per segment

room, using various concordancing options included in the EXMARaLDA package.

4.2 Classroom comparisons through further analysis

Using EXMARaLDA’s EXAKT concordancing programme, it is possible to pursue the comparisons between our two primary English classrooms. EXAKT enables the researcher to tally annotation codes on the dependent tiers according to speaker, to look at (and compare) the linguistic environment of target words or forms, and to carry out multi-tier analyses, combining a key word search on the transcription lines with information from the annotation tiers.

To compare the interactional patterns in our two lessons, we performed a simple count of the annotation codes on the ‘interaction’ coding tiers, after filtering out the L1 transcription lines. Results are given here for patterns directly related to teacher-learner interaction: directives, modelling, elicitation, response and acknowledgement. Asides and metastatements, often in French, are not included. Figure 4 shows interaction types in English in Year 1 and Year 3 for the teachers; Figure 5 compares learner interaction in the two classrooms (where both individual and learner-group interactions have been combined). See Table 1 above for an explanation of the interaction codes that are featured on the left of each chart.

The primary interactional difference between the two classes lies in the number of directives and of models for student production or repetition (coded pres-MOD) produced by the Year 1 teacher (Figure 4), and the (correspondingly) high

Heather E. Hilton & John Osborne

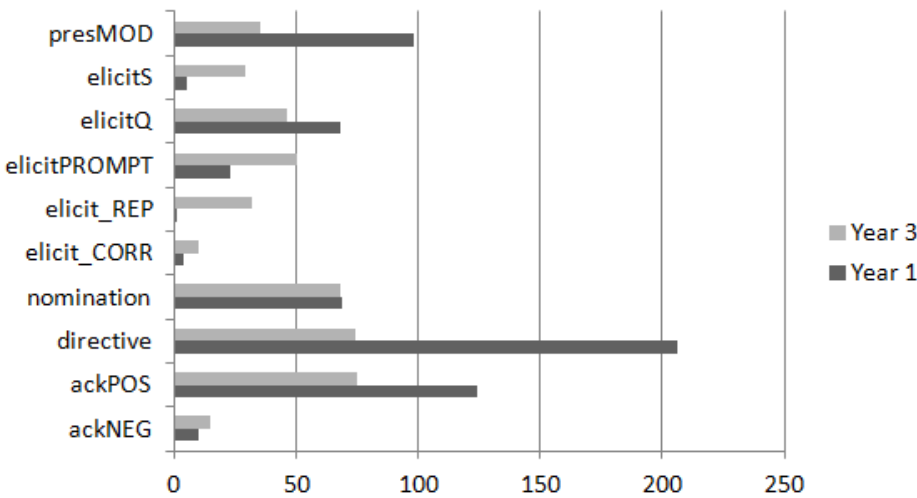


Figure 4: Teacher L2 interaction types, by classroom

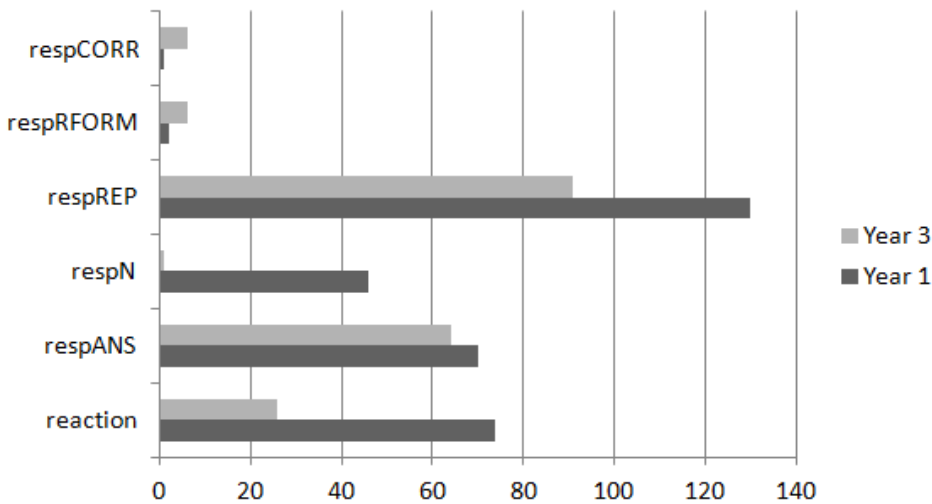


Figure 5: Learner L2 interaction types, by classroom

6 Analysing interaction in primary school language classes

proportion of naming responses (respN) produced by the Year 1 learners (Figure 5), as well as more repetition (respREP). In turn, these trigger a greater number of positive acknowledgements from the teacher.

The concordancing functions of EXAKT enable us to take a closer look at the linguistic contexts in which the new words appear in the two classrooms, and to compare the ways in which the two groups of learners structure utterances containing them. As the two lessons under examination here had food as the main topic, we are going to focus here on words related to this semantic domain.

The list of L1 and L2 food words occurring in the two lessons, obtained from a simple word-count in EXAKT, is as follows:

Year 1:

L2 words: *apple, banana, chicken, chocolate, egg, fish, ice-cream, orange, pea, pear, soup, spaghetti, string-beans*;

L2 associated words which do not designate types of food: *food, eat, hungry, mouth*;

L1 words: *chocolat; assiette, déjeuner, nourriture, aliments*.

Year 3:

L2 words: *butter, cake, egg(s), flour, lemon, milk, pancake(s), sugar*;

L2 associated words: *eat, taste*;

L1 words: *beurre, citron, crêpe(s), farine, lait, oeufs, sucre; casserole, cuisiner, goûter, ingrédients, recette*.

In the Year 3 class, each of the L2 food words, with the exception of *cake* (which occurs only once) is accompanied by an L1 equivalent, whereas in the Year 1 class, this is the case for only one word, *chocolate*. In both classes, occurrences of the target vocabulary items are distributed throughout the lesson, but in slightly different ways, as shown in Figures 6 and 7. The numbers on the horizontal axis indicate the position of occurrences during the lesson, with reference to the interval on the time-line in which they appear (in total, 1200 to 1300 intervals per 40-minute lesson, shown on the X axis in Figures 6 and 7).

In the Year 1 class (Figure 6) there is an intensive repetition of all the target vocabulary in the first third of the lesson, followed by re-use of all the items except *egg* towards the end of the lesson. In the Year 3 class (Figure 7), there are shorter bursts of repetition for some of the words – *pancake, egg, flour, milk* and *butter* – at slightly different moments, but otherwise the words are more randomly distributed between the beginning and the end of the lesson.

Appropriate use of food words in a complete English noun phrase depends partly on an appreciation of the mass-count distinction, since many foods can

Heather E. Hilton & John Osborne

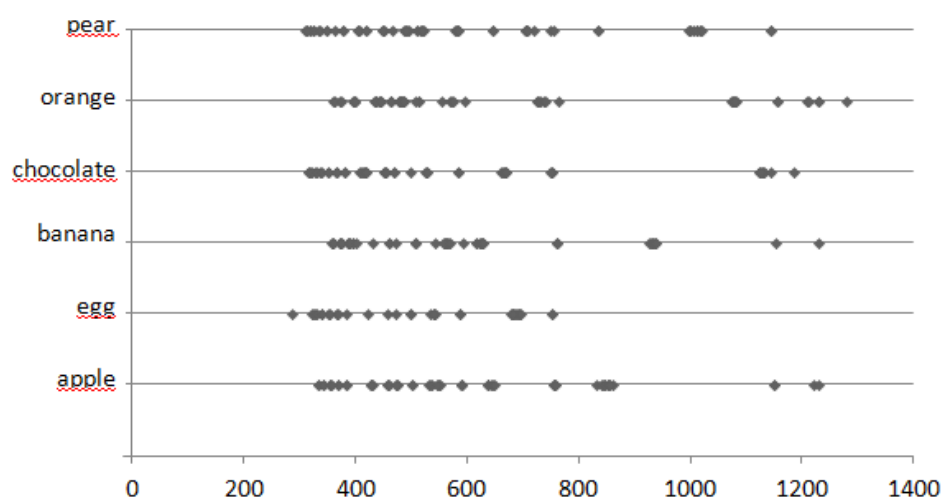


Figure 6: : Distribution of food words in Year 1 class

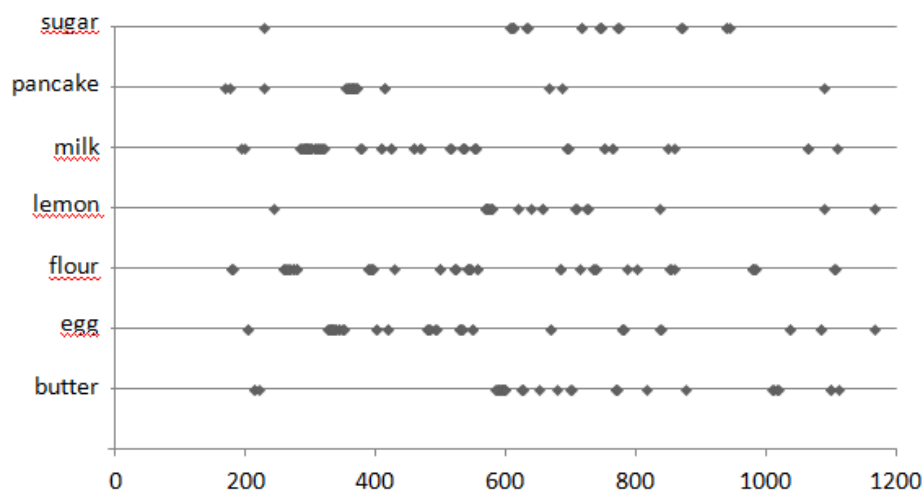


Figure 7: Distribution of food words in Year 3 class

6 Analysing interaction in primary school language classes

be presented and referred to either as substance or discrete units, often along a cline. The word ‘chocolate,’ for instance, can refer to individually wrapped chocolates, to a bar of chocolate, to cocoa, etc. – with consequences on its co-occurrence with determiners (\emptyset , *a*, *the*), with plural -s and with singular or plural verb forms. Consequently, learners have to discover how to map particular determiner+noun+verb combinations onto their possible meanings. Potentially, various types of information are available to them: linguistic exemplars, feedback on their own productions, metalinguistic input, L1 analogies, word-referent associations and physical contact. To what extent are these different kinds of information present in classroom interaction, how do they combine, how do they vary from one class to the other, and with what result on the language of the learners themselves? Contextual information about these occurrences can be obtained with EXAKT, which as a first step displays basic key-word-in-context (KWIC) concordance lines, listed by speaker and by order of occurrence. Figure 8 shows a concordance for the word *apple* as used by the teacher in the Year 1 class, in order of occurrence.

The concordance shows a progression from initial modelling of the word in isolation to uses in context, contained within a grammatical structure with a zero or other determiner (*it’s apple*, *that’s an apple*), apart from lines 21 and 24 where the word is once again repeated in isolation. The interesting thing about the grammaticalized occurrences is that they include references both to *apple* as discrete object (*it’s an apple*; *draw an apple*) and, during the blindfold tasting activity, to *apple* as substance (*it’s not apple*; *say if it’s orange or apple*), with corresponding use of determiners, *an* or \emptyset . This proves to be the case for all of the target food vocabulary in Year 1 (Table 4), with the exception of *egg*, which is only used to refer to an object, not to *egg* as substance.

Table 4: Grammatical contexts of food vocabulary, Year 1 (teacher + learner utterances)

	apple	egg	banana	chocolate	orange	pear
N (in isolation)	36	23	28	39	37	38
\emptyset + N	9	0	7	8	7	7
a(n) + N	14	13	14	1	15	16
some +N	0	0	0	2	0	0
the + N	0	0	0	0	0	1

In the Year 3 class, the target vocabulary consists predominantly of substance-type words (*butter*, *flour*, *milk*, *sugar*). Only *pancake* and *egg* are used countably.

Heather E. Hilton & John Osborne

1		apple .
2		apple .
3		apple .
4		apple .
5		apple .
6		apple .
7		apple .
8	is it (..) [*]	apple ?
9	is it	apple ? no: .it 's not apple what is it ?
10	is it apple ? no: .it 's not	apple what is it ?
11	it 's an	apple . is it an apple ? ye:s . it 's an appl
12	it 's an apple . is it an	apple ? ye:s . it 's an apple . very good .
13	apple . is it an apple ? ye:s . it 's an	apple . very good .
14	it 's an	apple .what is it ?
15	ybody on your board you [*] draw (..) an	apple . [/] you draw an apple . on your board
16	[*] draw (..) an apple . [/] you draw an	apple . on your board (..) you draw [*] (..) an
17	. on your board (..) you draw [*] (..) an	apple .
18	his is not +..is it a & hap [/] is it an	apple ?
19	.Dana (?) .and I want you to draw an (.)	apple . okay ? it 's not apple [*] .very good
20	u to draw an (.) apple . okay ? it 's not	apple [*] .very good . ye:s .shh .very goo:d
21		apple .what is it ?(banana) .and what is it ?
22	have to say if it 's (..) orange (..) or	apple (..) or banana (..) o:r orange [*] .okay
23	chu:t . shhh .is it	apple ?
24		apple . what is it ? shh !
25	it 's	apple . a:nd uh +..uh Maëlys what is it ?
26	it 's an	apple . what is it ?
27	is it	apple ?
28	ry good you can applaud [*] . that 's an	apple . goo:d .

Figure 8: KWIC concordance for *apple*, Year 1 teacher

For the other words, apart from an occurrence of *a butter*, the contexts are exclusively *N* in isolation (one-word utterances) and $\emptyset + N$ (*it's sugar*; *I need sugar*). In this class, isolated nouns, in repetitions or in one-word answers, represent 72% of the occurrences of the target food vocabulary, compared with 64% in Year 1. Figure 9 shows the teacher concordance lines for *egg* in the Year 3 lesson.

Compared with the Year 1 class, the build-up from word-in-isolation to word-in-context is less progressive. As the grammatical contextualisations are introduced, they include the structures previously used for *flour* (*it's N*, etc.) applied, non-grammatically, to pluralized *eggs* (*it's eggs*), then dubiously to *show me eggs*, legitimately to *I need eggs*, and finally to a hybrid *it is an eggs*.

By clicking on a concordance line in EXAKT, it is possible to jump to the corresponding point in the EXMARaLDA transcription, and in the linked video file to see the context. Another useful feature of EXAKT is the possibility of conducting multilevel searches by adding annotation columns to the concordance lines. For example, a search for *apple* or *egg* can be combined with simultaneous searches

6 *Analysing interaction in primary school language classes*

1		eggs .
2		eggs .
3		eggs .
4	listen !	eggs .
5	u:h (...) only: girl [*] .	eggs .
6		eggs .mm !
7	it 's:	eggs .
8	it 's [*]	eggs .what is it
9	show me::((1,2s))[/] show me	eggs .
10	show me	eggs : .
11	show me	eggs .ah yes !
12		eggs .
13		eggs .show me::((1,5s)) flou(r) [*] .
14	show me: (..)	eggs .
15		eggs .
16	show me (.)	eggs .
17		eggs .
18		eggs .yes .
19	I need	eggs .repeat .eggs .and you ?
20	I need eggs . repeat .	eggs .and you ?
21		eggs .and what 's missing ?
22	it is [*] an [*]	eggs ? ye:s !
23		eggs .qui se souvient d' autre chose ?

Figure 9: KWIC concordance for *egg*, Year 3 teacher

on the ‘interaction’ and ‘action’ tiers, to give concordance lines indicating not only the linguistic context, as in a standard KWIC concordance, but also what type of interaction each occurrence belongs to and what action (if any), accompanies it. Table 5 shows an example of a multi-level concordance for the first four occurrences of *egg* in Year 3.

Table 5: Extract from multi-level concordance for *egg*, Year 3.

KWIC concordance			interaction	Action
1	Teacher	eggs .	elicitS	pointing to text
2	Teacher	eggs .	elicitS	showing picture card
3	Teacher	eggs .	presMOD	
4	Teacher	listen ! eggs .	presMOD	

These four occurrences correspond to a short presentation sequence in which

Heather E. Hilton & John Osborne

the teacher first elicits a receptive response by pointing to a text on the whiteboard and holding up a picture card, and then by inviting a learner to come to the whiteboard and point to a picture of eggs (several loose eggs in a basket), followed by two repetitions where the teacher models the word *eggs* without any accompanying action. Most of the succeeding uses of *eggs* by the teacher (18/19) are elicitations (*show me eggs*), positive acknowledgements (*eggs; eggs, yes*), with just one elicitation in the form of a question, formulated as a declarative with rising intonation (*it is an eggs?*). Consequently, most of the learner uses of *eggs* (24/26) are one-word repetition responses, sometimes accompanied by the action of holding up a picture card. Overall, teacher and learners combined, the average length of utterances containing the target food vocabulary is shorter in Year 3 (7.2 transcribed characters) than in Year 1 (10.4 characters), where a similar sequence in Year 1 – this time for *apple* – begins in the same way with several teacher models, but then goes on to include more varied elicitation moves: elicitation questions (*is it apple?*), negative acknowledgement followed by a new elicitation (*no, it's not apple, what is it?*), eliciting a correction (*no, this is not...is it an apple?*, accompanied by the action of holding up a learner's slate). In turn, the learners' responses consist not only of repetitions but also of answers, either as isolated words (*apple*) or as structures or fragments of structures (*a apple, it's apple, it's a apple*).

The pointing and showing that accompany the first occurrences of *egg* are two of the most frequent actions (teachers and pupils combined) in both classes, along with gesturing, moving about the classroom and holding up cards, pictures or objects. Quantitative analysis of actions in the classroom is problematic, since one camera cannot capture everything that goes on, and among the many things which do appear in the frame, the transcriber will necessarily make selections as to what to annotate. Results of concordances on the 'actions' tier in these transcriptions are therefore more useful as pointers to other phenomena than for drawing conclusions about the frequency or distribution of the actions themselves. In this case, pointing, showing and holding up objects are clearly linked to techniques which the two teachers use to present and practice the words in association with their meaning and reference. Although the techniques are similar in the two classes, their relation with the language being produced is not quite the same. The concordance lines for Year 1 show a progression from initial presentation of target vocabulary (*pear, chocolate, egg, apple*, simultaneously pointing to picture cards), then through a *wh*-question sequence for active recall of the new words (*what is it?*, pointing to card), and finally to getting the learners to draw pictures on their slates (*you draw a banana on your board. boards up! good! this is a banana*, pointing to learner's board). Later in the lesson, when the chil-

6 Analysing interaction in primary school language classes

dren are blindfolded and have to guess what kind of food they are tasting, the teacher's 'feeding' action is accompanied by an instruction and a question (*open your mouth. what is it?*), while the learners' 'eating' action is accompanied by exclamation, laughter and *it's N* or *it's a N* constructions (*oh! it's a banana*). In the Year 3 class, where the new words are the ingredients needed to make pancakes, the teacher uses picture cards and pictures on the whiteboard to present the new vocabulary, but she tends to name the pictures herself, with less systematic recall effort from the learners. In the next lesson phase the teacher and learners manipulate the ingredients, taking them out of a shopping bag. Interestingly, this manipulation somewhat blurs the 'substance' meaning of the target words – *sugar, milk, butter, flour* – since what is actually being manipulated are jars and boxes of ingredients. At the same time, between the presentation phase and the manipulation of ingredients, the transition from *it's+N* forms to a new question and answer routine (*what's missing? / what do you need?*) results in exchanges of the type: *what do you need? it's a butter. what is it? what do you need? it's a egg. it's a eggs*.

Comparing the learners' productions in the two groups, we can focus on how they incorporate the new words into embryonic grammatical structures. Extracting all the learner utterances containing the target words and deleting all those that consist of only one word gives the inventory in Table 6; asterisks indicate non-target-like forms, either morphological (e.g. "a apple") or phonological (e.g. "pear" pronounced /pɛ/).

Year 1 learners produced more grammaticalised utterances, mostly on the pattern *it's a N*, sometimes appropriately, when naming pictures of fruit, but also inappropriately when referring to the fruit as substances, in the blindfold tasting activity. The Year 3 learners produced a greater variety of verb structures, not only the presentative *it's*, but also *show me* and *I need*. However, the determiner choices do not follow a clear pattern in relation to the type of reference (object vs. substance) or singular-plural distinction (*it's a eggs*).

In the Year 3 class, the food vocabulary items also occur in French translation. Implicit analogies with L1 are not detectable, but sometimes the learners produce spontaneous translations (Teacher: *it's flour it's flour it's flour* ; Learner: *c'est de la farine !*) or use the L1 for metalinguistic comments (*beurre ça ressemble à bu:t(ter); & fl flour [/] flour . on prononce flour [the French word beurre is like butter; it's pronounced 'flour']*). The teacher herself uses the L1 words for a final recapitulation sequence (*de la farine. comment on le dit? tu te souviens?* [*'de la farine' - How do you say that? Who remembers?*]), since when she asked the learners what new things they had just learned, they spontaneously gave the words in French.

Heather E. Hilton & John Osborne

Table 6: Comparison of target-word utterances by learners, Years 1 and 3

Year 1	Year 3
it 's a [*] apple .	show me:: (..) butter [*] .
(it 's) apple .	(it 's a) butter ?
& a (..) (it 's) a [*] apple .	we (..) need (..) butter . (?)
a [*] apple .	it 's butter ?
it 's a banana .	it 's a [*] egg .
(it 's a) banana .	it 's a [*] eggs .
(it) 's a banana [*] .	(these) are eggs [*] .
it 's a banana .	it 's & fro flour .
it 's [*] banana .	it 's uh flour .
it 's a banana .	it 's flour .
it 's a [*] chocolate .	I need lemon .
it 's a: [*] orange .	it is milk ?
it 's a [*] orange [*] .	(it 's) milk ?
it 's a [*] orange .	need milk .
it 's [*] orange .	(I need) milk .
it 's a pear .	it 's milk [*] .
it 's a pear .	
(a) yellow pear .	
a pear .	
a pear [*] .	
it ('s) a pear .	
(it was) pear .	

Compared with the Year 1 class, the learners make more asides in L1 but also more frequent use of L1 related to the actual content of the lesson.

5 Conclusions for language-teaching and acquisition research

The analyses presented here compare two similar lessons of beginning English, with two different teachers, and two different learner groups (Year 1 six-year-olds, and Year 3 eight-year-olds). We have tried to show how combinations of quantitative (and more partial) qualitative analyses, across different levels of transcrip-

6 Analysing interaction in primary school language classes

tion and annotation, can shed light on some of the factors at play in classroom interaction. The focus of this chapter is on research methodology and the tools which can assist it. From the small comparison that we have used to illustrate these methodological issues, it is not possible to draw wide-ranging pedagogical conclusions.

Nevertheless, a picture emerges, even from such a limited comparison, of two learning environments that are not equally effective. The learner-groups in Year 3 spoke half of the time in English, half in French. This is partially linked to frequency of off-task commentaries, but it is also revealing of the quality of memorisation taking place. At the end of their lesson, when asked during the metalinguistic wrap-up what new words they had learned, the Year 3 learners all gave the words initially in French. The Year 1 teacher used this final phase of the lesson to elicit, one last time, English words from picture cards (direct active recall), and the learners were able to provide the appropriate target words in English.

This outcome is observed at the end of a 30-40 minute lesson; it cannot be clearly attributed to any single cause, but probably results from an accumulation of differences in learning conditions between the two classes. It would be interesting to compare the same classroom over a longer time-span, following the techniques used by teachers to teach different sorts of language knowledge, to work on skills or culture. Another interesting line of observation would be to follow a small set of individual learners, much more closely than in the *Seine & Marne* project, focusing on precise behaviour during a lesson: how often did a learner produce, within which type of interaction; which words did the learner say out loud, which did she only hear, in which contexts, with what frequency and what periodicity? In a small-group study, this could be tied in with measures of emergent language knowledge and skill, as well as measures of individual characteristics of the learner, in a methodology devoted to analysing combinations of the numerous factors that make up the complex learning environment of a language classroom.

References

- Allen, Patrick, Maria Fröhlich & Nina Spada. 1983. The communicative orientation of language teaching: An observation scheme. In Jean Handscombe, Richard A. Orem & Barry P. Taylor (eds.), *The question of control: Selected papers from the Annual Convention of Teachers of English to Speakers of Other Languages*, 231–252. Washington DC: TESOL.
- Allwright, Richard. 1984. The importance of interaction in classroom language learning. *Applied Linguistics* 5(2). 156–171.

Heather E. Hilton & John Osborne

- Allwright, Richard. 1988. *Observation in the language classroom*. London: Longman.
- Arditty, Joseph & Marie-Thérèse Vasseur. 2005. Interaction et diversité des conduites d'apprentissage: présentation. *Acquisition et Interaction en Langue Étrangère* (22). 3–16.
- Bange, Pierre. 1992. À propos de la communication et de l'apprentissage de L2 (notamment dans ses formes institutionnelles). *Acquisition et interaction en langue étrangère* 1. 53–85.
- Bauman, Richard & Charles Briggs. 1990. Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology* 19. 59–88.
- Bowers, Roger. 1980. *Verbal behaviour in the language teaching classroom*. Reading. (Doctoral dissertation).
- Brebner, Mery. 1898. The method of teaching modern languages in Germany. In *The reform movement: Germany and France, 195–280*. Reprinted in 2002. London: Routledge.
- Brown, George. 1975. *Microteaching*. London: Methuen.
- Brundell, Patrick, Paul Tennent, Chris Greenhalgh, Dawn Knight, Andy Crabtree, Claire O'Malley, Shaaron Ainsworth, David Clarke, Ronald Carter & Svenja Adolphs. 2008. *Digital replay system (DRS): A tool for interaction analysis*.
- Cassidy, Steve & Thomas Schmidt. 2017. Tools for multimodal annotation. In Nancy Ide & James Pustejovsky (eds.), *Handbook of linguistic annotation*, 209–227. Dordrecht: Springer.
- Chapelle, Carole. 1990. The discourse of computer-assisted language learning: Toward a context for descriptive research. 24(2). 199–225. DOI: [10.2307/3586899](https://doi.org/10.2307/3586899).
- Chaudron, Craig. 1977. A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning* 27(1). 29–46.
- Chaudron, Craig. 1988. *Second language classrooms*. Cambridge: Cambridge University Press.
- de Pietro, Jean-François, Marinette Matthey & Bernard Py. 1989. Acquisition et contrat didactique : Les séquences potentiellement acquisitionnelles dans la conversation exolingue. In Dominique Weil & Huguette Fugier (eds.), *Actes du troisième colloque régional de linguistique*, 99–124. Strasbourg: Université des Sciences Humaines et Université Louis Pasteur.
- Develotte, Christine, Nicolas Guichon & Viorica Nikolaev. 2008. Apprendre et enseigner les langues en visioconférence. *Les Dossiers de l'ingénierie éducative* 63/64. 69–72.

6 Analysing interaction in primary school language classes

- Erickson, Frederick. 2004. Origins: A brief intellectual and technological history of the emergence of multimodal discourse analysis. In Philip Levine & Ron Scollon (eds.), 196–207. *Discourse and technology: Multimodal discourse analysis*. Washington, DC: Georgetown University Press.
- Firth, Alan & Johannes Wagner. 1997. On discourse, communication, and (some) fundamental concepts in SLA research. *The Modern Language Journal* 81(3). 285–300.
- Flanders, Ned A. 1970. *Analysing teaching behavior*. Reading Mass: Addison-Wesley.
- Gass, Susan. 1998. Apples and oranges: Or, why apples are not orange and don't need to be: A response to firth and wagner. *The Modern Language Journal* 82(1). 83–90.
- Glüer, Michael. 2018. Software for coding and analyzing interaction processes. In Elisabeth Brauner, Margarete Boos & Michaela Kolbe (eds.), *The Cambridge handbook of group interaction analysis*, 245–274. Cambridge: Cambridge University Press.
- Grandcolas, Bernadette & Nicole Soulé-Susbielles. 1986. The analysis of the foreign language classroom. *Studies in Second Language Acquisition* 8(3). 293–308.
- Hilton, Heather E. 2008. *Manuel de transcription, corpus PAROLE*. Chambéry: Université de Savoie.
- Hilton, Heather E. 2017. Individual differences and English L2 learning in two primary classrooms in France. In Janet Enever & Eva Lundgren (eds.), *Researching the complexity of early language learning in instructed contexts*, 127–144. Bristol: Multilingual Matters.
- Hilton, Heather E., Ewa Lenart & Naouel Zoghalmi. 2016. Compréhension et production en anglais L2 à l'école primaire. *Revue française de linguistique appliquée* XXI(2). 65–80.
- Hilton, Heather E. & Carine Royer. 2014. An analysis of complexity in primary school L2 English learning. *EUROSLA Yearbook* 14. 65–78.
- Jarvis, Gilbert A. 1968. A behavioral observation system for classroom foreign language skill acquisition activities. *Modern Language Journal* 52. 335–341.
- Jefferson, Gail. 1972. Side sequences. In David N. Sudnow (ed.), *Studies in social interaction*, 294–33. New York: Free Press.
- Jefferson, Gail. 1983. Issues in the transcription of naturally-occurring talk: Caricature versus capturing pronunciatonal particulars. *Tilburg Papers in Language and Literature* 34. 1–12.
- Jefferson, Gail. 1996. A case of transcriptional stereotyping. *Journal of Pragmatics* 26(2). 159–170. DOI: [10.1016/0378-2166\(96\)00010-0](https://doi.org/10.1016/0378-2166(96)00010-0).

Heather E. Hilton & John Osborne

- Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In Gene Lerner (ed.), *Conversation analysis: Studies from the first generation*, 13–31. Amsterdam: Benjamins.
- Jones, Rodney H. 2013. Data collection and transcription in discourse analysis: A technological history. In Ken Hyland & Brian Paltridge (eds.), *The Bloomsbury companion to discourse analysis*, 9–21. London: Bloomsbury.
- Kasper, Gabriele. 1997. “A” stands for acquisition: A response to Firth and Wagner. *The Modern Language Journal* 81(3). 307–312.
- Kipp, Michael. 2014. ANVIL: A universal video research tool. In J. Durand, U. Gut & G. Kristofferson (eds.), *Handbook of corpus phonology*, 420–436. Oxford: Oxford University Press.
- Lausberg, Hedda. & Han Sloetjes. 2009. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers* 41(3). 841–849.
- Long, Michael. 1997. Construct validity in SLA research: A response to Firth and Wagner. *The Modern Language Journal* 81(3). 318–323.
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analysing talk*. Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, Brian. 2010. A tale of two paradigms. In Michèle Kail & Maya Hickmann (eds.), *Language acquisition across linguistic and cognitive systems*, 17–32. Amsterdam: Benjamins.
- Markee, Numa. 2000. *Conversation analysis*. Mahwah NJ: Lawrence Erlbaum.
- Mary, Latisha. 2014. Pre-service primary teachers’ self-evaluations of language proficiency and self-efficacy for teaching English to young learners. In Janet Enever, Eva Lindgreen & Sergij Ivanov (eds.), *Conference proceedings from early language learning: theory and practice*, 86–90. Umeå: Institutionen för sprakstudier.
- Mondada, Lorenza. 2007. Commentary: Transcript variations and the indexicality of transcribing practices. *Discourse Studies* 9(6). 809–821.
- Mori, Junko. 2007. Border crossings? Exploring the intersection of Second Language Acquisition, Conversation Analysis, and foreign language pedagogy. *The Modern Language Journal* 91. 849–862.
- Moskowitz, Gertrude. 1976. The classroom interaction of outstanding language teachers. *Foreign Language Annals* 9(2). 135–157.
- Nunan, David. 1989. *Understanding language classrooms*. Hemel Hempstead: Prentice Hall.
- Ochs, Elinor. 1979. Transcription as theory. *Developmental pragmatics* 10(1). 43–72.

6 Analysing interaction in primary school language classes

- of Europe, Council. 2001. *The Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.
- Passy, Paul. 1885. *L'instruction primaire aux Etats-Unis: Rapport présenté au Ministère de l'Instruction publique*. Paris: Charles Delagrave.
- Pekarek Doehler, Simona. 2000. Approches interactionnistes de l'acquisition des langues étrangères: Concepts, recherches, perspectives. *Acquisition et interaction en langue étrangère* 12. 1–18.
- Poulisse, Nanda. 1997. Some words in defence of the psycholinguistic approach: A response to Firth and Wagner. *The Modern Language Journal* 81(3). 324–328.
- Rose, Yvan, Brian MacWhinney, Rodrigue Byrne, Gregory Hedlund, Keith Madocks, Philip O'Brien & Todd Wareham. 2006. Introducing Phon: A software solution for the study of phonological acquisition. In David Bamman, Tatiana Magnitskaia & Colleen Zaller (eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development*, 489–500. Somerville, MA: Cascadilla Press.
- Sacks, Harvey, Emanuel Schegloff & Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4). 696–735.
- Schegloff, Emanuel. 1982. Discourse as an interactional achievement: Some uses of 'uh uh' and other things that come between sentences. In Deborah Tannen (ed.), *Analyzing discourse: Text and talk*, 229–249. Georgetown University Round Table on Languages and Linguistics 1981. Washington, DC: Georgetown University Press.
- Schegloff, Emanuel. 2000. When 'others' initiate repair. *Applied Linguistics* 21(2). 205–43.
- Schegloff, Emanuel, Gail Jefferson & Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53(2). 361–382.
- Schmidt, Thomas & Kai Wörner. 2014. Exmaralda. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *The Oxford handbook on corpus phonology*, 402–419. Software: <http://www.exmaralda.org/>. Oxford: Oxford University Press.
- Seedhouse, Paul. 2004. *The interactional architecture of the language classroom: A conversation analysis perspective*. Oxford: Blackwell.
- Seedhouse, Paul. 2005. Conversation analysis and language learning. *Language Teaching* 38(4). 165–187.
- Seedhouse, Paul. 2010. Locusts, snowflakes and recasts: complexity theory and spoken interaction. *Classroom Discourse* 1(1). 4–24.
- Seedhouse, Paul. 2015. L2 classroom interaction as a complex adaptive system. In Numa Markee (ed.), *The handbook of classroom discourse and interaction*, 373–389. Malden, MA: Wiley-Blackwell.

Heather E. Hilton & John Osborne

- Seliger, Herbert & Michael Long (eds.). 1983. *Classroom oriented research in Second Language Acquisition*. Rowley, MA: Newbury House.
- Sinclair, John & Malcolm Coulthard. 1975. *Towards an analysis of discourse*. Oxford: Oxford University Press.
- Tognini-Bonelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Ullmann, Rebecca & Esther Geva. 1982. *The target language observation scheme (TALOS). York region board of education, core French evaluation project*. Toronto: Ontario Institute for Studies in Education.
- Ullmann, Rebecca & Esther Geva. 1984. Approaches to observation in second language classes. In Patrick Allen & Merrill Swain (eds.), *Language issues and education policies: Exploring Canada's multilingual resources* (ELT Documents 119), 113–128. Oxford: Pergamon.
- van Lier, Leo. 1988. *The classroom and the language learner*. London: Longman.
- Vasseur, Marie-Thérèse. 1989. Observables et réalité de l'acquisition d'une langue étrangère: séquences de négociation et processus d'acquisition. *Langage et Société* (50–51). 67–85. DOI: [10.3406/lso.1989.2466](https://doi.org/10.3406/lso.1989.2466).
- Véronique, Daniel. 1992. Recherches sur l'acquisition des langues secondes: Un état des lieux et quelques perspectives. *AILE* 1. 5–35.
- Willis, Jane. 1981. *Spoken discourse in the E.L.T. classroom: A system of analysis and a description*. University of Birmingham. (MA).
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (Irec 2006)*, 1556–1559.
- Wragg, Edward. 1970. Interaction analysis in the foreign language classroom. *Modern Language Journal* 54(2). 116–120.
- Young, Andrea & Latisha Mary. 2010. Une formation des professeurs des écoles en phase avec le 21^e siècle. In Lucile Cadet, Jan Goes & Jean-Marc Mangiante (eds.), *Langue et intégration: Dimensions institutionnelle, socio-professionnelle et universitaire*, 349–354. Brussels: Peter Lang.

Chapter 7

Transcribing interlanguage: The case of verb-final [e] in L2 French

Pascale Leclercq

Université Paul Valéry Montpellier 3

This chapter aims at shedding some light on the place of transcription in the data interpretation process. More specifically, it focuses on the example of verb-final [e] in oral second-language French, which causes interpretation problems when context does not provide disambiguation cues. Through an analysis of three studies on this phenomenon (Herschensohn 2001; Prévost 2007b,a; Granget 2015), displaying a variety of theoretical frameworks (generative versus functional) and transcription options (written- and spoken-centric approaches), I show that transcription choices, whether made intuitively or in a theory-constrained manner, are often problematic as they entail an early categorization of data, even before data coding and analysis, thereby introducing an interpretive bias (Mondada 2007). Finally, I draw some conclusions and offer suggestions regarding best transcription practices.

transcription, data interpretation, L2 French, verbal morphology, interlanguage

1 Introduction

This chapter stems from the author's questions and doubts while engaging in corpus-based second language acquisition (SLA) research, and more specifically when facing oral learner production that is ambiguous and requires the researcher to make conscious transcription decisions. I was particularly puzzled by the way some English learners of French use verb-final [e] as a generic verb ending, in spite of its highly polysemous value (e.g., *tomber* 'fall' infinitive, *tombait* 'was falling' imperfect, *tombé* 'fallen' past participle, etc.) and wondered how to transcribe and to interpret such forms. This phenomenon has been pointed out regularly over the last decades in the literature on the acquisition of morphosyntax



Pascale Leclercq

and has been studied from a variety of theoretical perspectives (Myles et al. 1998; Herschensohn 2001; Bartning & Schlyter 2004), making it an interesting case for a study on the links between interlanguage description and interpretation (see for example the negation phenomenon, Ortega 2014). The description of learner language is indeed a complicated business, which involves, to a certain degree, inferring what learners want to say in a given context. How do researchers deal with this when transcribing raw data? How do they make such inferences? To what extent are their choices theory constrained? This paper aims at shedding some light on these issues.

The concept of interlanguage (Selinker 1972; Han & Tarone 2014; Pallotti 2017) is key in SLA, as it reflects that a learner's language is a system in itself, and one which can vary in time and in contexts of use according to a number of variables (e.g., length and type of instruction, time spent in the target language country, motivation, aptitude, socioeconomic background, linguistic context, interlocutor). It depicts the language of learners as a dynamic unstable system, influenced by the patterns of the source language and/or other known languages, stabilizing at times, and sometimes subject to attrition or fossilization. Interlanguage development (notably speed and level of achievement) is constrained at the level of the individual, yet many researchers believe that there are shared itineraries (e.g., Bartning & Schlyter 2004), although this by no means implies that all developmental paths are the same. Researchers endeavouring to identify the dynamics of learning a language are faced with crucial methodological choices, regarding study design, data transcription and theoretical framework (Mackey & Gass 2012; Revesz 2012). In particular, they have to pay close attention to individual performance so as to find patterns, which may point to some general, and possibly universal, learning processes. Consequently, following Selinker's (1972) admonition to describe interlanguage before engaging in an explanatory process, SLA researchers have been working to map as accurately as possible the way learners of a new language develop their oral and written skills, whether in production or in comprehension (Ortega 2014). Oral production offers privileged access to the processes learners are engaged in when they utter messages in a foreign language, whatever their proficiency level: The repetitions, hesitations, reformulations that are typical of the oral modality may tell the researcher whether the learners are able to plan and structure their discourse and utterances, through more or less automatized access to the second-language (L2) lexicon and grammar, and whether they are able to monitor their speech for errors (Segalowitz 2010: 47 cited by Hilton 2014: 29; Kormos 2006).

This brings us to a major issue facing the SLA researcher: The interpretation of oral interlanguage data, a process in which transcription plays a major part.

7 Transcribing interlanguage

As acknowledged by Ochs (1979) in her pioneering article, and much later by Mondada (2000; 2002; 2007), transcription is a theory-laden interpretive procedure, which incorporates the researcher's theoretical assumptions in the way oral phenomena should be represented and converted to written form.

Although transcription procedures are at the very heart of research on spoken language, the transcription process itself has seldom been explored, and even less so from an SLA perspective. Yet, when transcribing learner oral production, the researcher has to make a series of strategic decisions regarding how to interpret ambiguous forms, such as, for example, verb-final [e] in L2 French, which can stand for homophonous infinitive (*tomber*), imperfect (*tombais*, *tombait*), past participle (*tombé*, *tombés*), or passé simple (*tombai*) marks in standard French, or it might even stand for something else in the learner's interlanguage, such as simple present (Granget 2015). How should such forms be represented in transcription? Should researchers use orthographic conventions and take a decision based on context and their knowledge of the target-language norms (for example, a past adverbial may lead researchers to adopt the interpretation of imperfect or *passé composé*), or should they leave the interpretation open and use phonetic transcription (MacWhinney 2000: 19; Saturno 2015)?

Against this backdrop, I purport to explore the way verb-final [e] in L2 French has been transcribed and analysed in two theoretical perspectives, with the aim of contributing to the current discussion on interlanguage description and interpretation (Ortega 2014), as well as offering some methodological reflections on data transcription.

Keeping in mind MacWhinney's (2000:19) warning that "perhaps the greatest danger facing the transcriber is the tendency to treat spoken language as if it were written language", I introduce the V-[e] transcription problem, and I reflect on the task the transcriber faces as well as on the use of the French writing system to transcribe oral learner data. Then I present the transcription, coding and interpretation choices made by three researchers (Herschensohn 2001; Prévost 2007b,a; Granget 2015) regarding the use of verb final homophonous [e] in L2 French. These three studies were selected as they feature an in-depth analysis of the [e] phenomenon, while offering different theoretical perspectives and transcription strategies (orthographic or phonetic). Based on these three approaches, and following Ortega (2014), I conclude by discussing the link between the choice of theoretical perspective and the description of linguistic data through the transcription process and providing a few guidelines for transcribers.

Pascale Leclercq

2 2. The verb final [e] problem in transcription

My first encounter with verb final [e] took place when I was preparing a paper on how learners' way of referring to time and space in narrative discourse developed over the course of L2 acquisition and whether the data corroborated the Aspect Hypothesis (Leclercq 2011).

Through a focus on motion predicates, in combination with tense and aspect markers, and within a crosslinguistic perspective, I wanted to provide a characterization of elementary, intermediate and advanced proficiency levels. The main research question was the following: How do learners' ways of referring to time and space develop over proficiency levels? I hypothesized that the selection of motion predicates was closely linked to morphological aspect marking (aspect hypothesis, Andersen & Shirai 1994; Robison 1995, also see Rohde 1996 for an overview), and that ambiguous forms as regards tense/aspect marking in English would decrease over proficiency levels.

The experimental design consisted of the administration of a biographical questionnaire, which included information regarding the language-learning history of the participants, and the completion of the Horse Story, an oral picture-based story-telling task developed within the Langacross project (a Franco-German research project funded by the French National Research Agency (ANR) and supervised by Maya Hickmann). The stimulus featured five pictures in which three entities (a horse, a cow and a bird) were localized with reference to a meadow and a fence. It triggered the retelling of motion events (running, jumping, falling, flying) (see Appendix A). It was initially used to study the acquisition of spatial reference in first-language (L1) and L2 French, Chinese, German and English (Hendriks 1998; Hickmann et al. 1998). The retellings were recorded and later transcribed by a trained researcher using CLAN conventions and the recommendations provided by the Langacross team in their coding manual (Hickmann et al. 2011), after which I checked them myself. In accordance with those recommendations, a @G line was inserted in the transcriptions to indicate the correspondence between the learners' utterances and the picture from the stimulus. For the purpose of the present chapter, the target verb final [e] forms are bolded in the following transcription.

I now examine example (1), which presents a retelling by an intermediate-level English learner of French, Mag, including verb-final [e] forms. I highlighted those forms in the transcription.

- (1) Horse Story, 2011 transcription, English L1, French L2 Mag, intermediate.
@G: pict_1

7 Transcribing interlanguage

*SUJ: *dans une première c' est un cheval* . 'in a first it is a horse'
 *SUJ: *qui court dans une prairie* . 'which runs in a meadow'
 @G: pict_2
 *SUJ: *mm c' est mm une vache dans le prochain image* . 'it is a cow in the next picture'
 *SUJ: *séparée par une mm un pré mm avec un oiseau* . 'separated by a field with a bird'
 @G: pict_3
 *SUJ: *mm alors dans le troisième le cheval essayE de monter le* . 'so in the third the horse try to climb the'
 *SUJ: *comment on dit* . 'how do you say'
 *EXP: *la barrière* . 'the barrier'
 *SUJ: *la barrière oui* . 'the barrier yes'
 *SUJ: *mm mais le barrière se casse* . 'but the barrier breaks'
 @G: pict_4
 *SUJ: *dans le quatrième est le cheval tombE* . 'in the fourth is the horse fall'
 @G: pict_5
 *SUJ: *et dans xxx le vache et l' oiseau euh aidE le cheval avec une boîte*
<de> [/] # de aide et # une bandage . 'and in xxx the cow and the bird help the horse with a box of help and a bandage'

What the learner produces is several instances of verbal forms with a verb-final [e] sound, which could either stand for an infinitive, or a past participle, or an inflected form (imperfect *aidait* 'helped' for example), or which could constitute a base form in the learner's interlanguage. What is interesting is that this particular learner used the [e] ending on a regular basis, but also used targetlike verb forms like present tense *se casse* 'breaks down'. The researcher, when facing forms that are ambiguous, has to make decisions regarding transcription and interpretation of data.

Example (1) shows that I chose to use regular orthographic spelling when phonological forms seemed to match the tense/aspect/person agreement rules, as is the case with *séparée* 'separated' on line 4 of the transcription, and the V-E symbol only for the forms which I identified as potentially ambiguous, either due to a position in the sentence that might be interpreted as necessitating an inflected verb (either present tense *aide* with a mute final <e> or passé composé *a aidé* 'helped'), or due to unusual word order (*est le cheval & tombE* 'is the horse fall-ED'). In this project I did not even consider that forms like *séparée* 'separated' could actually be non-targetlike in the mind of the learner, and based

Pascale Leclercq

on our knowledge of French graphonemics, the transcriber and I assumed that learners had produced target forms (a fairly naïve and controversial position). These intuitive transcription choices could be qualified, in the words of Ortega (2014), as pre-theoretical. Mondada (2000: 3) nevertheless points out that such intuitive choices are highly problematic when it comes to data categorisation and interpretation.

We will now try to shed some light on this phenomenon by considering the transcription process itself.

3 Transcribing as a situated practice and a cognitive challenge

Although it is a fundamentally theoretical enterprise (Ochs 1979), and a crucial part of the research process, transcription is a grey zone in most studies. This is reflected in the scarcity of research on the rationale and consequences of transcribers' choices (Mondada 2002: 46). In a series of papers using Ochs' study as a starting point and a framework for the analysis of the activity of transcription from an epistemological perspective, Mondada (2000; 2002; 2007) proposes an in-depth description of the transcription process. Bearing her analyses in mind, I discuss the practical and cognitive challenges awaiting the transcriber when it comes to the transcription of verb final [e] in L2 learner data.

3.1 What is transcription?

First, Mondada (2007) refers to transcribing as a “situated practice”, observing that it is “embedded within a series of research practices: data production, digitalization and compression, anonymization, storage and filing, representation and annotation, analysis, and so on [...]”. These practices configure and more radically ‘fabricate’ what we consider as ‘data’” (p. 810). She acknowledges the fact that transcripts on their own are not data since they cannot be autonomized from recordings (which constitute primary data). Transcripts are considered “secondary products of representation and annotation practices” (Mondada 2007: 810-811). Transcripts and recordings complement each other, particularly when data transcription software such as CLAN is used, as they enable the linking of transcription with the original audio or video recording. While transcripts enable researchers to focus on detail for analysis, recordings provide the possibility to listen again. Mondada therefore acknowledges the evolutive nature of transcripts: The transcriber can endlessly check, revise, reformat, for a specific analysis or

7 Transcribing interlanguage

for editorial purposes (p. 810). Another inherent feature of transcribing is variability (both within and in-between transcriptions, as illustrated in the different treatments of *séparée* and *essayE* in (1)).

On a more abstract level, Mondada (2000) describes the transcription process as an exploitation of writing resources in order to create a representation of oral discourse based on operations of filtering ‘noises’ (phenomena deemed non-meaningful by the transcriber), and of homogenisation through the use of systematic conventions (this latter point is particularly well exemplified in the CLAN and CHAT manuals, MacWhinney 2000). Mondada (2000) observes that the passage from oral to written form has consequences on the interpretability of the spoken language. On the one hand, the transcription appears as a structured account of oral speech, facilitating visual perception. On the other hand, having the possibility to listen again and again to the same recording provokes what Mondada refers to as a magnifier effect: The researcher can focus on a phenomenon which is ephemeral in real-time and might have passed unnoticed during normal conversation.

In line with Ochs (1979); Jefferson (1996) and Saturno (2015), Mondada (2000) indicates that transcribing is an inherently interpretive activity. She also highlights what she calls the circularity problem: Numerous interpretations of phenomena are incorporated a priori in the transcription, in spite of their being the purported aim of the a posteriori analysis of this transcription. In other words, the transcription choices made by the researchers already contain their interpretative choices, making the whole research enterprise dubious.

Mondada (2000: 8) also describes the transcriber’s job as “isolating, cutting out, identifying, making identifiable the recorded forms in a clear written form.”¹ Some notation systems enable the highlighting of indeterminate forms. Among those systems, she mentions the use of International Phonetic Alphabet (IPA) as opposed to orthographic representation. She claims that using IPA suggests a form is deemed non-identifiable by the researcher, while an orthographic rendering indicates that the researcher has already categorized the form. The use of phonetic versus orthographic conventions shows that transcribed forms can be categorised as “more or less comprehensible (transparent or opaque), more or less standard (according to their distance with the standard), [...] a social category, etc.” (Mondada 2000: 8, my translation). While complete IPA transcriptions are often deemed unpractical (and are rarely seen in SLA projects), Mondada points out that the punctual use of IPA allows us to avoid selecting an orthographic form, and therefore a specific language (a crucial point in interlanguage studies),

¹My translation.

Pascale Leclercq

but also to visually detach the transcribed content and consequently highlight it. I will refer to the orthographic option as a written-centric approach, and to the punctual IPA option as a spoken-centric approach. Although the spoken-centric option appears as more careful, as it leaves data interpretation open, it is not always chosen by researchers. Why is that? An interesting hypothesis is that the written code deeply influences literate speakers' representation and categorization of language units (Jaffré 2006, see below).

Keeping Mondada's reflections in mind, I now focus on the choice of orthographic transcription and what it might imply for the transcriber, particularly how phonological and graphical representations may interact and be treated by the transcriber. I use the verb final [e] phenomenon as a basis for the analysis.

3.2 Cognitive aspects of the learner and transcriber's task: Making the most of homophony in spoken French

While most practitioners and researchers agree that verb morphology, including verb-final [e] in French, is a major source of spelling confusion, whether for French children learning how to write (Brissaud & Sardon 1999; David et al. 2006; Fayol & Pacton 2006), for adults who master the orthographic system and use it daily, or for L2 learners in their oral and written productions (David et al. 2006; Brissaud et al. 2006; Prévost 2007b; Granget 2015), some researchers have started focusing on the impact of the orthographic experience on phonological awareness (Bassetti et al. 2015; Nimz & Khattab 2019). According to Detey & Ne-spoulous (2008: 68), several studies suggest that "orthographic representations might play a role in speech perception [...], at least through bidirectional activations between graphemes and phonemes [...]. Hearing a lexical unit might activate an orthographic representation, which might in turn influence phonological judgements." As a consequence, it is quite possible that experts' interpretation of oral learner speech might be largely influenced, directly or indirectly, by spelling knowledge. The transcriber, whether a native speaker or an expert user of the target language, as is the case in (1), often tackles the data with a number of assumptions about what the learner knows about the language, including morphographic representations.

3.3 Description of verb final [e] phenomenon in relation with the French spelling system

According to Jaffré (2006: 25), a spelling system is not a mere tool used for the sake of written communication: Centuries of usage have fostered tight links between spelling forms and their users, whose perceptive abilities they have shaped

7 Transcribing interlanguage

and constructed. He therefore argues that beyond the strict communicative need to disambiguate homophones, their very orthographic differentiation has progressively conditioned the cognitive representations of literate users. In that context, the written code can be considered as an autonomous linguistic representation, capable of exerting an influence on spoken language, as shown by the orthographic distinction of homophones in French. Of course, such an influence is only possible in a society where written communication forms part of the essentials of daily life, and where literacy is a basic skill most citizens master, as is the case in 21st century France.

According to Jaffré (2006), spelling systems have two main objectives. First and foremost, they aim at representing a given language phonographically, but they also have a semiological agenda: They aim at providing a visual representation of language, with tools that enable the disambiguation of spoken forms, for them to be readable and interpretable. In that regards, the use of orthographic spelling for L2 transcriptions compels transcribers to make choices informed by their knowledge of the target language grammar. Additionally, although languages such as French and English use an alphabetic system, their spelling systems often depart from a one-to-one sound/letter relation. In that context, heterography appears as a tool to tease apart homophones.

The French morphological verbal system, and more specifically final [e], provides an interesting case as regards homophony and heterography, and one which poses a particular challenge to the learner. As noted by Brissaud et al. (2006), a form such as /tʁuve/ can be spelt in ten different ways (*trouver, trouvé, trouvés, trouvée, trouvées, trouvaï, trouvais, trouvait, trouvaient, trouvez*), making the native writer and the L2 learner's task a daunting one. Under the umbrella of verb-final [e] forms, there are two types of combining morphemes: Mute endings (-e, -s, -t, -ent), referring to person, gender and number; and tense, aspect and morphemes (among others, -ai, -é, -er, respectively referring to imperfect, perfect and infinitive). Context plays a major role in the selection of the appropriate lexical item, and as a result, the appropriate spelling. However, it is often not sufficient to disambiguate homophonous conjugated verbs, as is the case between infinitive *jouer* and past participle *joué*, which are a frequent source of spelling mistakes, even among French native speakers. Scriptors then have to rely on their knowledge of grammatical rules to select the right target form (David et al. 2006; Brissaud et al. 2006). Regarding verb-final [e], French scriptors have usually received explicit instruction at school as to verb-agreement rules (subject-verb agreement, and tense/aspect endings), grammatical functions and categories (such as infinitive and past participle), and linguistic manipulations to disambiguate homophonous forms (for example, if /ale/ can be replaced by

Pascale Leclercq

/partir/, it has to be an infinitive form spelt *aller*.) They may also rely on their memorization of frequent co-occurring units (for example, *pour* is usually followed by an infinitive form, see [Brissaud et al. 2006: 78](#)). In a written-centric approach, the transcriber's interpretation of ambiguous oral forms consequently requires an analysis of context, and it taps into their phraseological knowledge as well as their knowledge of the target language orthographic rules, in order to apply the relevant contextual, categorical or morphological rules.

3.4 A challenge for the researcher

When transcribing learner data, using a written-centric approach is a leap of faith, as there is no guarantee at all that the learners master the system in the same way as a native speaker. Transcribing then becomes a game of inference: The researcher tries to infer what the learner had in mind when uttering a given phrase, and makes hypotheses regarding their choices, possibly in accordance with the selected theoretical framework of analysis. Transcription problems are consequently inherent to the fact that transcribers often assume that learners share their knowledge of the target language, including its writing system. [Flavell \(1977, cited by Nickerson 1999: 739\)](#) observes that “we are usually unable to turn our own viewpoint off completely when trying to infer the other's, and it usually continues to ring in our ears while we try to decode the other's. It may take considerable skill and effort to represent another's point of view accurately through this kind of noise, and the possibility of egocentric distortion is ever present.” It is therefore important for the researcher to keep in mind that they do not know the extent of the L2 learner's knowledge of the target language and culture. Shared knowledge, and an understanding of what the interlocutor knows, is at the heart of the communication process ([Nickerson 1999; Keysar et al. 2003](#)). We assume our interlocutors share basic communication principles. However, as [Nickerson \(1999\)](#) puts it, “overimputation” of one's knowledge (i.e., attributing learners' knowledge about the target language that they do not necessarily possess), or lack of ability to adopt another perspective than one's own, can cause communication difficulties and lead to wrong interpretation of the interlocutor's message. The researcher has to be careful about “overimputation” and has to remember that nothing can be taken for granted in the realm of L2 acquisition. We cannot assume that an L2 learner thinks in the exact same way as a native speaker, nor that they use verbal forms with the same degree of mastery of form/function relations. When learners use verb-final [e] forms, do they try (with mitigated success) to retrieve targetlike morphology, or do they create a new ending to compensate for their lack of procedural knowledge? When in doubt,

using a spoken-centric approach (i.e., using IPA to transcribe ambiguous forms) might be a good option, as it does not entail early categorisation of the data.

Keeping this in mind, we will examine how verb-final [e] was transcribed and analysed in three different studies (Herschensohn 2001; Prévost 2007b,a,² Granget 2015), displaying a variety of theoretical frameworks (generative vs functional) and transcription options (written and spoken-centric approaches), so as to shed light on the link between data transcription and interpretation, and offer reflections regarding best transcription practices.

4 Making theoretically informed transcription and interpretation choices

As stated by Ortega (2014: 186), “the formal linguistic study of L2 development puts theory first and is driven by the quest to understand the role that Universal Grammar or abstract linguistic knowledge plays in the acquisition of human language across the life span.” Generative researchers therefore favour a top-down approach, with overarching research questions focusing on finding proof in favour of (or against) the theoretical constructs under scrutiny through data analysis (Lardi  re 2012). Data are used to inform or contradict theoretical premises. This is the case with L2 morphosyntax and particularly with the development of inflection (Herschensohn 2001), which have been largely explored within a generative, theory-constrained framework (among others, Pr  vost & White 2000; Pr  vost 2007a; Herschensohn 2001; Hawkins 2004). On the other end of the theoretical spectrum, functionalist researchers (see Lenart & Perdue 2004) take a rather bottom-up, data-driven approach: Based on learner discourse, they try to account for the way learners acquire and make use of formal linguistic levels of organisation (morphology, morphosyntax, lexicon) in a given context of use. In the words of Klein & Perdue (1989) (cited by Lenart & Perdue 2004), researchers have to solve “the learner’s problem of arranging words” to produce a contextually meaningful message. Functionalists have also provided detailed accounts of the verb-final [e] in French L2, through longitudinal studies (for example, studies based on the ESF project, such as Colette et al. 1995; V  ronique 2004), or cross-sectional studies (Bartning & Schlyter 2004; Granget 2015). I compare theory-constrained approaches and data-driven approaches to see what each contributes to the debate on how to interpret, and thus transcribe,

²Pr  vost (2007b) and Pr  vost (2007a) present complementary information about the study under consideration in this chapter.

Pascale Leclercq

oral interlanguage productions, as regards the analysis of verb-final [e]. This section was also largely inspired by the work of [Granget \(2015\)](#), who paved the way for the following analysis by citing the work of [Herschensohn \(2001\)](#) and [Prévost \(2007b\)](#).

4.1 Theory-constrained and written-centric approaches: Defective or missing surface inflection hypothesis

The verb-final [e] phenomenon is presented by [Herschensohn \(2001\)](#) as part of a wider debate on the development of inflection in learner grammars and what it reveals about access of L2 users to Universal Grammar. Within a generative framework of analysis, Herschensohn seeks to determine whether the empirical data support theoretical claims regarding the reason why L2 learners' verbal inflections are so often defective (i.e., not targetlike) especially at intermediate levels. She re-examines the relationship between the acquisition of morphology and functional categories: Some researchers claim that morphology and syntax are developed conjointly in the L2 grammar (co-dependence, [Eubank 1993](#); [Vainikka & Young-Scholten 1996](#); [1998a,b](#) cited by [Herschensohn 2001](#)), in the same way as in L1 grammar, while others reject such a link and propose that morphology and syntax develop independently ([Schwartz & Sprouse 1996](#); [Lardière 1998](#) cited by [Herschensohn 2001](#)). In this view, a lack of morphological marks is attributed to processing difficulties. L2 learners may display evidence of syntactic competence, but deficient morphological production because "under the Missing Surface Inflection Hypothesis, the L2er has a grammar with complete functional projections but incomplete morphology" ([Herschensohn 2001](#): 280).

Within this framework, [Herschensohn \(2001\)](#) argues that inflectional deficits (i.e., non-finite verbs, or other morphological errors) support neither the co-dependence nor the independence hypothesis. She claims that "the French data rather indicate that deficiencies in morphological mapping, not defective syntax (functional categories) are the cause of L2 failed inflection" (p. 273). What is very interesting in this paper, and which provides one of the main reasons to include it in the current analysis, is that the author explicitly motivates her choice of using the infinitival form in the transcription, making her approach a typically written-centric one. Tapping into the abundant literature on the acquisition of morphosyntax, she situated her transcription choice of attributing an infinitive value to verb-final [e] forms from the very onset of the paper: "the second language (L2) use of *infinitives in contexts of obligatory tense* – amply documented and discussed in the literature [...] – is of theoretical interest [...]" (p. 273) (emphasis mine).

Verb-final [e] is examined in this study based on data collected from two stay-abroad participants, Chloe and Emma, through a series of interviews that took

7 Transcribing interlanguage

place before, at midpoint and after their six-month period in France. The interviews include discussion of topics in the present tense, and questions that refer to past and future actions.

The data were transcribed by the author herself and checked by a French phonologist. It yielded “several hundred tokens of verb morphology” (Herschensohn 2001: 285), among which “a number of verb errors are transcribed as infinitives, although the infinitival form is homophonous with the past participle and the *vous* form of the present in most cases” (p. 286). Herschensohn explains the rationale behind her choice of transcription for these forms: “In the cases where the form is transcribed as an –er infinitive, the context clearly excludes the possibility that the inflection is a past participle” (p. 286). She provides the two examples in (2) and (3).

- (2) A quelle heure est-ce que le musée **ouvrir**? (p. 286)
 At what time Q the museum open
- (3) A quelle heure est-ce que le musée **fermer**? (p. 286)
 At what time Q the museum close

She explains this choice as coming from her interpretation of such interlanguage forms as “the regularization of the irregular infinitives *ouvrir* (‘to open’)” rather than as the imperfect *ouvrait* (‘it was opening’)” (p. 286). She notes that when she transcribed the [e] form as an infinitive, the other options (*vous* form *ouvrez*, past participle, imperfect form *ouvrait*) were excluded by contextual information. The author also explicitly rejects an interpretation of those forms as occurrences of the imperfect (*ouvrait*, *fermait*), as she claims that “the use of any tense other than present would be inappropriate” (p. 286). However, what does “inappropriate” mean in the context of a learner production? In spite of the author’s claim, it seems difficult to rule out the imperfect interpretation, which would be grammatically correct, though less expected than simple present in the cited context. This might well be a case of “overimputation” on the part of the researcher (Nickerson 1999), who might have been influenced by her knowledge of French written rules, as proposed by Jaffré (2006).

In sum, Herschensohn (2001) showcases how theory constrains transcription and interpretation choices: The author’s decision to make theoretically informed choices regarding the spelling, and therefore the grammatical status, of final verb [e] forms, entails that these morphological forms are ascribed a specific (potentially erroneous) value even before the data description stage, thereby constraining interpretation. This is what Mondada (2000: 2) refers to as the circularity problem: The researcher makes theory-informed data interpretation decisions

Pascale Leclercq

at the transcription level and then analyses the transcribed phenomenon in the light of the same theoretical framework.

Another example of theoretically informed labelling is provided in [Prévost \(2007b\)](#), in which he refers to verb-final [e] forms as “default forms” in his study on the influence of the source language verbal paradigm on morphological variability in L2 French. He describes L2 French infinitives as recurring erroneous forms used in contexts where other forms are expected in the target language (“1.a. *Ah je voyager* >à> [/] *à des Etats* <U> [/] *Unis*.”). He notices that these forms are used in a similar way by child and adult learners, of various mother tongues (English, Turkish or Chinese), and at different proficiency levels (beginners to advanced). He also points out the fact that these forms are not the only ones that are used by learners and that they can co-exist with targetlike forms, often within the same sentence. He calls this phenomenon morphological variability. Like [Herschensohn \(2001\)](#), he situates his research within a generative framework, and asks whether morphological variability reflects some sort of deficiency of interlanguage grammars, especially as regards functional categories such as Infl(ection).

[Prévost \(2007a\)](#) analyses verbal errors produced by 21 Anglophone learners of French, at four proficiency levels (beginner to upper intermediate 3). He observed that learners produced infinitives in contexts where an inflected verb was expected, such as after a lexical subject, but they also used inflected forms in contexts where an uninflected form was expected, for example after a preposition, an auxiliary or another verb. [Prévost \(2007a: 360\)](#) explains his transcription choices in the following way: “[A]n infinitival verb was considered nonfinite unless evidence of the contrary existed. In other words, verbs ending in [e], which is ambiguous between the infinitival marker *-er*, the past participle marker *-é*, and the second-person plural ending *-ez*, were categorized as nonfinite unless they appeared with the second-person plural pronoun *vous* ‘you’.” Just like [Herschensohn \(2001\)](#), [Prévost](#) adopts a written-centric approach and does not envisage the possibility of an imperfect meaning or of another undetermined value. What is more, his explanation perfectly illustrates the fact that transcription choices imply a categorisation of data (and hence, constrain data analysis) ([Mondada 2007](#)).

The examples provided in the paper show that at the level of transcription, the author chose the *-er* infinitive ending whenever an ambiguous form occurred:

- (4) a. Je parler avec administration (Ann, G1)
 I speak.INF with administration
- b. Ils visiter moi. (Mike, G2)
 they visit.INF me

7 Transcribing interlanguage

- c. On aller au centre d'achats (Jill, G3)
we go.INF to mall
- d. Il retourner à la maison (Sandra, G4) (p. 363)
he return.INF to the house

He justifies his choice by stating that similar constructions also occur with verbs of the second (*ouvrir*) or third (*boire*) group, as in (5) a, b, c and are also found in his data with negation as in (6):

- (5) a. Il boire (Edward, G1) (p. 361)
he drink.INF
- b. Il prendre des vêtements (John, G4) (p. 361)
he take.INF some clothes
- c. ... quand ils ouvrir les cadeaux (Kate, G3) (p.364)
when they open.INF the presents
- (6) Je ne expliquer pas (Mark, G1) (p. 365)
I NEG explain.INF NEG
- (7) a. J' ai difficile demande des questions (Ann, G1)
I have.1s difficult ask.s some questions
- b. Comment tu vas arrive à mon travail ?
how you go.s to my work (Jen, G2)
- c. ... qui uh j' ai rencontre à Nouvelle Ecosse (Jill, G3)" (p.369)
who I have.1s meet.s in Nova Scotia

Examples from [Prévost \(2007a\)](#), presented in [Prévost \(2007b: 61\)](#)

Just like Herschensohn, [Prévost \(2007a\)](#) refers to those forms as errors. What is very interesting in [Prévost \(2007a\)](#) is that he observes and describes substantial inter-individual variation in the use of uninflected forms where an inflected form is expected, and in the use of inflected forms where uninflected forms are expected as in (7). Based on the fact that the learners in this study are Anglophones and that their L1 does not possess a particular marker for infinitive, he interprets their choice of inflected forms where uninflected ones are expected as some sort of base uninflected form for the learners. He also very perceptively observes that many errors share an ambiguous phonological form (*je/tu/il/elle/on/il-s/elles chant[t]* 'I/you/he/she/it/they sing'), and that they could very well be uninflected forms in the mind of learners. Although his overarching research questions are driven by the generative-approach agenda, and in spite of transcription

Pascale Leclercq

choices that are highly constraining as they constitute a pre-analysis and pre-interpretation of the data, [Prévost \(2007a\)](#) proposes a careful description of the data under scrutiny and is wary of overinterpretation. We will now turn to a more strictly data-driven approach with the analysis proposed by [Granget \(2015\)](#).

4.1 Data-driven and spoken-centric approaches

[Granget \(2015\)](#) investigates the acquisition of the French present tense, also called “présent simple”, and more particularly the emergence of the morphological expression of this tense in L2 French, and she adopts a radically different approach. She used a database of 36 oral picture story retellings (Loch Ness stimulus from the French Learner Language Oral Corpora (FLLOC) project, [Marsden et al. 2002](#)) by Anglophone teenagers learning French, at three different institutional levels (3 to 5 years of instruction, 12 learners per level).

[Granget \(2015\)](#) described the database under investigation, and emphasised its interactional nature: During the task the learners interacted with the interviewer, who provided either positive feedback (*très bien, c’est bien* ‘very good, it’s good’ etc.) to answer the learners’ vocabulary requests or helped them move the narrative forward (*en bateau très bien, qu’est-ce qu’elle fait?* ‘In a boat very good, what is she doing?’). She then explained her transcription choices: Although the FLLOC recordings were transcribed following CLAN procedures, Granget modified the original transcriptions where ambiguous forms occurred. More specifically, she “demorphologized” ambiguous verbal forms and transcribed them phonetically in order to question the learners’ morphological representation of these forms. For example, when the learner asks the interviewer “what’s fishing?” (p. 123), the reply [peʃ] is transcribed phonetically because Granget did not suppose a priori that the learner used an inflected form, as the use of *pêche* ‘to fish, simple present, 3rd person singular’ or *pêchent* ‘to fish, simple present, 3rd person plural’ would indicate. Non-ambiguous irregular verb forms such as *sont* ‘be 3rd person plural’, *est* ‘be 3rd person singular’, *fait* ‘do 3rd person singular’ are nevertheless transcribed using orthographic conventions.

[Granget \(2015\)](#) sought to answer the following research question: To what extent can we consider that L2 learners’ verbal forms include morphological markers? For example, in (8), is [sorte] ‘go out’ an inflected form (base form [sort] + *e* morpheme)? In (9), [di] sounds like a native-like inflected form (*dit*, ‘says’) but is this the case from the learner’s view point?

- (8) le mère et le petit garçon [sorte] les villes
 ‘the mother and the little boy go_out the city’

7 Transcribing interlanguage

- (9) les enfants et le grand-mère [di] au revoir
 ‘the children and the grandmother say bye’

She performed a qualitative analysis using the Finiteness framework (Klein 2006). According to Klein (2006); Gretsch & Perdue (2007); Perdue et al. (2002), finiteness “is usually associated with the morphosyntactic categories of person and tense” (Perdue et al. 2002: 853). However, Perdue et al. distinguish between the finiteness concept and the markers used to express it in the world’s languages. Klein (1994) relates finiteness to assertion (i.e., “the speaker’s making a claim about a time span” Perdue et al. 2002: 853). This implies temporal and aspectual anchoring (Klein 1994). The authors distinguish two types of finiteness:

- S(ematic) finiteness is a property of the whole utterance and is often marked by discursive or lexical means.
- M(orphological) finiteness denotes finite verbal morphology.

Within this framework, the learner’s tasks are the following: (1) noticing and acquiring the means that the target language provides for the expression of S-finiteness, and (2) finding out whether there are grammaticalised means to express (M) finiteness. Development from S (i.e., discursive or lexical means) to M (i.e., morphological means) finiteness denotes a progression towards native norm, as illustrated by these examples from the ESF project: (examples (10) and (11) come from Véronique 2004, example (12) comes from Granget 2015).

(10) Lexical and pragmatic means : [e] la dame la cassette comme ça. (Zarah)

(11) Lexical means : moi [iparle] comme la cassette (Zarah)

(12) Morphological means : Ma fille elle va déjà au lycée (Alfonso)

In (10) and (11), Zarah describes a French class she attends. She relies on her interlocutor’s capacity of inference, and on pragmatic means (in (10) she uses a gesture to show how the lady plays the cassette), while using lexicalised form [iparle]. In (12), Alfonso uses a targetlike verb form.

When coding, Granget (2015) was careful to use labels that do not predetermine the finite/non-finite status of the target elements: V-/e/, V, Aux + V. Her analysis suggests that the same verb is often used within the same production with phonological variations that might reveal morphological variation ([rəgard]/[rəgarde]), an observation also made by Prévost (2007b,a). She tries to account

Pascale Leclercq

for the distribution of such allomorphs in the corpus and envisages several possible explanations for this phenomenon. First of all, the presence of a final [e] morpheme in some but not all verbal form makes it difficult to decide whether [e] can be interpreted as a flexional morpheme or whether forms such as [rəgard]/[rəgarde] both belong to the mental lexicon. She highlights the extreme difficulty in using target-language functional categories such as tense/aspect to describe learners' interlanguage. She then evokes the Aspect Hypothesis to account for the use of verb final [e] in the data. According to this hypothesis, verb final [e] would occur more frequently with predicates denoting bounded events. However, she observes that it is difficult to determine the lexical aspect of some of the verbal forms used by learners. In (13), it is difficult to assert whether [rəturne] is telic (taking the buoys out of the lake) or atelic (describing the state when the buoys are on the bank).

- (13) le grand-mère [rəturne] les bouées de le lac
(? 'the grandmother turn_round the buoys of the lake')

Granget (2015) therefore decides against the Aspect Hypothesis to account for the use of those forms and concludes that these forms should be treated as non-finite and non-analysed and that free variation is the rule in learner discourse (p. 132).

- (14) A30: *l'enfance et (.) le le mère (.) euh [rəgarde] la monster dans la lac*
ADR : *mmm*
A30 : *euh (.) un [?] euh (.) un journaliste et touriste*
ADR : *mmm*
A30 : *euh [rəgard] la monster euh*
A30 : *euh (.) maintenant euh l'enfance le enfles enfants [rəgarde] la tele*

Granget (2015) finally states that the data show high variability in the means of temporal anchoring and linking within and between narratives and advocates for the use of descriptive labels that avoid a pre-analysis assignment of linguistic category, in an effort not to overinterpret learner data. She suggests V-/e/ forms can be interpreted as 'verboidial means of assertion' (i.e., lexical forms with syntactic properties of finite verbs). As for whether these forms display morphological properties of inflected verbs or not, Granget interprets them as "non-inflected and non-analysed verboidial forms" (p. 132).

Although these three studies focus on a common phenomenon (i.e., the use of verbal inflection in the oral production of L2 learners of French), the authors' decisions regarding data transcription were all firmly anchored in their theoretical

framework. Our comparison shows that decisions at the transcription level condition the description of a given phenomenon and its interpretation. In fact, the transcription stage is strongly dependent on theoretical assumptions. Depending on their overarching goals and on their theoretical framework, researchers may opt for transcription choices that reveal a pre-categorisation of the data (e.g., adopting the *-er* infinitive spelling for verb-final [e], a written-centric approach), or for earmarking ambiguous forms for future analysis, such as using the phonetic symbols instead of deciding on a specific spelling (spoken-centric approach), thus keeping all interpretive options open.

5 Conclusion

Through this chapter, I aimed to shed light on the transcription process in the context of research based on learner corpora, more specifically when oral production tasks are involved. Transcription indeed constitutes a crucial step in the constitution of oral corpora, as it shapes the data and makes them ready for subsequent analyses. The current study briefly described the cognitive aspects of transcription and focused on the methodological implications of transcription choices (written vs spoken-centric approaches).

First, I tried to describe the difficult task facing the researcher when transcribing learner data, and more particularly in the case of ambiguous homophonous forms, such as verb-final [e] in French L2. Data transcribers often work on the assumption that learner language can be safely mapped onto the target written orthographic system and often rely on phonological and contextual cues provided by the learners to process and make sense out of ambiguous sounds. However, this entails a risk of “overimputation” on the part of the researcher (Nickerson 1999). This in turn creates methodological problems, as data description, analysis and interpretation are highly dependent on the transcription process itself (Mondada 2007). For example, transcribing ambiguous verb-final [e] as an infinitive *-er* form, as Herschensohn (2001) and Prévost (2007b,a) do, reveals pre-analysis choices that entail a labelling of those forms as errors, thus conditioning the subsequent analysis. This subsequently leads us to questioning the validity of analysing learner data in the light of native speakers’ productions. In Dewaele’s (2003) homage to Larry Selinker, one of the founding fathers of the interlanguage concept, he notices that in spite of the enormous success of this concept, linguists persist in comparing interlanguage with native speakers’ systems and tend to analyse any deviation from the norm as a deficiency on the part of the learner. He then praises Cook’s (2002) plea for learners to be considered

Pascale Leclercq

as language users rather than as failed native speakers. We believe this is sound practice and we wish to encourage researchers to beware of overimputation or overinterpretation of learner data. When possible, we assume the best transcription practice is to adopt a spoken-centric approach and earmark the problematic forms through the use of IPA, without taking any decision regarding interpretation at the transcription level, as proposed and exemplified by [Mondada \(2002\)](#); [Granget \(2015\)](#) and [Saturno \(2015\)](#).

As for interpretation, I hope to have shown that it heavily depends on theoretical framework. It is also closely linked to the transcriber's intimate knowledge of the writing system and the assumption that such knowledge is shared by learners. However, we have no way of verifying this assumption from listening to a recorded production. Indeed, it remains to be understood why learners of French use verb-final [e]. None of the three papers under scrutiny in the present chapter provides a definite or convincing answer to this intriguing phenomenon. Let's go back to the starting point of the current chapter: It is clear from example (1) that Mag is a learner who has created her own idiosyncratic verbal system, which includes endings reminiscent of target verb forms. What is not known is the extent to which she is aware of the grammar rules that enable the French native speakers to differentiate, for example, an infinitive and a past participle. She started learning French at school, age 15. She spent 6 months as an au pair in Paris when she was 18. She had been at a French university for 3 months when her interlanguage was recorded. She might have been taught the relevant rules in an instructed setting but not been able to access them when producing oral discourse, thus displaying a lack of proceduralization of the rules. Or, she might not have understood the rules in a correct fashion and therefore have opted for a creative solution that is compatible with a variety of interpretations on the part of her interlocutor. Or, she might not have been taught those particular sets of rules and rely on frequency effects from the input. It would have been interesting to ask the participant herself what she thought these forms stood for, consequently involving her in data construction and interpretation, as suggested in [Revesz \(2012\)](#). The use of a think aloud protocol ([Leow & Morgan-Short 2004](#)) could help gain access to speakers' representations and what they have in mind when using [e] forms, yet it could only take place as a retrospective task, by asking them to transcribe their own production or by having the participant listen to their production and ask them to comment on what they think they meant by the use of such forms. Notwithstanding, we have no way of making sure that the participant's representation is stable and that they know how to explain it to the researcher ([Norris & Ortega 2012](#); see also [Gass & Mackey 2000](#) on a presentation of learner introspection and retrospection techniques).

To fulfill SLA's objective of describing and understanding the dynamics of interlanguage development, we need adequate transcription procedures to propose a valid interpretative framework for data analysis. In that regard, the spoken-centric approach seems a good fit for the purpose. Finally, we need to thoroughly document data transcription and data coding procedures. I believe that bearing those issues in mind when designing experimental settings is crucial to provide meaningful research results and contribute to a sound description of learner language development.

Acknowledgements

I wish to thank the series editors as well as the anonymous reviewer and my fellow co-editors for their detailed and insightful remarks on earlier version of this paper. They helped me considerably improve the shape and content of this chapter.

References

- Andersen, Roger W. & Yasuhiro Shirai. 1994. Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition* 16. 133–156.
- Bartning, Inge & Suzanne Schlyter. 2004. Itinéraires acquisitionnels et stades de développement en français L2. *French Language Studies* 14. 281–299.
- Bassetti, Benedetta, Paola Escudero & Rachel Hayes-Harb. 2015. Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics* 36. 1–6.
- Brissaud, Catherine, Jean-Pierre Chevrot & Pascale Lefrançois. 2006. Les formes verbales homophones en /E/ entre 8 et 15 ans: Contraintes et conflits dans la construction des savoirs sur une difficulté orthographique. *Langue Française* 151. 74–93.
- Brissaud, Catherine & Jean-Michel Sandon. 1999. L'acquisition des formes verbales en /E/ à l'école élémentaire et au collège: Entre phonographie et morphographie. *Langue Française* 124. 40–57.
- Colette, Noyau, Et-Tayeb Houdaïfa, Marie-Thérèse Vasseur & Georges Daniel Véronique. 1995. The acquisition of temporality in French. In W. Klein Dietrich R. & C. Noyau (eds.), *The acquisition of temporality in a second language*, 145–210. Amsterdam: John Benjamins Publishing Company.
- Cook, Vivian (ed.). 2002. *Portraits of the L2 user*. Clevedon: Multilingual Matters.

Pascale Leclercq

- David, Jacques, Odile Guyon & Catherine Brissaud. 2006. Apprendre à orthographier les verbes: Le cas de l'homophonie des finales en /E/. *Langue Française* 151. 109–126.
- Detey, Sylvain & Jean-Luc Nespoulous. 2008. Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua* 118. 66–81.
- Dewaele, Jean-Marc. 2003. Compte rendu – hommage: l'œuvre de L. Selinker. 49. 153–159. DOI: [10.4000/linx.564](https://doi.org/10.4000/linx.564).
- Eubank, Lynn. 1993. On the transfer of parametric values in L2 development. *Language Acquisition* 3(3). 183–208. DOI: [10.1207/s15327817la0303_1](https://doi.org/10.1207/s15327817la0303_1).
- Fayol, Michel & Sébastien Pacton. 2006. L'accord du participe passé: Entre compétition de procédures et récupération en mémoire. *Langue Française* 151. 59–73.
- Flavell, John H. 1977. *Cognitive development*. Englewood Cliffs, NJ: Prentice Hall.
- Gass, Susan M. & Alison Mackey. 2000. *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum.
- Granget, Cyrille. 2015. Pourquoi le présent n'est-il pas si simple d'un point de vue psycholinguistique? *Language, Interaction and Acquisition* 6(1). 107–148.
- Gretsch, Petra & Clive Perdue. 2007. In. 2007. Finiteness in first and second language acquisition. In F. Planck and I. Nikolaeva (eds), *Finiteness. Theoretical and empirical foundations*, 432–484. Oxford: Oxford University Press.
- Han, ZhaoHong & Elaine Tarone. 2014. Introduction. In E. Tarone Z. Han (ed.), *Interlanguage: Forty years later*, 1–6. Amsterdam: John Benjamins Publishing Company.
- Hawkins, Roger. 2004. The contribution of the theory of universal grammar to our understanding of the acquisition of French as a second language. *French Language Studies* 14. 233–255.
- Hendriks, Henriëtte. 1998. Comment il monte le chat? en grimant! *Acquisition et Interaction en Langue Étrangère* 11. 147–190.
- Herschensohn, Julia. 2001. Missing inflection in second language French: Accidental infinitives and other verbal deficits. *Second Language Research* 17(3). 273–305.
- Hickmann, Maya. 1982. *The development of narrative skills: pragmatic and metapragmatic aspects of discourse cohesion*. Chicago, IL: University of Chicago. (Doctoral dissertation).
- Hickmann, Maya, Henriëtte Hendriks & Françoise Roland. 1998. Référence spatiale dans les récits d'enfants français: Perspective inter-langue. *Langue Française* 118. 104–123.

7 Transcribing interlanguage

- Hickmann, Maya, Henriëtte Hendriks, Annie-Claude Demagny, Helen Engemann, Tatiana Iakovleva, Yinglin Ji, Anna-Katharina Ochsenbauer & Efstathia Soroli. 2011. *La représentation de l'espace - études expérimentales et translinguistiques*. Manuel de codage.
- Hilton, Heather. 2014. Oral fluency and spoken proficiency: Considerations for research and testing. In Pascale Leclercq, Amanda Edmonds & Heather Hilton (eds.), *Measuring L2 proficiency: Perspectives from SLA*, 27–53. Bristol: Multilingual Matters. DOI: [10.21832/9781783092291-005](https://doi.org/10.21832/9781783092291-005).
- Jaffré, Jean-Pierre. 2006. Pourquoi distinguer les homophones? *Langue Française* 151. 25–40.
- Jefferson, Gail. 1996. A case of transcriptional stereotyping. *Journal of Pragmatics* 26(2). 159–170. DOI: [10.1016/0378-2166\(96\)00010-0](https://doi.org/10.1016/0378-2166(96)00010-0).
- Keysar, Boaz, Shuhong Lin & Dale J. Barr. 2003. Limits on theory of mind use in adults. *Cognition* 89. 25–41.
- Klein, Wolfgang. 1994. *Time in language*. Routledge. London.
- Klein, Wolfgang. 2006. On finiteness. In Veerle Van Geenhoven (ed.), *Semantics in acquisition*, 245–272. Dordrecht: Springer.
- Klein, Wolfgang & Clive Perdue. 1989. The learner's problem of arranging words. In Elizabeth Bates & Brian MacWhinney (eds.), *The cross-linguistic study of sentence processing*, 292–327. Cambridge: Cambridge University Press.
- Kormos, Judit. 2006. *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Lardière, Donna. 1998. Dissociating syntax from morphology in a divergent L2 end-stage grammar. *Second Language Research* 14(4). 359–375. DOI: [10.1191/026765898672500216](https://doi.org/10.1191/026765898672500216).
- Lardière, Donna. 2012. Linguistic approaches to second language morphosyntax. In Susan M. Gass & Alison Mackey (eds.), *The Routledge handbook of second language acquisition*, 106–126. Oxon: Routledge.
- Leclercq, Pascale. 2011. *How do learners of French and English include space and time reference in narratives?* Paper presented at the EUROSLA 21 conference, Stockholm, 8–10 September 2011.
- Lenart, Ewa & Clive Perdue. 2004. L'approche fonctionnaliste: Structure interne et mise en œuvre du syntagme nominal. *Acquisition et Interaction en Langue Etrangère* 21. 85–121.
- Leow, Ronald P. & Kara Morgan-Short. 2004. To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition* 26(1). 35–57.
- Mackey, Alison & Susan Gass. 2012. *Research methods in second language acquisition: A practical guide*. Chichester: Blackwell Publishing.

Pascale Leclercq

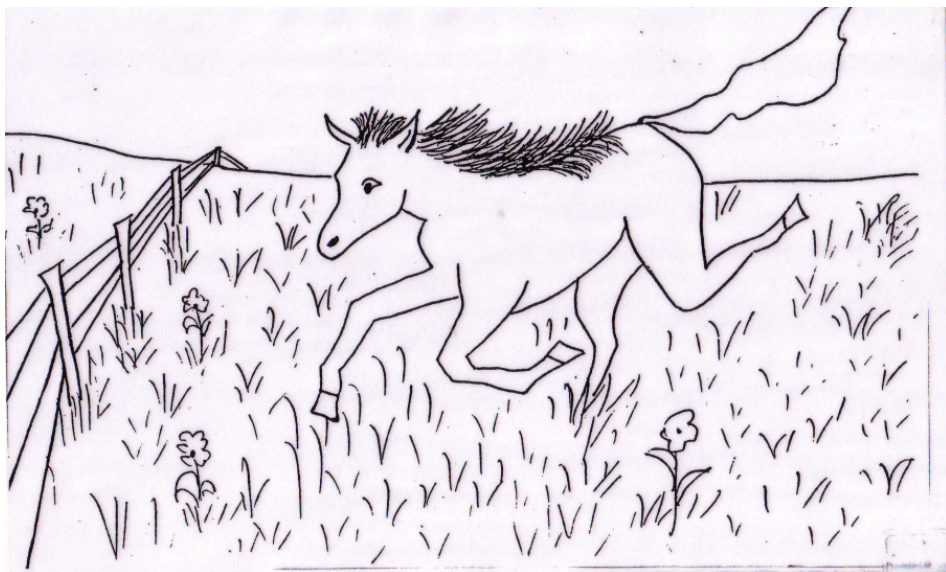
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analysing talk*. Mahwah, NJ: Lawrence Erlbaum.
- Marsden, Emma, Rosamond Mitchell, Sarah Rule & Florence Myles. 2002. *Oral French interlanguage corpora: Tools for data management and analysis*. Occasional Paper No. 58.
- Mondada, Lorenza. 2000. Les effets théoriques des pratiques de transcription. *Linx* 42. 131–146.
- Mondada, Lorenza. 2002. Transcrire l'interaction. *Cahiers de Praxématique* 39. 45–75.
- Mondada, Lorenza. 2007. Commentary: Transcript variations and the indexicality of transcribing practices. *Discourse Studies* 9(6). 809–821.
- Myles, Florence, Janet Hooper & Rosamond Mitchell. 1998. Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning* 48(3). 323–364.
- Nickerson, Raymond S. 1999. How we know - and sometimes misjudge - what others know: Imputing one's own knowledge to others. *Psychological Bulletin* 125(6). 737–759.
- Nimz, Katharina & Ghada Khattab. 2019. On the role of orthography in L2 vowel production: The case of Polish learners of German. *Second Language Research* 28. 387–413. DOI: [10.1177/0267658319828424](https://doi.org/10.1177/0267658319828424).
- Norris, John M. & Lourdes Ortega. 2012. Assessing learner knowledge. In Susan M. Gass & Alison Mackey (eds.), *The Routledge handbook of second language acquisition*, 573–589. Oxon: Routledge.
- Ochs, Elinor. 1979. Transcription as theory. *Developmental pragmatics* 10(1). 43–72.
- Ortega, Lourdes. 2014. Trying out theories on interlanguage: Description and explanation over 40 years of L2 negation research. In ZhaoHong Han & Elaine Tarone (eds.), *Interlanguage: Forty years later*, 173–201. Amsterdam: John Benjamins.
- Pallotti, Gabriele. 2017. Applying the interlanguage approach to language teaching. *International Review of Applied Linguistics* 55(4). 393–412.
- Perdue, Clive, Sandra Benazzo & Patrizia Giuliano. 2002. When finiteness gets marked: The relation between morphosyntactic development and use of scopal items in adult language acquisition. *Linguistics* 40(4). 849–890.
- Prévost, Philippe. 2007a. Knowledge of morphology and syntax in early adult L2 french: Evidence for the Missing surface Inflection Hypothesis. In, 353–377. Liceras, Juana M. and Zobl, Helmut and Goodluck, Helen. Mahwah, NJ: Lawrence Erlbaum.

7 Transcribing interlanguage

- Prévost, Philippe. 2007b. L'influence du paradigme verbal de la langue maternelle sur la variabilité morphologique en acquisition du français L2 par des apprenants adultes. *French Language Studies* 17. 49–80.
- Prévost, Philippe & Lydia White. 2000. Missing surface inflection or impairment in second language? Evidence from tense and agreement. *Second Language Research* 16(2). 103–133.
- Revesz, Andrea. 2012. *Coding second language data validly and reliably*. Alison Mackey & Susan M. Gass (eds.). Chichester: Blackwell Publishing. 203–221.
- Robison, Richard E. 1995. The aspect hypothesis revisited: A cross-sectional study of tense and aspect marking in interlanguage. *Applied Linguistics* 16. 344–370.
- Rohde, Andrea. 1996. The aspect hypothesis and the emergence of tense distinction in naturalistic L2 acquisition. *Linguistics* 34. 1115–1137.
- Saturno, Jacopo. 2015. La trascrizione di varietà di apprendimento iniziali. In Marina Chini (ed.), *Il parlato in (italiano) L2: Aspetti pragmatici e prosodici. [italian L2 spoken discourse: Pragmatic and prosodic aspects]*, 117–138. Milano: FrancoAngeli.
- Schwartz, Bonnie D. & Rex A. Sprouse. 1996. L2 cognitive states and the full transfer/full access model. *Second Language Research* 12(1). 40–72. DOI: [10.1177/026765839601200103](https://doi.org/10.1177/026765839601200103).
- Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. New York, NY: Routledge.
- Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics* 10. 209–231. DOI: [10.1515/iral.1972.10.1-4.209](https://doi.org/10.1515/iral.1972.10.1-4.209).
- Vainikka, Anne & Martha Young-Scholten. 1996. Gradual development of L2 phrase structure. *Second Language Research* 12(1). 7–39. DOI: [10.1177/026765839601200102](https://doi.org/10.1177/026765839601200102).
- Vainikka, Anne & Martha Young-Scholten. 1998a. Morphosyntactic triggers in adult SLA. In Maria-Luise Beck (ed.), *Morphology and its interfaces in second language knowledge*, 89–114. Amsterdam: John Benjamins.
- Vainikka, Anne & Martha Young-Scholten. 1998b. The initial state in the L2 acquisition of phrase structure. In Suzanne Flynn, Gita Martohardjono & Wayne O'Neil (eds.), *The generative study of second language acquisition*, 17–34. Mahwah, NJ: Laurence Erlbaum.
- Véronique, Daniel. 2004. The development of referential activities and clause-combining as aspects of the acquisition of discourse in French as l2. *French Language Studies* 14. 257–280.

Pascale Leclercq

Appendix A: The horse story stimulus (**Hickmann 1982**).



Picture 1

7 Transcribing interlanguage: The case of verb-final [e] in L2 French



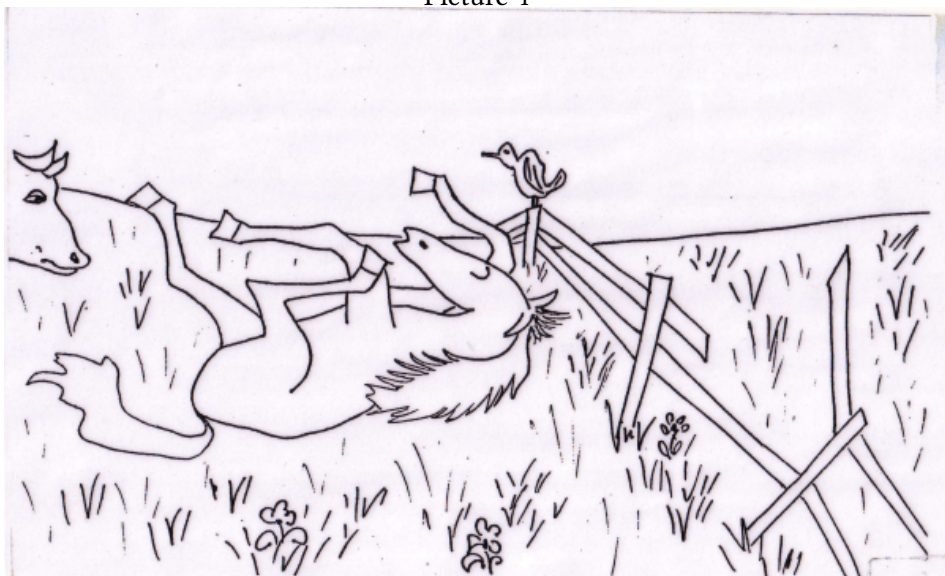
Picture 2

Picture 3



Pascale Leclercq

Picture 4



Picture 5

Chapter 8

Potential pitfalls of interpreting data from English-French tandem conversations

Sylwia Scheuer

University of Paris 3 – Sorbonne Nouvelle

Céline Horgues

University of Paris 3 – Sorbonne Nouvelle

The chapter focuses on methodological issues involved in analysing, coding and interpreting data from the *Spécificités des Interactions verbales dans le cadre de Tandems linguistiques Anglais-Français* (*Characteristics of English/French spoken tandem interactions*) corpus of English-French tandem exchanges. Each of the 21 tandem pairs recorded consisted of a native speaker of English and a native speaker of French. The participants were video and audio recorded while performing tasks (conversation and reading) in both languages. So far, two major threads of research on the corpus data have emerged: corrective feedback and communication breakdowns. We have attempted to gain insights as to when or why corrective feedback is given to the non-native tandem partner and when or why communication between the partners gets compromised. Findings from those previous thematic areas serve as the basis for the present study. The major challenge we have encountered in conducting the analyses is the ambiguity and complexity of our conversational data. Both corrective feedback and communication breakdowns may have multiple – and not always obvious – causes and may or may not be clearly signalled by the participants. In the chapter, we discuss the various problems we faced and addressed while coding the data, as well as how the methodological choices we made affect our results and conclusions. The discussion is amply illustrated with examples from the corpus.

tandem learning, corrective feedback, communication breakdowns, data coding, NS-NNS interactions, English L1 and L2 speakers, French L1 and L2 speakers



Sylvia Scheuer & Céline Horgues

1 Introduction

Tandem learning is “an arrangement in which two native speakers of different languages communicate regularly with one another, each with the purpose of learning the other’s language” (O’Rourke 2005: 434). Consequently, tandem interactions constitute a unique collaborative language-learning environment, which is based neither on the socially institutionalised teacher-learner hierarchy nor on the exact symmetry of peer interactions, where learners share their first language (L1) and their target second language (L2). Instead, it is based on role-reversibility and solidarity between the two tandem partners, each of whom will construct two roles throughout the conversation exchange and, more generally, throughout their tandem history: the role of the (relative) expert when speaking in their mother tongue and the role of the learner, or the less proficient speaker, when speaking in the L2. The fact that each participant gets to wear the hat of both the native speaker (NS) and the non-native speaker (NNS) at some point in the interaction makes their relationship essentially non-hierarchical. Language-expertise asymmetry is only contextual (the conversation invariably switches from one’s L1 to L2, or the other way round, within a short period of time), which makes tandem exchanges also different from the classic NS-NNS conversational setting, where the expert-novice relationship is not reversible.

The database that the present contribution draws on to discuss such interactions is the *Spécificités des Interactions verbales dans le cadre de Tandems linguistiques Anglais-Français* (SITAF: *Characteristics of English/French spoken tandem interactions*) corpus, in which we collected linguistic material – both video and audio recorded – from face-to-face conversational exchanges held by 21 pairs of undergraduate students at the University of Paris 3 – Sorbonne Nouvelle. Each such tandem consisted of a NS of English and a NS of French. By virtue of containing largely unscripted L1-L2 productions, the corpus offers ample opportunities for various types of analyses of NS-NNS interactions, including studies of corrective feedback (CF) and communication breakdowns (CBs). It is those two, overlapping, research areas that the chapter focuses on, with a view to presenting various methodological challenges that researchers can face when coding and interpreting data.

We equate corrective feedback with the verbal provision of negative evidence. Negative evidence, in turn, is defined as “the type of information that is provided to learners concerning the incorrectness of an utterance” (Gass 2003: 225) – in other words, information as to what is not possible, or not deemed acceptable, in a given language. This can be illustrated with the following example from the

8 Potential pitfalls of interpreting data from E-F tandem conversations

SITAF corpus, where an American participant comments on his French partner's renditions of the 'th' sounds:

- (1) NS: The only suggestion that I could make for you was the /θ/ sound [...]. I could completely understand you, and everyone else could, but... erm... instead of ['zi] it's /'ði:/.

Here, the L2 English learner gets corrected on a pronunciation issue which, by the NS's own admission, did not cause any communicative turbulence.

By definition, communication breakdowns do hamper communication, at least at some point in the conversation. Our conception of CBs includes all cases where the listener has difficulty or is incapable of grasping the meaning of an utterance as seemingly intended by the speaker, and makes that difficulty somehow visible or audible. Naturally, what the speaker truly means can be a matter of speculation, although the study of the broader context in which the interaction takes place usually sheds sufficient light on the matter. The following exchange serves as an example of a successfully resolved communication breakdown in our corpus:

- (2) NNS: You know, when people are calling you you are sometimes hungry [* ['hʌŋɡri]].
 NS: [laughing] Wait, the person you are calling is hungry?
 NNS: No, no, no, no, the person which is called.
 NS: Oh, is it angry [hyperarticulation]?
 NNS: Yeah! Sorry, sorry for my accent.

As example (2) demonstrates, communication breakdowns arising from the speech of a NNS will, to a large extent, also involve corrective feedback. Very often, a CB instance will actually trigger an input-providing corrective sequence, such as the hyperarticulation of the mispronounced adjective above. However, because of this extra load brought about by unintelligibility, we deemed it appropriate and informative to conduct a separate, additional analysis of CB instances, which followed a different coding protocol. The reasons for this decision are revisited later in this section, as well as in sections 33 and 5.1.

Our working definition of a communication breakdown, given above, is a broad one. It incorporates cases which other scholars may also term "misunderstanding" (e.g., [Mauranen 2006](#)) or "miscommunication" (e.g., [Dascal 1999](#)), and also includes non-understanding (as defined e.g., in [Jenkins 2000](#)). In a similar vein, we adopt an all-embracing definition of intelligibility. We follow [Bamgbose \(1998: 11\)](#) in taking it to mean "a complex of factors comprising recognizing an

expression, knowing its meaning, and knowing what that meaning signifies in the sociocultural context”. Other linguists, however, make a distinction between those three aspects, labelling them as intelligibility, “comprehensibility”, and “interpretability”, respectively (e.g., [Smith & Nelson 1985](#); [McKay 2002](#)).

In general, breakdowns are more common in NS-NNS than in NS-NS dyads due to the fact that NSs and NNSs “may have radically different customs, modes of interacting, notions of appropriateness and, of course, linguistic systems”, which renders them “multiply handicapped” in interactions with one another ([Varonis & Gass 1985](#): 327, 340). NS discourse may present processing challenges to the NNS interlocutor, for example by virtue of showing insufficient accommodation to the needs of the latter. Embracing these needs ideally means avoiding “slang, opaque idioms, rapid speaking rates, and culture-specific references” ([Trudgill 2005](#): 82). On the other hand, among the major difficulties inherent to NNS output, one can invoke their insufficient mastery of the linguistic system (however that mastery, or indeed the linguistic system to be mastered, is defined), which may result in what [Varonis & Gass \(1985: 334\)](#) term “noise” in the speaker’s utterance, produced by, for instance, accent or ungrammaticality. This, in turn, will often act as a trigger of a corrective episode.

The main rationale behind examining CF instances, with special attention given to L2-speech-induced communication breakdowns, is their potential for carrying pedagogical implications. Establishing the types of non-targetlike linguistic structures that tend to invite corrective feedback, especially if they contribute to communication breakdowns outside of the classroom, could inform L2 teaching priorities (here: for English and French). Assuming that rendering L2 speech communicatively effective is a top priority in most types of L2 instruction, attempts to identify the types of errors which compromise intelligibility hardly need vindication. On the other hand, certain non-standard productions do not lead to communication breakdowns but still trigger corrective feedback from the interlocutor, as shown in example (1). They may therefore be argued to also merit special pedagogical attention, although possibly less so than those non-target forms that are communicatively more salient. In analysing both types of sequences, however, the key problem we have had to address is data ambiguity. This stems from the fact that the interlocutors’ intentions and motivations – unlike the literal meaning of their utterances – are often far from evident, even when considered within a larger context and supplemented by visual cues. This is compounded by the conversational nature of our data, where a simple confirmation check or a genuine question on the part of the NS may easily be misinterpreted as an interrogative recast (i.e., CF) or even a sign of non-understanding.

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

The following sections will give more details on the SITAF corpus (Section 2) before offering a literature review (Section 3) and presenting our analyses of overall CF and then CBs found in the data (sections 4 and 5, respectively), homing in on various dilemmas we have encountered while coding and subsequently interpreting our findings (Section 6).

2 The SITAF tandem corpus

The SITAF corpus is a bilingual database of tandem exchanges collected at the University of Paris 3 – Sorbonne Nouvelle in 2013. The corpus, described at length in [Horgues & Scheuer \(2015\)](#), consists of around 25 hours of audio- and video-recorded, face-to-face interactions held by 21 pairs of native-French speaking and native-English speaking partners. The participants were all students at our university, aged between 17 and 22, none of whom were balanced English-French bilinguals.

To the best of our knowledge, no video corpus of spoken, face-to-face tandem exchanges had previously been compiled. The available language tandem corpora have mainly focused on written L2 production and/or made use of technology-mediated forms such as e-tandem or, more generally, telecollaboration (e.g., [O'Rourke 2005](#); [O'Dowd & Ritter 2006](#); [Ware & O'Dowd 2008](#)). Filling this gap, especially in terms of collecting real-time spoken material illustrating English/French tandem exchanges,¹ was the overall primary objective behind the SITAF project. From that perspective, we believed the SITAF corpus would have three principal assets. Firstly, with the added benefit of video recording, it allows for multimodal studies of real-time interactional phenomena, including non-vocal ones such as gestures or facial mimicry. Secondly, it provides a rich and comprehensive collection of speech data: apart from the NS-NNS exchanges, which constitute its crucial part, the corpus also contains L1-L1 data in both English and French, produced by the same set of participants. Each speaker therefore contributes three types of speech: L1 in an interaction with a fellow NS, L1 in an interaction with a NNS, and also L2 use in an interaction with a NS. In addition, the speech tasks and corresponding speech styles are varied, ranging from semi-spontaneous conversation (both narrative and argumentative) to text reading (see below). Thirdly, the SITAF database is longitudinal, allowing for the observation of a potential evolution in a learner's linguistic output and/or partners' interactional strate-

¹The choice of the English/French combination stemmed from the fact that the project was led by researchers from the English department at our university, i.e. a French university.

gies, possibly affording more insight about language and communication development during tandem learning.

The candidates for the SITAF project were all recruited on a voluntary basis, as part of an optional programme of autonomous tandem exchange run throughout the second semester of the academic year 2012/2013. The recruitment was performed with the help of an on-line questionnaire, which aimed to gauge – through self-assessment – dimensions such as their linguistic background (all languages spoken), level of proficiency in English (for NSs of French) or French (for NSs of English), as well as matters like interests and preferences regarding potential conversation topics and special requests as to their ideal tandem partner. Aside from the researcher's need to establish the participants profiles with a view to interpreting and qualifying future findings, this information was deemed vital in the context of the pairing-up task, performed by the SITAF team members prior to the introductory meeting, during which all participants met their suggested partners. 45 tandem pairs were formed in this way. Of those, 25 subsequently took part in the first recording session, and 21 went on to attend the second session three months later (that is, completed the entire cycle). It is the data obtained from those 42 speakers that makes up the central core of the corpus and that the present study is limited to. The remaining pairs either did not respond to our invitations to the recording studio, or were unable to participate.

The 21 native French-speaking students (subsequently coded F01 to F21, to ensure anonymity) were English language majors for the most part, with a self-assessed level in L2 English of 7.2/10 for their mean proficiency and 6.8/10 for oral expression in particular. The 21 English-speaking students (coded A01 to A21) studied various disciplines and came from various Anglophone countries (United States, Canada, United Kingdom, Ireland). They self-assessed their level in L2 French as 6.9 out of a maximum of 10 for the mean proficiency and 6.6/10 for oral expression.² The above scores were crucial factors informing our pairing-up decisions, as we aimed to match candidates with similar self-assessed L2 proficiency levels, even if those were not necessarily expected to be very accurate reflections of the participants' actual abilities. The significance of proficiency pairing is acknowledged in the research on collaborative learning (e.g., *Storch & Aldosari 2012*), although tandem learning, again, does not present a standard case here, in that two different L2s are involved in each pair. Still, our rationale

²Out of the five self-evaluations (oral expression, oral comprehension, written expression, written comprehension, mean score), oral expression was the only score where the difference between the two speaker groups – native English vs. native French – reached statistical significance, with the latter group reporting a higher level of L2 proficiency ($p < .05$).

8 Potential pitfalls of interpreting data from E-F tandem conversations

was that a stark imbalance between the partners' L2 skills might lead to unnecessary frustrations on the part of one or both participants, which we preferred to avoid. All 21 native French speakers were female, whereas the Anglophone group consisted of 16 female and 5 male members.

The speakers were recorded on two occasions – in February (Session 1) and May 2013 (Session 2). Needless to say, in keeping with the principle of autonomy (e.g., [Brammerts & Calvert 2003](#)), the tandems were free to meet as often as they wished outside of the recording sessions. The questionnaires that the speakers filled out after completing the entire recording cycle suggest that the average tandem had met 12 times over the 3-month period in question, in line with the programme's recommendations for weekly conversations autonomously planned by the tandem partners. However, the individual numbers ranged from two to 23 meetings, thus pointing to substantial variation among the pairs.

Predictably, one of the dilemmas we faced while developing the experimental design was how to strike a balance between spontaneity and homogeneity of the data to be sampled. The latter quality is particularly valuable in the case of pronunciation studies – the main language area the authors specialise in – where having control over the phonemic makeup of the utterances greatly facilitates the researcher's subsequent analyses. As a result, we settled on three types of collaborative tasks, which came with a uniform set of written instructions in the participant's L1, to make sure each pair followed the same protocol. Two of them were communication activities: *Liar-Liar* (Game 1; expected to elicit a narrative style and the most spontaneous speech out of the three tasks) and *Like Minds* (Game 2; debating style with a pre-determined topic). The last one was a partially monitored reading task. In Game 1, the L2 learner had to tell a story containing three lies that the native-speaking partner had to identify by asking questions. In Game 2, both participants had to give their opinion on a potentially controversial subject – e.g., "Prisoners should not have the right to vote" – before assessing the degree of like-mindedness (in other words, convergence of opinions) between them. As regards potential metalinguistic interventions on the part of the NS during Game 1 and Game 2, the guidelines given to the Anglophones read: "When your partner speaks in English, let them do so as much as possible. However, feel free to help or correct them if they can't find the right word or expression, or if you think what they are saying needs correcting". The French participants were instructed accordingly about the French tasks. The text used for the reading task was "The North Wind and the Sun" ("La bise et le soleil" in French), which is a reference text in studies on phonological variation. The NNS speaker read the text twice. The first reading encouraged help and feedback from the NS partner (hence *monitored reading*), whereas no interruption was supposed to occur

during the second reading, which immediately followed.

We insisted on separating and balancing the use of the two languages, in that the entirety of the spontaneous tasks – i.e., both Games – had to be performed first in English and then in French, or vice versa.³ For the most part, the two recording sessions followed the same pattern, outlined above. However, Session 1 actually started with L1-L1 interactions (Games 1 and 2) before moving on to the L1-L2 exchanges, and Session 2 ended with text reading performed by the NS partner. Also, care was taken that in Game 2 each tandem discussed a different topic in each recording session and in each language condition. This was meant to ensure the novelty of the opinions being confronted, and therefore to promote a higher level of engagement of the participants.

Since the central focus of this chapter is data ambiguity, it is the findings from the communicative Games 1 and 2 that are discussed in the following sections. The reading task, being scripted and therefore essentially lacking in spontaneity, was deemed much less suitable for this type of analysis.⁴ It could be described – following Long (1991) – as a focus-on-forms task, since it is the linguistic (and more specifically, the phonetic) form of the learner’s output that the activity almost exclusively focuses on. Game 1 and Game 2, on the other hand, fall under the category of focus-on-form tasks, which are primarily concerned with communication, but which are nevertheless punctuated by the participants’ attending to linguistic issues (i.e., engaging in a Language-Related Episode (LRE)). This ties in with Loewen’s (2018: 2750) definition of focus-on-form practices as ones consisting of “primarily meaning-focused interaction in which there is brief, and sometimes spontaneous, attention to linguistic forms”. The following section presents a brief review of some of the relevant studies on LREs, and more specifically, CF. Consequently, it also outlines our framework for studying communication breakdowns.

3 LREs, CF and communication breakdowns

In this section, we attempt to clarify the relationship between Language-Related Episodes (LREs), corrective feedback (a subset of LREs), and communication breakdowns (to a large extent, a subset of CF).

³In principle, we alternated between English-first and French-first (i.e., every other tandem started in English and the others in French).

⁴Naturally, this is not to say that the interpretation of the reading data does not pose problems of its own (some of those are discussed in Horgues & Scheuer (2014)). However, the nature of the task and the issues associated with it is sufficiently different to warrant a separate treatment.

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

Following Swain & Lapkin's (1998: 326) classic definition, an LRE is understood to be a part of interaction during which the participants "talk about the language they are producing, question their language use, or correct themselves or others". A few years earlier, the same authors offered a slightly differently worded definition, which explicitly stated that each LRE "is related to a *problem* the student had with the production of the target language" (Swain & Lapkin 1995: 379, *italics added*), thus pointing to the fact that the main driving force of such episodes is a potential gap between the target form and the form actually produced, or the absence of the latter. It is this original definition that is the default in the present chapter. So far, LREs have often been studied in classroom settings, during collaborative tasks performed by learners having the same L2 (e.g., Storch 1998; Basterrechea & García Mayo 2013; Basterrechea & Leiser 2019). As for studies of LREs in expert-novice interactions, which may not be the default LRE experimental context but which are of direct relevance in this chapter, CF either features prominently (e.g., Ballinger 2012) or is all but equated with LREs (e.g., Ware & O'Dowd 2008). Ballinger (2012: 79-80) clarifies that "all CF can also be categorized as LREs" but, in her study, giving and receiving CF (as well as partner-directed questions) were analysed separately from LREs in general "because they were deemed the most important for the promotion of collaborative interaction and for reciprocal language learning". This approach is replicated in our studies. Our decision to carry out a separate analysis of communication breakdowns, even though CBs in the context of L2 speech largely fall under the CF umbrella, follows the same logic. Communication breakdowns provide data deemed invaluable for the understanding of how NS-NNS (un)intelligibility works, and – consequently – for reciprocal communicative language learning and teaching practices. We therefore believe they occupy a pre-eminent position within CF episodes.

The importance of supplying CF, in one form or another, and the ways in which it can be beneficial to the L2 learner have been a recurrent theme in second language acquisition (SLA) literature, Nassaji & Kartchava (2017; *In press*) being examples of entire volumes devoted to the subject. Ellis (2017: 4), drawing on an early article by Hendrickson (1978), groups the key aspects of CF, both in terms of teacher experience and research findings, under the following five headings:

1. Should learners' errors be corrected?
2. When should learners' errors be corrected?
3. Which errors should be corrected?
4. How should errors be corrected?
5. Who should do the correcting?

The bulk of the findings from CF studies fall under one or more of the above categories, and they are briefly discussed in the following paragraphs.

First, the question as to whether learners' errors should be corrected at all. As Ellis (2017) points out, contrary to some suspicions expressed by the advocates of certain language teaching methods such as the Audiolingual Method or the Natural Approach,⁵ there has now been a wealth of research showing that CF does assist L2 acquisition (e.g., Li 2010; Lyster & Saito 2010). As for the theoretical grounding of the benefits of CF, Saito (In press) attributes them to CF's "ability to promote learners' awareness, noticing and understanding of linguistic form, especially when using their L2 for meaning conveyance".

The question of the timing of CF has been scarcely investigated, although immediate feedback appears to be preferred on theoretical grounds, for instance by virtue of providing the learner with a window of opportunity during which to map a specific form onto the meaning conveyed (Doughty 2001 cited in Ellis 2017: 7).

To date, most studies have investigated CF provided by teachers on L2 morphosyntax and vocabulary (Lyster & Ranta 1997; El Tatawy 2002; Mackey 2006; Lyster et al. 2013; Kartchava 2019). Meanwhile, some other studies have pointed to pronunciation and vocabulary CF being more noticeable for learners than morphosyntactic CF, which was found to be less likely to lead to uptake (Mackey et al. 2000; Saito & Lyster 2012; Saito In press). Learner uptake is understood, following Lyster & Ranta (1997: 49) definition, as "a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way" to that feedback.

As regards the question of how errors should be corrected, several typologies of CF strategies have been proposed with a view to establishing what the most frequent and the most effective type(s) are. Generally, researchers have classified these strategies on a continuum ranging from the most explicit to the most implicit CF (see Sheen (2006) for a relevant discussion). Lyster et al. (2013) distinguish strategies that offer negative evidence only: prompts (which include – with increasing explicitness – clarification requests, repetition of learner error, paralinguistic signals, elicitation and metalinguistic clues) from those which offer both negative and positive evidence: reformulations (including – with increasing explicitness – conversational recast, didactic recast, explicit correction, explicit correction + metalinguistic explanation). Lyster & Ranta (1997: 46) classic definition of a recast describes it as a strategy involving "the teacher's reformulation of

⁵For example, the Audiolingual Method believed in "strict control of learner output, thus removing the need for CF, which was viewed as a form of punishment that can inhibit learning" (Ellis 2017: 4).

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

all or part of a student's utterance, minus the error". In terms of the relative effectiveness of CF techniques (in the sense of being beneficial to L2 learning), studies have so far yielded variable results. [Kartchava & Ammar \(2014\)](#), for example, set out to determine the relative effectiveness of recasts, prompts and combinations of the two, in terms of both CF noticeability and L2 learning. The study was conducted on selected morphosyntactic structures in a classroom setting. Rather predictably, recasts proved to be the least noticeable of the three, although no significant differences across the CF types were found in terms of learning outcomes. [Sato & Loewen \(2018: 514\)](#) comment on previous studies by observing that "at least in the classroom setting, output-prompting corrective feedback has been found to better facilitate L2 development, compared to input-providing corrective feedback", e.g. recasts. This was corroborated by their own study, where the former category was found to be more effective than the latter. There, however, the effectiveness of both types of CF was mediated by the linguistic structure concerned. The superiority of the output-prompting over the input-providing type was found to be statistically significant only in the case of the more perceptually salient structure under consideration (possessive determiners versus third-person singular -s). [Saito \(In press\)](#) summarises the available research by stating that explicit/output-prompting feedback may be particularly effective in a classroom context, whereas in laboratory settings, "where L2 learners can receive individualized attention from their interlocutors, all CF techniques seem to be equally salient and effective".

The last question on [Hendrickson's](#) (1978) list – "Who should do the correcting?" – has not as yet received a straightforward answer either. Even though most CF studies to date have looked into CF provided by teachers, the benefits of peer feedback – in the sense of learner-to-learner exchanges – have been receiving more and more attention in recent years (e.g., [Adams 2007](#); [Sato & Lyster 2012](#); [Sato 2017](#)). [Sato \(In press\)](#) observes that learners feel more comfortable working on a task with their peers compared to with the teacher or a native speaker. This is conducive to producing a higher amount of output, which in itself is of benefit to L2 learning. On the other hand, learners may not feel comfortable providing CF to their classmates, since that may be considered a socially inappropriate, face-threatening act (e.g., [Foster 1998](#); [Ballinger 2015](#)). What is more, even if peer CF is provided, its quality may be problematic and its quantity insufficient. [Mackey et al. \(2003\)](#), for example, draw attention to its possible shortcomings on both counts (quality and quantity), even though other researchers have obtained results suggesting a longer lasting effect of peer, as opposed to teacher, CF (e.g., [Sippel & Jackson 2015](#)). Importantly, [Sato \(2017\)](#) links the nature and effectiveness of peer CF to the extent to which the social dynamics between the peers

are collaborative⁶. If the learners fail to construct a collaborative relationship, there may be “social awkwardness in providing feedback, and embarrassment over being corrected by their peers” (Sato 2017: 27).

To conclude the above discussion, we may state that CF is a highly complex issue where no overall ideal strategy might necessarily be identified (e.g., Lyster et al. 2013), even though its benefits to L2 learning have been well documented, and “[l]earners almost invariably express a wish to be corrected” (Sheen & Ellis 2011: 606). This review of studies on corrective feedback serves as a backdrop against which to view research into the CF found in the SITAF database, summarised in Section 4.

4 Research on CF in the SITAF corpus

This part presents the main results of the CF analyses carried out so far on the SITAF data, before moving on to some of the key methodological challenges encountered in the process. More specifically, Section 4.1 expounds the criteria and parameters according to which we coded each corrective episode, and Section 4.2 offers the main findings in relation to those same parameters. Both sections set the scene for the discussion of coding issues that follows in 4.3.

4.1 Computing and coding CF episodes

Using the definition given in Section 1 as a starting point, we employ the term CF to refer to the verbally expressed negative evidence given by the NS participant to their NNS tandem partner during the recorded interactions. Naturally, “correcting others” (as per Swain & Lapkin’s 1998 definition) is just one among several possible types of LREs to be explored in the SITAF corpus, but it is the one that this chapter focuses on, with special attention given to communication breakdowns. Language-related episodes revolving solely around positive feedback (acceptance or acknowledgement of the correct form produced by the learner) or self-corrections, for example, are not discussed here.

The video recordings corresponding to the two communicative activities – Game 1 and Game 2 – in both recording sessions and in both languages were examined for the occurrences of CF. The simultaneous visual and auditory analysis was conducted by the two authors, who split the work but consulted one another (and, if necessary, other team members) about difficult or dubious cases

⁶In his discussion of the social dynamics between peers, Sato draws on Storch’s model of dyadic interactions classified along the dimensions of equality and mutuality (e.g., Storch 2002).

8 Potential pitfalls of interpreting data from E-F tandem conversations

and subsequently reached a consensus. Each CF occurrence thus identified was annotated and coded according to at least four parameters. This coding protocol expanded on some of Hendrickson's (1978) categories discussed in Section 3, notably aspects 3 and 4: "Which errors should be corrected?" and "How should errors be corrected?". The four parameters were the following:

1. CF focus: morphosyntax, pronunciation, vocabulary, or mix of any of the three
2. CF strategy (see below)
3. Presence or absence of CF request: whether the feedback was somehow solicited by the learner (explicitly or implicitly) or it was offered spontaneously
4. CF uptake: total, partial, failed or none.

Furthermore, CF episodes contributed by selected tandem pairs were also coded according to the multimodal resources employed by the participants, for instance types of gestures or specific vocal non-verbal content (hyperarticulation, rising tone, etc.).

As regards CF strategies, we have simplified and tailored Lyster et al.'s (2013) typology, presented in Section 3, to better fit the context of tandem exchanges. Importantly, we do not make use of categories such as elicitation or repetition of the learner's error, which are absent from our peer-to-peer interactions. These CF strategies seem to be restricted to teachers' corrective style and avoided by tandem participants possibly because they reinforce the asymmetry between the two partners.

Consequently, we have distinguished three basic CF strategies in our analyses:

1. Explicit comments, where the NS provides metalinguistic information explicitly, as in example (3) below, where the native English speaker is explaining to her unconvinced French partner why she made a mistake when she used *sunbath* as a verb:

(3) NNS: 'Cause you say bath.

NS: Right, bath, but then to bathe oneself, so sunbathe.

2. Clarification requests, as in example (4) below, where the NS reacts to her American partner's use of the word *flip-flops* in the middle of a French sentence:

Sylvia Scheuer & Céline Horgues

- (4) NS: Les flip-flops, c'est quoi? (French)
 'What are flip-flops?' (*flip-flops* not being a French word)

3. Recasts, e.g.:

- (5) NNS [talking about a past event]: And I miss my plane.
 NS: You missed your plane?
 NNS: Yeah, yeah.

Such input-providing corrections as recasts – unlike output-prompting strategies – effectively supply the novice “with reformulations in response to their errors, thereby providing positive evidence, that is, linguistic information about what is allowed in the target language” (Sato & Loewen 2019: 32). Yet, their *raison d'être* is the provision of *negative* evidence in the sense of signalling the incorrectness of the learner's output.

4.2 Summary of the main findings

The most comprehensive report to date on the corrective feedback found in the SITAF database is provided in Scheuer & Horgues (2020). The summary offered below is organised according to the parameters outlined in Section 4.1, i.e.: CF focus, strategy, presence or absence of request, uptake and multimodality.

All in all, we have identified 492 CF instances in the approximately 15 hours of conversational exchanges held as part of Game 1 and Game 2 in both recording sessions. Of those, 156 were found in the English, and the remaining 336 (i.e., over twice as many) in the French part of the data. The primary focus of the corrective interventions is vocabulary: about half of the CF instances found in the conversations – 52.5% in English and 49% in French – target missing or incorrectly used words, expressions or collocations. In the case of the French data, vocabulary errors also include the wrong grammatical gender, as in:

- (6) NNS: Pour mes, ma Noël... (French)
 For my.PL my.F.SG Christmas
 NS: Ton Noël.
 your.M Christmas
 NNS: Je veux parler de ma Noël.
 I want to talk about my.F.SG Christmas

The runner-up category in English was pronunciation, which accounted for 20% of all CF instances (e.g., the NS recasting the incorrect stress pattern in his partner's rendition of *pri'soners), followed by morphosyntax with 12.5%, as in:

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

(7) NNS: Well, it depends of... the crime.

NS: On the, yeah.

NNS: On? It depends on the crime.

In French, morphosyntax ranked second, with 19% of hits, compared to 15% garnered by pronunciation. The remaining corrective episodes (15% in English and 17% in French) were classified as having mixed focus, as they revolved around learner utterances that were erroneous in more ways than one. These will be discussed further in Section 4.3.

As regards the corrective strategies employed by the SITAF participants, by far the most common one was recast, which accounted for 84% of CF in the English, and 89% in the French conversations. The remaining cases were split almost equally between explicit comments and clarification request.

Feedback was solicited by the learner roughly as often as it was not (i.e., the NS intervened unprompted in nearly 56% of the English cases, and just over 47% of the French ones). An example of solicited CF, which took the form of an explicit comment, is given in (8):

(8) NNS: And after I celebrated the happy new year, you know [...] the new year, not the happy new year, yeah?

NS: New year [head nod]. “Happy New Year” is what you say!

The two extremities of the uptake spectrum – total uptake and no uptake – jointly account for nearly 90% of all CF episodes in the two languages. However, there is a sharp difference between the English and the French conversations in terms of the relative share of each of the most frequent categories. In English, total uptake (shown in example 7) occurred in just 36.5% of cases, whereas no uptake followed 52.6% of the corrective interventions, with the remaining cases representing either partial or failed uptake. The French figures are almost identical but in reverse order: 52.4% for total, and 36.9% for no uptake. Example (6) illustrates the latter category: the recast of ‘Noël’, with the gender-appropriate possessive determiner, does not affect the NNS’s subsequent utterance in any way.

Finally, corrective episodes proved to be highly multimodal activities. 94% of the CF occurrences studied were multimodal (i.e., combined verbal, vocal and visual resources). The remaining 6% were verbal and vocal only (Debras et al. 2015).

4.3 Some methodological issues encountered

The results presented in the previous section are the fruit of an analysis that was rich in methodological challenges. Whether or not a given episode constitutes an occurrence of CF is not always a straightforward matter. Furthermore, the exact strategy employed by the expert and the focus of their corrective intervention can also be hard to determine. Below we discuss and give examples of some of the challenges we faced while coding the data.

As could be expected from conversational data, the exact nature and purpose of the participants' output is not always easy to establish, and "the categorization of an utterance can be ambiguous since researchers are not privy to the speakers' intentions" (Ballinger 2015: 44). Example (5), where the NS produced an echo question albeit with the correct(ed) grammatical tense, serves as a good illustration of the most fundamental methodological issue we have been grappling with when coding CF instances in the SITAF database: deciding whether we are dealing with a corrective intervention in the first place. Recast is, by definition, discreet, indirect and non-threatening, not to mention easy to dispense. As such, it is particularly well suited to the tandem type of peer-to-peer interaction, where neither partner tends to particularly want to reinforce their short-lived dominant position. Therefore, it comes as no surprise that the vast majority of what we ultimately considered to be CF endeavours on the part of the NS (including example 5) were carried out by means of recast. However, precisely because of this discreetness, recasts may easily be misconstrued as the interlocutor's innocuous contribution to the activity. The problem is exacerbated by the conversational nature of our data (i.e., the fact that the participants engage in a genuine exchange of stories and ideas, where backchannelling, confirmation checks, repetitions and reformulations are common in both L1-L1 and L1-L2 interactions). The issue – although viewed from the perspective of the addressee of the hypothetical CF – has been highlighted by a number of researchers. Carpenter et al. (2006: 209), for example, observe that "recasts might be ambiguous to learners; that is, instead of perceiving recasts as containing CF, learners might see them simply as literal or semantic repetitions without any corrective element". This ties in with Sato & Loewen's (2019: 38) definition of a conversational recast as any teacher response to an error that includes "the correct linguistic form without any emphasis". In the context of tandem exchanges, where the NS partner is not a teacher but rather an empathetic peer, the remedial element of a conversational recast may be all the easier to miss.

In the short tandem exchange cited in (5), the NS's question, which we labelled as a recast, might have been nothing more than a confirmation check or a com-

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

miserative reaction triggered by the news of her partner having missed her plane. Those are particularly valid assumptions in the case of a communicative task like our narrative Game 1 (during which the exchange took place), where the NS listener was meant to make a mental note of the details of the NNS partner's story and was therefore expected to try to verify the information given in that story. The problem is also evident in the following example, where the French speaker erroneously pronounces the word 'castle' with a [t]:

- (9) NNS: My aunt organised a big party in a castle,
 NS: OK
 NNS: with all the family, with the cousins...
 NS: In, in a castle, you said?
 NNS: Yeah, a little casTle [gesture representing a castle] ...
 NS: OK [smiles].

Clearly, the NNS takes her partner's question at face value, and subsequently provides a gesture-enhanced confirmation – “yeah, a little casTle” – where she repeats her original error.

Needless to say, if recasts are ambiguous to the potential recipients, they may also be ambiguous to the researchers analysing the exchanges, because the researchers are unable to tap into the speakers' mindset. What they do have access to, however, is the remainder of the conversation, which may, retrospectively, shed more light on the NS's intentions. For example, knowing that the American speaker quoted in (9) tried to elicit the word *castle* from his partner later in the conversation gives us extra reassurance that his “in a castle, you said?”, uttered 3 minutes earlier, was indeed an interrogative recast rather than a genuine question. Still, our decisions to classify many cases like the above as CF attempts, although often reinforced by the study of the surrounding context, are nevertheless not entirely unquestionable. Ultimately, whether the NSs in (5) and (9) were actually trying to correct their partners' pronunciation and/or syntax, rather than simply making sure they had correctly understood the discourse, will never be established beyond doubt. After all, it is not every day that one misses a plane or one parties in a castle.

On the other hand, such potential mistakes, where a CF label was applied to an utterance devoid of corrective intention, may have been offset by a number of instances where the opposite error may inadvertently have been committed. The error would have consisted in actual recasts being miscoded as conversational turn-taking. In particular, this was likely in cases where the NS repeated all, or some, of the NNS's utterance word-for-word, phoneme-for-phoneme, as in the following exchange:

Sylvia Scheuer & Céline Horgues

- (10) NNS: Il y avait trois. (French)
 ‘There were three’.
 NS: Trois, d’accord.
 ‘Three, all right’.⁷

Much as the above looks like an innocuous repetition, confirmation or simply an effort to maintain the conversation flow, however one cannot rule out the possibility that the expert was, in fact, recasting a suprasegmental or sub-phonemic detail of the non-native pronunciation (e.g., the /r/ in ‘trois’ /trwa/, which the NNS all but deleted) that they judged worthy of discreet correction. Technically, the NSs’ interventions like the one shown in (10) do satisfy our broad definition of recast – repeating all or part of the novice’s utterance minus the (pronunciation) error – even though the presence of actual corrective intention is far from evident.

Even if a CF instance has been rightly acknowledged as such, this does not automatically mean its classification – both in terms of the exact CF strategy employed and of its intended focus – is a clear-cut matter. Both aspects may be complex to interpret since a correction may involve multiple moves (e.g., a recast followed by a clarification request) and may target various linguistic levels (phonology, lexis, morphosyntax) at once. While coding errors relating to a CF strategy can be argued to be relatively inconsequential, failure to identify the linguistic trigger of the NS’s remedial reaction could potentially skew certain pedagogical implications that the researcher may wish to glean from the CF analyses. The validity of such implications hinges on determining a causal link between the specific characteristics of the NNS’s output and the NS’s corrective intervention. Assuming that the experts are relatively likely to intervene when the novice’s mistake is prominent in terms of compromising communication (see Section 5) or violating a norm that they consider important, being able to dissect the nature of that mistake is of great potential value to language teachers. Namely, it helps pinpoint the types of interlanguage errors that bother NSs more than others, which might, in turn, inform teaching priorities. Given that not every aspect of the L2 system can possibly be accorded the same amount of time and attention in the L2 classroom, being selective about what to teach first and foremost is a sheer necessity. Even the most committed language instructor has therefore to make choices, which may well be guided by the perceived

⁷In fact, the correct version of the NNS’s utterance would have been ‘il y *en* avait trois’ [‘there were three of them’], but this syntactic inaccuracy clearly did not bother the French NS enough to rectify it.

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

gravity of non-target forms. If so, examining some of the factors which might potentially determine this gravity seems like a worthwhile endeavour.

This endeavour is certainly not without its problems. One of the pitfalls inherent in attempting to establish relative error gravity is confusing different dimensions, such as the dimension of accuracy with that of communicative effectiveness. Namely, it may be tempting to conclude that an error which may potentially lead to a communication breakdown is somehow more ‘erroneous’ than an error that does not jeopardise intelligibility. Meanwhile, as pointed out by Pallotti (2009: 592), both types of error have the same impact on the accuracy of an utterance, since “a 100-word production with 10 errors not compromising communication is not more ‘accurate’ than a text of the same length with 10 errors hindering comprehension, but just more ‘understandable’ or ‘communicatively effective’”. Nonetheless, it can be argued that the types of errors that, for whatever reason, tend to command the NS’s corrective attention also deserve special pedagogical attention, at least in contexts where the learner is consciously oriented towards the NS model, as is indeed the case with tandem exchanges. To borrow Pallotti’s logic: a 100-word production with 10 errors not bothering the listener is no more ‘accurate’ than a text of the same length with 10 errors bothering the listener, but it nevertheless ranks higher than the latter on the dimension of acceptability, which is no small matter.

Since the vast majority of CF instances were performed by means of recast and the NNSs’ utterances were often incorrect in more ways than one, it comes as no surprise that the specific motive behind the NSs’ corrective intervention was not always evident, neither to the recipient, nor subsequently to the researcher. Example (11) is a case in point:

(11) NNS: There were interior in leather [* [’li:ðər]].

NS: Oh, the leather, oh, so there’s leather interior.

NNS: Yeah, because in Ferrari there’s leather [* [’li:ðər]] in the car.

NS: Yeah, yeah, it’s just what you said [gesture representing switching] you, just switch the words so it’s leather interior, not interior leather.

The American speaker first uses a recast (“leather interior”), then explicitly insists on the correct word order (“just switch the words”), then praises his French partner for eventually getting the order right, and subsequently resumes the flow of the conversation. However, throughout this episode, the French speaker persists in her erroneous rendition of the vowel, thereby only correcting her original output to “leather [* [’li:ðər] interior”. It is impossible to know to what extent the pronunciation problem troubled the NS and whether it would have triggered his

Sylvia Scheuer & Céline Horgues

corrective intervention at all, had it not been accompanied by the syntactic error. The fact that he chose not to revisit the NNS's utterance once the syntax had been fixed does not necessarily mean the wrong vowel was not considered an issue. Rather, the expert might have chosen not to overwhelm his partner with too many corrections directed at one short utterance, which might have been unhelpful to her L2 acquisition process (cf. Ellis et al. 2008, in the context of written CF).

Unlike the example above, there were numerous cases where a recast was not accompanied by an explicit comment. These instances were more enigmatic and therefore more problematic. This is illustrated by the following exchange:

- (12) NNS: We putted our ski.
 NS: [nodding] Put your skis on.
 NNS: And we ...

Here, the exact reason for the NS's remedial reaction – was it the wrong past tense verb form, the missing particle, or the missing plural marker? – is unclear, even though it stands to reason that the cumulative effect of the various issues may have been what incited the expert to intervene by recasting all the issues in one move. We believe that our decision to label such occurrences as having a 'mixed' focus (in the example above: a mix of lexis and morphosyntax) has the advantage of making our observations more objective, through minimising the need for the researcher's personal judgement and interpretation of what the speaker truly meant to correct while offering the correction.

Having presented the main results of our studies of general CF in the SITAF corpus, as well as the major problems encountered in the process of obtaining them, we will now turn to the other – related – research focus of direct relevance to the chapter: communication breakdowns.

5 Study of communication breakdowns

5.1 Communication breakdowns versus CF in general

The detailed study of communication breakdowns is a major thread of research to emerge from analysing the SITAF data on the back of CF analyses (e.g., Horgues & Scheuer 2018b). Our working definition of a CB, provided in Section 1, encompasses all cases where the listener demonstrably has difficulty or is incapable of understanding the meaning of an utterance as intended by the speaker. This tallies with Mauranen's (2006: 128) definition of the term "misunderstanding",

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

taken to denote “a potential breakdown point in conversation, or at least a kind of communicative turbulence”. The reasons why CBs receive a separate treatment in our analyses are mostly to do with their particular communicative, and therefore potential pedagogical, relevance. The matter was already addressed in Sections 1 and 3, and is further clarified below.

The relationship between communication breakdowns and CF is not entirely straightforward. CF is often provided even though comprehension is not at stake, as examples (5–12) demonstrate. The reverse is also true: a communication breakdown in NS-NNS conversations may contain no CF overlay at all. This typically occurs when it is the NS’s discourse that is not understood, as in (13):

- (13) NS: On va pas le défendre; on va plutôt le sermonner. (French)
 ‘One will not defend him; one will rather lecture him.’
 NNS: Sermonner, qu’est-ce que ça veut dire?
 ‘‘Sermonner’’ [to lecture], what does it mean?’
 NS: [explains the meaning of ‘sermonner’]

Clearly, there is no corrective intention behind the NNS’s clarification request, and the NS’s subsequent explanation only serves to provide the NNS with positive – rather than negative – evidence, even if the whole sequence uncovered a lexical gap on the part of the NNS. Another type of episode that can be potentially classified as an instance of communication breakdown but not of CF arises when the confused recipient is sending visual signals only. Non-verbal strategies, especially face expressions (frowns, squints) or shifts in gaze, may well be indicative of non-understanding. However, in accordance with our definition given in Sections 1 and 4.1, they do not, in or by themselves, count as CF in the present analysis. Finally, a CB may go undetected by either participant, even though it may be evident to an external observer. In such cases the NS expert, unaware of the true meaning of the NNS’s utterance, will not be able to provide correction. Only one such instance has been identified in the SITAF corpus. In one other episode, a CB very nearly went undetected: the French participant mispronounced the word “tuition” in “tuition fees” so that her NS partner misinterpreted it as “teaching fees” (i.e., teacher salary). A prolonged misunderstanding sequence follows (2’30 minutes long), where the two interactants run parallel but disconnected argumentations (what the students should pay to study versus how much the teachers should get paid). The problem finally gets resolved almost by accident, when the NNS makes a remark – “for studies” – which alerts the NS to the fact that she has been misunderstanding her partner all along:

Sylvia Scheuer & Céline Horgues

- (14) NNS: Here, not a lot of people can afford 400 [euros] per year for, for studies, so [...]
 NS: Oh, hang on [looks at the slip of paper with “tuition fees” printed on it]. Tuition fees, OK. [...] I thought, teaching fees. Instead of tuition fees. [...] I thought the teacher only gets paid 400.
 NNS: [laughter] Ah, no, no [...].

Despite such divergences, there is a considerable area of overlap between communication breakdowns and corrective feedback, in that verbally signalled CBs arising in the context of NNS speech are largely a subset of CF instances. The ‘flip-flops’ example (4) illustrates this point: the very fact that the attentive NS (who, within a tandem setting, will by default be cooperative⁸) does not understand a lexical item will, in most cases, constitute negative evidence already: either the word itself is incorrect, or there is something wrong with the NNS’s rendition of it.

5.2 Computing and coding communication breakdowns

Needless to say, each CB occurrence which also constituted corrective feedback had previously been annotated according to the CF coding protocol (Section 4.1). In addition, each instance of communication breakdown identified in the data – whether or not it coincided with CF – was annotated and coded according to parameters such as:

- Whose speech got misunderstood: NS or NSS?
- Miscommunication trigger: morphosyntax, pronunciation, vocabulary or mix of any of the three?
- Who detected the misunderstanding: the main speaker, the interlocutor or both?
- Timing of detection: instantaneous (next turn) or delayed?

Comparisons were also made between Session 1 and Session 2.

Section 5.3 presents our main findings, along the lines outlined by the above parameters.

⁸The expectation that tandem partners should be cooperative and therefore willing to understand each other stems from one of the fundamental principles of tandem learning, i.e. reciprocity. This means “the reciprocal dependence and mutual support of the partners” (Brammerts 1996: 11). The listener’s willingness to understand the interlocutor, which is not always to be taken for granted, is a crucial factor in mutual intelligibility, highlighted by Chambers & Trudgill (1998: 4)

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

5.3 Summary of the main findings

Since one of our goals was to determine whose output was misunderstood in each case, we deemed Game 1 (storytelling) less suitable for this type of quantitative analysis, given that disproportionately more speaking time was naturally given to one of the participants (the storyteller). Quantifying communication breakdowns encountered in the course of that activity might therefore have skewed the overall results.

Having quantitatively analysed the data from the debating Game 2 (approximately 5h), we identified a total of 72 cases of detectable communication breakdowns in the two language conditions. Of those, 41 were found in the English and 31 cases in the French conversations. A total of 40 (55.6%) arose in connection with the NNS discourse, which means that in the remaining 32 (44.4%) cases it was the NS who was misunderstood or not understood. Vocabulary proved to be the main stumbling block when it came to processing NS discourse (21 cases out of 32; 65.6%), as opposed to pronunciation in the case of NNS speech (14 cases out of 40; 35%). In about two-thirds of occurrences it was the interlocutor (recipient) who signalled the communication breakdown, whereas in the remaining cases the problem was detected by both participants, roughly simultaneously. In keeping with the collaborative spirit of tandem learning, CB detection was largely instantaneous, occurring in the next turn (60 out of the 72 instances), occasionally delayed (11 instances), and missing altogether in just one case. The number of communication breakdowns dropped between the two recording sessions, from 39 to 33, although the difference was only statistically significant in the English conversations (from 26 to 15).

5.4 Some methodological issues encountered

It comes as no surprise that in many respects the methodological challenges posed by the identification and subsequent interpretation of communication breakdowns resemble those encountered while exploring general corrective feedback. Not only can it be difficult to pinpoint the exact cause of a CB, but it is also frequently impossible to determine with a fair degree of certainty that mutual comprehension was hindered.

Our analysis of CB instances has necessarily been confined to cases where a comprehension issue is somehow signalled. The problem is that such signals will inevitably vary in clarity and will therefore be more or less legible to the observer. Example (13) represented the clear end of the spectrum, as does (15), this time with the NS in the role of the non-understander:

Sylvia Scheuer & Céline Horgues

- (15) NNS: Cela rend les gens plus seuls [*[sul]]. (French)
 ‘This makes people more lonely.’
 NS: [at first, silence and blank face] Plus quoi?
 ‘More what?’
 NNS: Seuls [*[sul]].
 ‘Lonely.’
 NS: Ah, plus seuls [[søl]]!
 ‘Oh, more lonely!’

The unambiguous clarification request on the part of the French participant (“more what?”) is a clear sign of her struggling to make sense of her NSS partner’s utterance.

On the other hand, the interlocutor may react to an utterance with signals so subtle as to leave the researcher in doubt as to their true significance, as in (16):

- (16) NNS: If it’s just uh...
 NS: [keeps nodding]
 NNS: if you’re just as thief [*[’vif]]
 NS: [stops nodding]
 NNS: who, who go to prison,
 NS: [now nodding only slightly]
 NNS: maybe you could? [...]
 NS: Uhm. I don’t know where I stand on this.

Here it is the disappearance of non-verbal response on the part of the interlocutor that is suggestive of her confusion: she had been nodding for some time but stopped doing so upon hearing the [’vif] utterance. Yet again, the surrounding context is helpful: the NS’s subsequent verbal contribution, which is rather non-committal and does not build on her partner’s discourse, provides further support for this interpretation.

Apart from being economical with cues, the recipient may also be sending conflicting signals as to whether the tandem partner’s discourse has been understood. The following exchange can serve as an example of this coding dilemma:

- (17) French speaker:⁹ *In French we say familiarly une boîte à fric, je sais pas si [...]*
 ‘a scam, I don’t know if [...]
 American speaker: OK [gazes sideways].

⁹Due to code mixing, it would be confusing to use the labels NS and NNS in this case.

8 Potential pitfalls of interpreting data from E-F tandem conversations

The conversation proceeds in English but, due to a lexical gap, at some point the French speaker resorts to an expression in her L1 (*une boîte à frie*). On a purely literal level, her American partner seems to have no difficulty processing her output (he utters “OK”), but the non-verbal cues he provides tell a slightly different story: his tone of voice is hesitant and he gazes sideways, suggesting that, at least at that particular moment, the exact meaning of the colloquialism *boîte à frie* is unclear or the sudden language switch caught him off-guard.

In addition to the dilemmas outlined above – determining whether a communication breakdown did indeed occur, in the absence of tangible or consistent cues – the researcher is faced with the other major coding issue discussed in the context of CF: what brought the problem about. Identifying the linguistic triggers of communication breakdowns in NS-NNS interactions is potentially of even greater pedagogical importance than is the case with the remaining body of CF. After all, the primary function of language is communication. If that is jeopardised, one is justified in trying to eliminate the source of the problem before moving on to somewhat higher-level considerations potentially triggering CF provision, such as sounding aesthetically pleasing to the listener. Fortunately, the majority of communication breakdowns identified in the SITAF corpus leave the researcher in little doubt as to the linguistic source of the problem. Explicit information is often provided by the participants themselves during the relevant episodes. This is shown in example (18):

- (18) NNS: Et la cousine [*[kyzin]] de de mon père d'accueil n'a pas mangé du veau parce que elle a dit que... (French)
 'And the cousin of of my host father did not eat the veal because she said that...'
 NS: [confused facial expression at first] Ah, la cousine! [...] D'accord, j'avais compris la cuisine; la cousine, OK!
 'Oh, the cousin! [...] All right, I'd understood *the kitchen*, the cousin, OK!'

The NS makes it fairly clear that the problem was her partner's erroneous fronting of the first vowel in the word *cousine* (/kuzin/), which made her perceptually confuse it with *cuisine* (/kɥizin/). Unlike in (15), however, the NS offers her partner (and the researchers) the added bonus of an explicit comment on how exactly she misunderstood the NNS utterance. It is also worth noting that (18) is a representative example of how a communication breakdown may serve as a starting point of a corrective episode, a phenomenon alluded to earlier on in the chapter.

Sylvia Scheuer & Céline Horgues

Despite the prevalence of relatively straightforward (in terms of the linguistic trigger) cases like (18), the cause of the communication breakdown was not always easy to pin down. As a result, in around 18% of instances (13 out of 72) we ended up labelling the CB as being due to a combined trigger, for reasons similar to those mentioned in the context of the mixed focus CF occurrences, and with the same corollary of making our observations less informative than we might perhaps have wished them to be. The NNS in (19), using the expression *for per'petuity* to talk about prisoners sentenced to life in jail is one such instance:

- (19) NNS: And if you are in jail for... I don't know how to say that... for
 **per'petuity*
 NS: What?

Not only is the phrase a calque from French (the word *perpetuity* not being used in this legal sense in English), but the NNS also mispronounces the word by stressing the second syllable instead of the third. The NS does not know what her partner is trying to say (“What?”), but a successful clarification attempt follows. This involves the French speaker first switching to her L1 (“perpetuité”) and then reformulating her initial proposition as “you’re gonna die in prison”. It is perhaps tempting to propose that the pronunciation issue played the key role in generating this temporary communication breakdown. After all, the English word *perpetuity* is not far removed semantically from the French term, so the NS would likely have gotten the idea had she been able to simply recognise the word, but this is impossible to verify.

Other instances where a CB is clearly evident present an even more complex picture in terms of their underlying cause(s). A case in point is a French conversation where what appears to be the keyword in the NNS’s discourse – *le lieu* [/lə lʃø/], ‘the place’ (where she got sick) – is erroneously rendered as *[la ly]:

- (20) NNS: *La lieu [*[ly]] où je suis tombée malade. (French)
 The-FEM place where I got sick.
 NS: [confused facial expression]
 NNS: *La lieu [*[ly]].
 NS: [la ly]???

The cumulative effect of the two issues, the wrong vowel and the wrong grammatical gender, ensures that the NS is at a loss as to what her partner means. She clearly does not understand, which is revealed not only by her confused facial expression and interrogative echoing of the offending sequence, but also by her unsuccessful attempt at paraphrasing it, which follows: “Ah! T’es pas tombée

8 Potential pitfalls of interpreting data from E-F tandem conversations

malade!” [‘Ah! You didn’t get sick!’]. Matters are not helped by the American speaker’s subsequent use of a false friend – *la location* [‘the hire’] – in a bid to clarify the meaning of her original sentence, and the CB remains unresolved. Again, it would be valuable to know whether one of the two mistakes involved in *[la ly] was actually more salient than the other, in the context of establishing error hierarchies. For example, it stands to reason that the mispronounced vowel, but not the incorrect gender, might have single-handedly pushed the native speaker over the line of non-understanding. If so, this would highlight the importance of attending to pronunciation details (here: the exact quality of rounded vowels) in the L2 French classroom, where correct gender assignment might receive considerably and undeservedly more pedagogic attention in comparison. What is worth noting is that no input-providing CF can be given in example (17), since the novice’s utterance remains cryptic to the NS, while being clearly incorrect. The latter adopts a ‘let-it-pass’ strategy and chooses to simply move on, uttering a rather unconvinced “d’accord” [‘all right’] in the process. Her American partner emerges from this episode none the wiser as to the grammatical gender and the phonemic shape of the word *lieu*, and she has reason to believe that her discourse, if not entirely accurate, was at least communicatively effective.

As shown in this section, dissecting the nature and identifying the exact trigger of a communication breakdown may be a highly challenging task, in ways that are similar to those previously discussed in the context of general CF instances.

6 Discussion and conclusion

This final section offers a summary and a further discussion of the methodological issues highlighted in the chapter, the ways in which we have tried to address them, as well as a conclusion hinting at potential future perspectives.

6.1 Methodological challenges encountered

As demonstrated in the chapter, coding LREs such as CF episodes and communication breakdowns occurring in semi-spontaneous NS-NNS interactions is no straightforward task. The participants’ output and reactions can be both complex and ambiguous, often making it impossible for the researchers, or indeed the interlocutors themselves, to perceive and decode the speakers’ intentions with a fair degree of certainty. This ambiguity has manifested itself particularly at the levels of identifying the interactional function of speech turns (e.g., teasing apart corrective sequences from conversational moves such as confirmation checks or

topic continuations) and pinning down the triggers that led to a given CF or CB instance.

The added layer of complexity stems from the fact that the various functions, as well as the various triggers, may appear in combination with one another. Our method of coding the two interactional phenomena – CF and communication breakdowns – acknowledges this complexity. This is reflected in our extensive use of the labels mixed and combined when more than one factor seemed to be at play. However, this kind of cautious coding will inevitably influence and, to some extent, constrain the interpretation of our findings. Since one of our stated objectives in studying CF and communication breakdowns has been to obtain data that could inform L2 teaching priorities, an optimal end result would be to provide unambiguous answers as to what types of non-native productions are likely to cause communicative turbulence. Intelligible speech is certainly one of the most highly desirable learning outcomes in any L2 classroom, which means that those non-target productions that tend to jeopardise intelligibility should perhaps receive the teacher's attention before anything else. On the other hand, non-native output which simply triggers corrective feedback without hindering communication will probably rank considerably lower in that hierarchy, while still being more worthy of remedial action in the classroom than other types of inaccurate productions. For those reasons, mixed and combined CF/CB instances will be of more limited pedagogical relevance, since they will be harder to interpret in terms of specific remedial actions. Sequence (20) may serve as an example here: the fact that the non-target vowel in *lieu* is intermingled with the wrong grammatical gender, to some extent downplays the importance of either issue, as it is uncertain whether either of them would have triggered the communication breakdown on its own. Instead of possibly serving as a prime example of how incorrect vowel quality may single-handedly hamper intelligibility, this instance will lose some of its significance by feeding into the rather fuzzy mixed category.

Another problem with interpreting quantitative findings obtained from a group like ours is the fact that they are generalised across pairs that are far from homogeneous. The social dynamics between the two partners will naturally be slightly different within each tandem, and that will inevitably affect the way CF is dispensed and received and the way communication breakdowns are signalled and resolved. This means that the data contributed by different tandems may not always be directly comparable. As observed by Horgues & Tardieu (2015), certain SITAF participants are hyper-correctors and others are hypo-correctors, and there is no straightforward correlation between the level of L2 competence and the amount of CF received. For example, of the 336 CF instances we identified in the French section of the corpus, one participant (F11) contributed 52 (15.5%)

8 Potential pitfalls of interpreting data from E-F tandem conversations

cases of CF. On the other hand, two other French speakers produced just 1 instance (0.3%) each. Foster (1993: 25) highlights a similar issue in the context of her own study of collaborative tasks performed in an L2 classroom: “The range in the individual scores is so wide, and the lack of participation by some students is so striking as to make statistics based on group totals very misleading”. This is another reason why the interpretation of such group observations in the context of gleaning pedagogical insights should be carried out with utmost caution. Due to idiosyncratic linguistic preferences and individual corrective styles, certain relatively minor inaccuracies might get overrepresented and therefore ascribed disproportionately more importance than they deserve, if they happened to fall on the over-sensitive ears of an over-corrector. F11 with her 52 corrective interventions is a case in point. On one occasion, she corrects a collocation (*heureuse* to refer to *période*) that her American partner has directly copied from the topic the pair was given in writing at the beginning of the task. The topic, which read “L’adolescence est la période la plus *heureuse* de la vie” (‘Adolescence is the *happiest* period of your life’), had previously been prepared and approved by her fellow native French speakers. In other words, the NNS gets corrected on something that for all intents and purposes is correct in L1 French, which might make this corrective intervention appear of little didactic value. On the other hand, it could be argued that it is precisely this sort of rather unexpected and unconventional results that make our findings most interesting. If one takes the participants’ perspective, one gets a chance to see what individual speakers treat as an error, or at least what sort of forms they find annoying and worth eradicating, on top of the “real” errors that one could identify and code by simply referring to handbooks and dictionaries.

6.2 Solutions adopted

In view of the fact that there seem to be no available studies of corpora of video recorded, face-to-face tandem interactions, we have had to grapple with challenges that have not necessarily been adequately addressed in the SLA literature. The frameworks previously developed for analysing LREs in L2 classroom settings do not entirely fit our context of expert-novice, yet peer-to-peer, interactions. Therefore, one of the basic steps we needed to take was to adjust the descriptive categories previously employed in CF studies, to better capture the specificities of our data. Crucially, ours was a setting where there was no need to account for corrective moves characteristic of teacher discourse, but where the roles of (relative) expert and novice within each conversation section were clearly defined.

Sylvia Scheuer & Céline Horgues

While coding conversational data, which is invariably complex and ambiguous, the risk of exercising excessively subjective judgement is ever present. We have endeavoured to minimise the role of this subjectivity through taking various measures. One basic and commonly adopted step, in addition to developing a detailed coding protocol, was to have the particularly challenging cases analysed by two or more team members. The multimodal nature of our data also provided further opportunity to objectivise our analysis. Supplementing the subtle – or even non-existent – verbal cues potentially signalling a communication breakdown with the vocal and visual cues (rising intonation, changes in speech rate, hesitation, facial expressions, gestures) proved extremely helpful in deciding on the most plausible interpretation of the sequences in question. Moreover, our aim was always to consider the CF/CB episodes within their larger contexts and therefore to benefit from the wisdom of hindsight. That meant looking not only at the turns immediately preceding and following the episode under scrutiny, but also taking into account the rest of the conversation. As our comments on examples (9) and (16) demonstrate, the participants' subsequent utterances may provide precious insights into their intentions and thus lend support – or not – to our hypotheses concerning the nature of the actions performed several seconds or minutes earlier. Needless to say, being able to watch the exchanges numerous times offers the researchers various opportunities for refining their hypotheses – another considerable advantage over the real-time processing that the participants themselves needed to execute. Making use of the rather vague labels mixed and combined when coding complex CF or CB instances represents further efforts on our part to minimise the effect of subjective judgement as to what the underlying linguistic triggers were. Such categories tally with the reality of L2 speech production and perception, where the intermingling of issues from various linguistic levels (phonetics, syntax, semantics) is the norm rather than the exception.

Lastly, there is an issue which is more or less implicit in the account of our CB data coding and which represents a challenge and a solution at the same time: the fact that we have only taken into consideration those communication breakdowns that are somehow overtly marked. As a result, a potentially large amount of covert communicative turbulence may have been left unaccounted for. Signalling non-understanding – just like giving CF – may be regarded as a face-threatening act. This means that certain participants may have refrained from sending distress signals as a deliberate strategy to prioritise fluid and friendly communication, in the hope that the meaning intended by their partners would get clarified later in the conversation. In a bid to keep our coding process as objective as possible, we chose not to speculate about – and, consequently, not

8 *Potential pitfalls of interpreting data from E-F tandem conversations*

to quantify – such likely avoidance phenomena.¹⁰ This approach, however, undoubtedly affects the interpretation of our findings, in that our quantitative data almost certainly suggest that the participants misunderstood each other less often than they actually did. Yet again, though, it could be argued that this apparent shortcoming puts our results more in line with real life speech processing than might otherwise be the case, as – according to [Keysar \(2007\)](#) – speakers routinely believe that what they say is accurately understood by the addressee more often than it really is.

6.3 Conclusion

The SITAF tandem corpus captures conversational exchanges between various types of speakers in all their inherent complexity, multimodality and ambiguity. The fact that, by definition, tandem partners do not share an L1 makes matters even more complex and our data even more challenging to interpret, especially at the level of negotiation of form, than would presumably be the case with NS-NS dyads. Throughout the paper we have shown how we attempted to deal with the various aspects of data ambiguity, and how our decisions impact our conclusions.

The analysis of our corpus data could certainly be refined in the future, mainly by going further beyond the verbal and literal information contained in the participants' output. Research paths that could be explored in a bid to enrich our findings include issues like stance taking, power dynamics within individual tandem pairs, a variety of face-saving strategies employed, notions of politeness and appropriateness, affective and empathetic reactions, as well as task effects. In the event of compiling a new, similar corpus of tandem interactions in the future, data ambiguity could to some extent be reduced through employing a stimulated recall protocol (as done by [Mackey et al. 2000](#), for example). This would enable the researchers to watch the recorded interactions with both participants, with a view to discussing their perceptions of the LREs they have just engaged in. Despite such measures, an element of ambiguity is still bound to remain when it comes to perceiving and interpreting real people's actions, intentions and emotions. There will therefore always be more to a database of human interactions than will meet the researcher's cautious eye.

¹⁰On numerous occasions it was tempting to engage in such speculations (for instance, when the participant was speaking very fast or indistinctly or their L2 production was extremely dysfluent).

Sylvia Scheuer & Céline Horgues

Acknowledgements

The SITAF project was financed thanks to a research grant obtained from the Conseil Scientifique de l'Université Sorbonne Nouvelle (Projet Jeunes Chercheurs, 2012-2014). Part of the orthographic transcription and corpus finalisation was financially sponsored by Labex EFL program (ANR-10-LABX-0083) and Ircor/Ortolang. For their participation and support, we are very grateful to all the SITAF tandem participants, the SITAF team members, our research team (Sesylia, Prismes EA 4398) and the university's engineers.

References

- Adams, Rebecca. 2007. Do second language learners benefit from interacting with each other? In Alison Mackey (ed.), *Conversational interaction in second language acquisition: A collection of empirical studies*, 29–51. Oxford: Oxford University Press.
- Ballinger, Susan. 2012. *Towards a cross-linguistic pedagogy: Bilingual and reciprocal learning strategies in French immersion*. Montreal. (Doctoral dissertation).
- Ballinger, Susan. 2015. Linking content, linking students: A cross-linguistic pedagogical intervention. In Jasone Cenoz & Durk Gorter (eds.), *Multilingual education. between language learning and translanguaging*, 35–60. Cambridge: Cambridge University Press.
- Bamgbose, Ayo. 1998. Torn between the norms: Innovations in world Englishes. *World Englishes* 17(1). 1–14.
- Basterrechea, María & María del Pilar García Mayo. 2013. Language-related episodes during collaborative tasks: A comparison of CLIL and EFL learners. In Kim McDonough & Alison Mackey (eds.), *Second language interaction in diverse educational contexts*, 25–43. Amsterdam: John Benjamins.
- Basterrechea, María & Michael J. Leiser. 2019. Language-related episodes and learner proficiency during collaborative dialogue in CLIL. *Language Awareness* 28(2). 97–113.
- Brammerts, Helmut. 1996. Tandem language learning via the internet and the international email tandem network. In David Little & Helmut Brammerts (eds.), *A guide to language learning in tandem via the Internet (CLCS Occasional Paper no. 46)*, 9–22. Dublin: Trinity College.
- Brammerts, Helmut & Mike Calvert. 2003. Learning by communicating in tandem. In Tim Lewis & Lesley Walker (eds.), *Autonomous language learning in tandem*, 45–59. Sheffield: Academy Electronic Press.

8 Potential pitfalls of interpreting data from E-F tandem conversations

- Carpenter, Helen, K. Seon Jeon, David MacGregor & Alison Mackey. 2006. Learners' interpretations of recasts. *Studies in Second Language Acquisition* 28(2). 209–236.
- Chambers, Jack K. & Peter Trudgill. 1998. *Dialectology*. Cambridge: Cambridge University Press.
- Dascal, Marcelo. 1999. Introduction: Some questions about misunderstanding. *Journal of Pragmatics* 31. 753–762.
- Debras, Camille, Céline Horgues & Sylwia Scheuer. 2015. The multimodality of corrective feedback in tandem interactions. *Procedia – Social and Behavioural Sciences* 212. 16–22. DOI: [10.1016/j.sbspro.2015.11.292](https://doi.org/10.1016/j.sbspro.2015.11.292).
- Doughty, Catherine. 2001. Cognitive underpinnings of focus on form. In Peter Robinson (ed.), *Cognition and second language instruction*, 206–257. Cambridge: Cambridge University Press.
- El Tatawy, Mounira. 2002. Corrective feedback in second language acquisition. *Working Papers in TESOL & Applied Linguistics* 2. 1–19.
- Ellis, Rod. 2017. Oral corrective feedback in L2 classrooms: What we know so far. In Hossein Nassaji & Eva Kartchava (eds.), *Corrective feedback in second language teaching and learning*, 3–18. New York: Routledge.
- Ellis, Rod, Younghee Sheen, Mihoko Murakami & Hide Takashima. 2008. The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System* 36(3). 353–371.
- Foster, Pauline. 1993. Discoursal outcomes of small group work in an EFL classroom: A look at the interaction of non-native speakers. *Thames Valley University Working Papers in English Language Teaching* 2. 1–32.
- Foster, Pauline. 1998. A classroom perspective on the negotiation of meaning. *Applied Linguistics* 19(1). 1–23.
- Gass, Susan M. 2003. Input and interaction. In Catherine J. Doughty (ed.), *The handbook of second language acquisition*, 224–255. Oxford: Blackwell.
- Hendrickson, James M. 1978. Error correction in foreign language teaching: Recent theory, research, and practice. *The Modern Language Journal* 62(8). 387–398.
- Horgues, Céline & Sylwia Scheuer. 2014. “i understood you, but there was this pronunciation thing..”: L2 pronunciation feedback in English/French tandem interactions. *Research in Language* 12(2). 145–161.
- Horgues, Céline & Sylwia Scheuer. 2018a. Paper presented at EuroSLA, Münster, 4–8 September 2018. Coming to an understanding: Miscommunication management in the course of English/French tandem interactions.

Sylwia Scheuer & Céline Horgues

- Horgues, Céline & Sylwia Scheuer. 2018b. “what are you not understanding?” Managing communication breakdowns in native/non-native English/French tandem interactions. In Elena Babatsouli (ed.), *Crosslinguistic research in monolingual and bilingual speech*, 129–146. Chania: Institute of Monolingual & Bilingual Speech.
- Horgues, Céline & Claire Tardieu. 2015. *Positionnement correctif dans les interactions orales en tandem anglais-français*. Paris: Association des chercheurs et enseignants didacticiens des langues étrangères ACEDLE.
- Horgues, Céline & Sylwia Scheuer. 2015. Why some things are better done in tandem? In Jose A. Mompeán & Jonás Fouz-González (eds.), *Investigating English pronunciation: Current trends and directions*, 47–82. Basingstoke: Palgrave Macmillan.
- Jenkins, Jennifer. 2000. *The phonology of English as an international language*. Oxford: Oxford University Press.
- Kartchava, Eva. 2019. Noticing oral corrective feedback in the second language classroom: Background & evidence: Lexington Books.
- Kartchava, Eva & Ahlem Ammar. 2014. The noticeability and effectiveness of corrective feedback in relation to target type. *Language Teaching Research* 18(4). 428–452.
- Keysar, Boaz. 2007. Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics* 4(1). 71–84.
- Li, Shaofeng. 2010. The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning* 60(2). 309–365.
- Loewen, Shawn. 2018. Focus on form versus focus on forms. In John I. Lontas (ed.), *The TESOL encyclopedia of English language teaching*, 2750–2755. Hoboken, NJ: John Wiley & Sons, Inc. DOI: [10.1002/9781118784235.eelt0062](https://doi.org/10.1002/9781118784235.eelt0062).
- Long, Michael H. 1991. Focus on form: A design feature in language teaching methodology. In Kees De Bot, Ralph Ginsberg & Claire Kramsch (eds.), *Foreign language research in cross-cultural perspective*, 39–52. Amsterdam: John Benjamins.
- Lyster, Roy & Leila Ranta. 1997. Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition* 19(1). 37–66.
- Lyster, Roy & Kazuya Saito. 2010. Oral feedback in classroom SLA. 32(2). *Studies in Second Language Acquisition*, 265–302.
- Lyster, Roy, Kazuya Saito & Masatoshi Sato. 2013. Oral corrective feedback in second language classrooms. *Language Teaching* 46(1). 1–40.
- Mackey, Alison. 2006. Feedback, noticing and instructed second language learning. *Applied Linguistics* 27(3). 405–430.

8 Potential pitfalls of interpreting data from E-F tandem conversations

- Mackey, Alison, Susan Gass & Kim McDonough. 2000. How do learners perceive interactional feedback? *Studies in Second Language Acquisition* 22. 471–497.
- Mackey, Alison, Rhonda Oliver & Jennifer Leeman. 2003. Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning* 53(1). 35–66.
- Mauranen, Anna. 2006. Signaling and preventing misunderstanding in English as a lingua franca communication. *International Journal of the Sociology of Language* 177. 123–150.
- McKay, Sandra L. 2002. *Teaching English as an international language*. Oxford: Oxford University Press.
- Nassaji, Hossein & Eva Kartchava (eds.). 2017. *Corrective feedback in second language teaching and learning: Research, theory, applications, implications*. New York/London: Routledge.
- In press. In Hossein Nassaji & Eva Kartchava (eds.), *The Cambridge handbook of corrective feedback in language learning and teaching*. Cambridge: Cambridge University Press.
- O'Dowd, Robert & Markus Ritter. 2006. Understanding and working with 'failed communication' in telecollaborative exchanges. *CALICO Journal* 23(3). 623–642.
- O'Rourke, Breffni. 2005. Form-focused interaction in online tandem learning. *CALICO Journal* 22(3). 433–466.
- Pallotti, Gabriele. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics* 30(4). 590–601.
- Saito, Kazuya. In press. Corrective feedback and the development of L2 pronunciation. In Hossein Nassaji & Eva Kartchava (eds.), *The Cambridge handbook of corrective feedback in language learning and teaching*. Cambridge: Cambridge University Press.
- Saito, Kazuya & Roy Lyster. 2012. Investigating the pedagogical potential of recasts for L2 vowel acquisition. *TESOL Quarterly* 46(2). 387–398.
- Sato, Masatoshi. 2017. Oral peer corrective feedback. In Hossein Nassaji & Eva Kartchava (eds.), *Corrective feedback in second language teaching and learning*, 19–34. New York/London: Routledge.
- Sato, Masatoshi. In press. Metacognitive instruction for collaborative interaction: The process and product of self-regulated learning in the Chilean EFL context. In Craig Lambert & Rhonda Oliver (eds.), *Using tasks in diverse contexts*. Clevedon: Multilingual Matters.
- Sato, Masatoshi & Shawn Loewen. 2018. Metacognitive instruction enhances the effectiveness of corrective feedback: Variable effects of feedback types and linguistic targets. *Language Learning* 68(2). 507–545.

- Sato, Masatoshi & Shawn Loewen. 2019. Methodological strengths, challenges, and joys of classroom-based quasi-experimental research: Metacognitive instruction and corrective feedback. In Robert M. DeKeyser & Goretti Prieto Botana (eds.), *Doing SLA research with implications for the classroom: reconciling methodological demands and pedagogical applicability*, 31–54. Amsterdam: John Benjamins.
- Sato, Masatoshi & Roy Lyster. 2012. Peer interaction and corrective feedback for accuracy and fluency development: Monitoring, practice, and proceduralization. *Studies in Second Language Acquisition* 34(4). 591–626.
- Scheuer, Sylwia & Céline Horgues. 2020. Corrective feedback in English/French spoken tandem interactions. In Claire Tardieu & Céline Horgues (eds.), *Redefining tandem language and culture learning in higher education*, 147–160. New York, NY: Routledge.
- Sheen, Younghee. 2006. Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research* 10(4). 361–392.
- Sheen, Younghee & Rod Ellis. 2011. Corrective feedback in language teaching. In Eli Hinkel (ed.), *The handbook of research in second language teaching and learning*, 593–610. New York/London: Routledge.
- Sippel, Lieselotte & Carrie N. Jackson. 2015. Teacher vs. peer oral corrective feedback in the german language classroom. *Foreign Language Annals* 48(4). 688–705. DOI: [10.1111/flan.12164](https://doi.org/10.1111/flan.12164).
- Smith, Larry E. & Cecil L. Nelson. 1985. International intelligibility of English: Directions and resources. *World Englishes* 4(3). 333–342.
- Storch, Neomy. 1998. A classroom-based study: Insights from a collaborative text reconstruction task. *ELT Journal* 52(4). 119–158.
- Storch, Neomy. 2002. Patterns of interaction in ESL pair work. *Language Learning* 52(1). 291–300.
- Storch, Neomy & Ali Aldosari. 2012. Pairing learners in pair work activity. *Language Teaching Research* 17(1). 31–48.
- Swain, Merrill & Sharon Lapkin. 1995. Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics* 16(3). 371–391.
- Swain, Merrill & Sharon Lapkin. 1998. Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal* 82(3). 320–337.
- Trudgill, Peter. 2005. Native-speaker segmental phonological models and the English Lingua Franca Core. In Katarzyna Dziubalska-Kołaczyk & Joanna Przedlacka (eds.), *English pronunciation models: A changing scene. A democratic accent for the global village?*, 77–98. Frankfurt am Main: Peter Lang.

*8 Potential pitfalls of interpreting data from English-French tandem
conversations*

- Varonis, Evangeline M. & Susan Gass. 1985. Miscommunication in native/nonnative conversation. *Language in Society* 14(3). 327–343.
- Ware, Paige D. & Robert O'Dowd. 2008. Peer feedback on language form in telecollaboration. *Language Learning & Technology* 12(1). 43–63.

