

## ABSTRACT

Title of Dissertation:                   CROSS-LINGUISTIC DIFFERENCES IN THE  
LEARNING OF INFLECTIONAL  
MORPHOLOGY: EFFECTS OF TARGET  
LANGUAGE PARADIGM COMPLEXITY

Ekaterina Solovyeva, Doctor of Philosophy,  
2020

Dissertation directed by:           Professor Robert M. DeKeyser, Second  
Language Acquisition program

Inflectional morphology poses significant difficulty to learners of foreign languages.

Multiple approaches have attempted to explain it through one of two lenses. First, inflection has been viewed as one manifestation of syntactic knowledge; its learning has been related to the learning of syntactic structures. Second, the perceptual and semantic properties of the morphemes themselves have been invoked as a cause of difficulty.

These groups of accounts presuppose different amounts of abstract knowledge and quite different learning mechanisms. On syntactic accounts, learners possess elaborate architectures of syntactic projections that they use to analyze linguistic input. They do not simply learn morphemes as discrete units in a list—instead, they learn the configurations of feature settings that these morphemes express. On general-cognitive accounts, learners do learn morphemes as units—each with non-zero difficulty and more or less independent of the others. The “more” there is to learn, the worse off the learner.

This dissertation paves the way towards integrating the two types of accounts by testing them on cross-linguistic data. This study compares learning rates for languages whose

inflectional systems vary in complexity (as reflected in the number of distinct inflectional endings)—German (lowest), Italian (high), and Czech (high, coupled with morpholexical variation). Written learner productions were examined for the accuracy of verbal inflection on dimensions ranging from morphosyntactic (uninflected forms, non-finite forms, use of finite instead of non-finite forms) to morpholexical (errors in root processes, application of wrong verb class templates, or wrong phonemic composition of the root or ending). Error frequencies were modeled using Poisson regression.

Complexity affected accuracy differently in different domains of inflection production. Inflectional paradigm complexity was facilitative for learning to supply inflection, and learners of Italian and Czech were not disadvantaged compared to learners of German, despite their paradigms having more distinct elements. However, the complexity of verb class systems and the opacity of morphophonological alternations did result in disadvantages. Learners of Czech misapplied inflectional patterns associated with verb classes more than learners of German; they also failed to recall the correct segments associated with inflections, which resulted in more frequent use of inexistent forms.

CROSS-LINGUISTIC DIFFERENCES IN THE LEARNING OF INFLECTIONAL  
MORPHOLOGY: EFFECTS OF TARGET LANGUAGE PARADIGM  
COMPLEXITY

by

Ekaterina Solovyeva

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
2020

Advisory Committee:

Professor Robert M. DeKeyser, Chair

Professor Steven J. Ross

Dr. Polly O'Rourke

Dr. Amir Zeldes

Dean's Representative: Professor Ralph Bauer

© Copyright by  
Ekaterina Solovyeva  
2020

## Acknowledgements

Clichés are clichés for a reason. It does take a village, and mine extends in space and time. I have a lot of thanks to give:

I thank my advisor, Robert DeKeyser, for the gifts of close reading, intellectual engagement, and generously letting me explore and speculate.

Steven Ross, Polly O'Rourke, Amir Zeldes, and Ralph Bauer for insightful questions and stimulating thoughts about the meaning of the data and its limitations.

The support provided by the dissertation grant from the National Foreign Modern Language Teachers Association jointly with the National Council of Less Commonly Taught Languages.

Countless students of German, Italian, and Czech who—by choice or out of necessity—braved the journey of language learning and participated in the data collection that served as the basis of the corpus used in my study.

At my first academic home in the US, the University of Northern Iowa: Ardith Meier—for modeling a life of curiosity, service, and high standards (as well as assuring me that I am not going to “get stupider”). Siegrun Wildner, John Balong, Reinhold Bubser, Otto Maclin—for giving me the tools and space to engage in scholarship and giving me my start on this path.

My cohort at the University of Maryland, College Park: Ilina Kachinske, Stephen O'Connell, Susan Benson. The camaraderie of our first year is among my fondest memories.

My tribe at the University Career Center: Rachel Wobrak, Becky Weir, Erin Rooney-Eckel, Pamela Allen, Erin Brault. I had a home on this campus thanks to all of you.

The Graduate School Writing Center, Linda Macrì—for providing the space, support, and snacks for the *practice* of scholarship throughout all stages of this and other projects.

Zach Hebert—for coffee, walks, eye rolls, your humor and sanity.

Rene Jones—I'd say that our friendship cannot be measured, but it can: in minutes, pages written and edited, deadlines blown, late nights, and trees.

My oldest friends—Maya Kosova, Olga Vasileva, Olga Avsyukevich. Your friendship has taken many forms since we met, but what has not changed is your warmth, sincerity, and you just “getting” me.

My parents, Lyubov Solovyova and Sergei Solovev—for buying me all those books throughout the years, unwavering support of my ambitions, and making peace with my nomadic life.

## Table of Contents

Acknowledgments .....	ii
List of Tables .....	vi
List of Figures .....	viii
Introduction: What Makes Inflectional Morphology Hard? .....	1
Chapter 1: Regularities in L1 and L2 Learning of Inflectional Morphology .....	7
1.1 Morphological development in L1 .....	7
1.2 Morphological development in L2: English .....	10
1.3 Morphological development in L2: other target languages .....	17
Chapter 2: Accounts of Morphological Development .....	26
2.1 Syntactic Competence .....	26
2.2 General-cognitive Approaches .....	35
Chapter 3: Research Questions and Motivations .....	42
3.1 Benefits of Paradigm Complexity .....	42
3.2 Potential Trade-offs between Learning and Processing .....	46
3.3 The Current Research .....	48
3.4 Target languages and their inflectional systems .....	50
Chapter 4: Methods—Corpus Study of Written Learner Productions .....	73
4.1 Data Source and Learner Backgrounds .....	73
4.2 Procedure .....	81
4.3 Error Categories and Their Significance .....	82
4.4 Cleaning and Coding of Data .....	86
Chapter 5: Results—Cross-linguistic Differences in Inflection Error Frequency .....	96
5.1 Regression Model Specification and Model Selection .....	99
5.2 Regression Model Results .....	102
5.3 Cross-Validation .....	124
Chapter 6: Results—Production of Verbal Inflection in German: Phonological Environments .....	128

6.1 Methods.....	133
6.2 Results.....	139
6.3 Conclusions.....	151
Chapter 7: Discussion and Conclusions.....	154
7.1 Key research aims and findings .....	154
7.2 Theoretical implications and takeaways.....	157
7.3 Limitations to consider in future research .....	162
7.4 Contributions.....	169
References.....	174



## List of Tables

Table 1 Syllable structure and permissible coda clusters in target and first languages .....	59
Table 2 Corpus frequencies of inflected forms in Czech (spoken, written) .....	61
Table 3 Written frequencies of inflected forms in a web corpus of Italian .....	64
Table 4 Frequencies of German inflected forms in a written corpus.....	65
Table 5 Comparison of rank orders of inflected forms in written German, Italian, and Czech.....	71
Table 6 Summary of key differences between the morphological systems of target languages.....	71
Table 7 Merlin corpus statistics: Number of texts re-rated at each CEFR level.....	74
Table 8 Error types adopted in the coding scheme, with examples from each TL.....	85
Table 9 Examples of data excluded during data cleaning.....	90
Table 10 Structure of the Data .....	97
Table 11 Variables used in the analysis .....	99
Table 12 Summary of Poisson Model Dispersion and Model Fit Values.....	101
Table 13 Regression model results predicting error rates in German, Italian, and Czech.....	103
Table 14 Significance of model contrasts integrated by variable and interaction .....	105
Table 15 Contributions of interaction terms to model fit (assessed by single term deletions).....	105
Table 16 Summary of pairwise comparisons of error rates between target languages by type.....	110
Table 17 Rank orders of error types by target language and proficiency level .....	117
Table 18 Rank orders of error types by target language, averaged across all proficiency levels .....	119
Table 19 Prediction accuracy of regression models when tested on unseen test data .....	126
Table 20 First language backgrounds of learners in the sample.....	134
Table 21 Classification schemes for predicate type.....	137
Table 22 Effects of predicate type (four coding schemes) on inflection accuracy .....	141
Table 23 Effects of syllabicity on accuracy of production .....	143

Table 24 Effects of previous segment class on accuracy of inflectional ending: Obstruents versus sonorants .....	144
Table 25 Effect of previous segment on inflection accuracy: Manner of articulation.....	145
Table 26 Effect of following segment on inflection accuracy .....	146
Table 27 Joint effects of phonological environment on inflection accuracy .....	147
Table 28 Combined effects of syllabicity of ending and phonological environment on inflection accuracy .....	149

## List of Figures

Figure 1. Learner L1 backgrounds by Target L2, aggregated across all proficiency levels. ....	77
Figure 2. L1-TL contingency table Chi-square test residuals (left) and their % contribution to total statistic (right). ....	78
Figure 3. Model residuals plotted against fitted values for the model predicting error counts from: target language, CEFR, error type, and interactions between—TL*error type; TL*CEFR; CEFR*error type.....	101
Figure 4. Plots of aggregated model effects. Top panel: target language by error type interaction; middle panel: target language by proficiency interaction; bottom panel: proficiency by error type interaction.....	106
Figure 5. Model-predicted rates by type for German, Italian, and Czech across proficiency levels A2 through B1+.....	107
Figure 6. Relativized (per number of texts) observed frequencies of error types in German, Italian, and Czech across proficiency levels A2-B1+.....	108
Figure 7. Summary of pairwise comparisons among error type rates within each TL, averaged across all proficiency levels.....	117
Figure 8. Cross-over pattern in morphosyntactic and morpholexical errors depending on target-language complexity.....	120
Figure 9. Error rates by type and target language across CEFR proficiency levels .....	121
Figure 10. Interaction between class of following phonological segment (x axis) and previous phonological segment (y axis) in affecting inflection accuracy.....	149

## Introduction: What Makes Inflectional Morphology Hard?

Morphological difficulties have been perhaps the most salient hallmark of adult language learning since the inception of its study. Morphology and morphosyntax continue to dominate our notions of L2 proficiency, ultimate attainment, and fossilization (e.g., DeKeyser, 2000; Johnson & Newport, 1989; Lardiere, 1998), even as our thinking on communicative competence evolves and is enriched by considerations of sociocultural, pragmalinguistic factors, or phenomena at the interface between pragmatics and syntax, among others.<sup>1</sup>

Some of the claims downplaying morphosyntactic difficulties are based on the first emergence of grammatical forms used contrastively, rather than a preponderance of grammatical forms used correctly, or even on the learners' ability to successfully *comprehend* the feature of interest (as demonstrated through sentence interpretation tasks, for example). Thus, a feature considered "acquired" based on this type of analysis may, in fact, fail to be realized in the majority of learner productions. Such assertions also emphasize the primacy of competence over performance, arguably distorted by "noise", such as processing limitations, retrieval failure, and memory bottlenecks. And yet, learner errors are not as random as would be expected on account of processing resource breakdowns.

It is this non-randomness in learners' errors that has been invoked (White, 2003, p. 196) to argue against a global breakdown in abstract syntactic competence (e.g.,

---

<sup>1</sup> Recent examinations of near-native learners' difficulties, by contrast, have emphasized phenomena at the interface of syntax with pragmatics and semantics as main areas of difficulty, not morphosyntax and syntax per se (e.g., Sanchez, Camacho, & Ulloa, 2010; Sorace, 2011; Sorace & Filiaci, 2006). However, considering the small proportion of learners who reach near-nativeness among a vast majority who do not, it seems fair to say that morphology and syntax are far from trivial for the *average* learner.

Clahsen, 1988; Meisel, 1997) as the root cause of persistent morphological errors. Attempts to explain morphological difficulties by positing gaps in lexical learning and memory retrieval are also insufficient upon closer inspection. They seem to merely replace the phenomenon to be explained with a new one, leaving us with a different formulation of the same question: “Why are some inflected forms more easily accessible (readily retrievable) than others?”.

The question of how accurate use of inflected forms develops is interesting in its own right, whether one accepts it as reflecting syntactic competence or calls it by any other name. Regardless of its place as part of syntactic competence or outside of it, target-like use of inflected forms requires extensive learning. Even though such learning has been outsourced to the lexicon in fairly recent syntactic thinking in SLA (Herschensohn, 2001; Lardiere, 1998), the need for it is undisputed even in the strongest Universal Grammar literatures (Lidz & Gagliardi, 2015; Yang, 2002).

The details of this learning have been filled in, to a certain extent, by general-cognitive proposals (DeKeyser, 2005; Goldschneider & DeKeyser, 2001), accounts invoking the limitations of the production processor (Pienemann, 2015), and psycholinguistic approaches that posit developmental shifts between the reliance on whole-form storage of inflected forms and compositional assembly (from former to latter—Clahsen & Felser, 2006; from latter to former—Gor & Jackson, 2013; Portin, Lehtonen, & Laine, 2007).<sup>2</sup> However, these theoretical pieces do not add up to a

---

<sup>2</sup> The two directions are not mutually exclusive in the sense that within one learner, representations of both kinds are likely to coexist and develop non-linearly. The two positions taken by psycholinguistic researchers and cited here claim that the overall progression tends to be in one or the other direction.

complete puzzle, not least because they only hint at how their explanations might fit with the others when some pattern in the data cannot be explained from within a theory itself.

The task at hand, therefore, is to study the progression of this learning, moving from the consideration of difficulty intrinsic to a particular morpheme towards accounting for the acquisition of inflectional paradigms—that is, systems of contrasts among morphemes. Accounting for the totality of forms to be learned is a way to respect—and model—current linguistic descriptions of the adult native speaker endstate, which treat it as a system of structural relations between elements and not lists of independent elements, following a long-standing structuralist tradition (e.g., Saussure, 1966).

Even though evidence on second language (L2) morphological development has by now accumulated for a number of target languages (TLs) other than English, an explicit comparison of rates of morphological learning has not been pursued. Research programs that include multiple TLs are often carried out to validate an existing—and one-TL based—account of difficulty (e.g., Pienemann, 2003) rather than to seek potentially disconfirming evidence or to accommodate cross-linguistic data in a principled way. In this sense, the spirit of multi-language research efforts, such as those based on the Processability Theory (PT) or the Shallow Structure Hypothesis (SSH), is anything but cross-linguistic.

One of the dimensions on which TLs differ in ways potentially consequential for theory building is paradigm complexity. By presenting radically different learning problems, languages ranging in paradigm richness can serve as a testing ground for competing theories of linguistic complexity or even competing conceptualizations of the

learning mechanism itself. Considering entire linguistic systems, rather than isolated grammatical phenomena, has the advantage of reflecting more closely the reality of learning: the language presents itself to the learner at once, even in conditions of the strictest instructional control over input. Even when such control is present, instruction of the focus-on-forms (Long, 1991) variety tends to respect paradigms: it is hard to imagine a pedagogical approach, for instance, that would selectively rely on drills for just one feature combination (e.g., second person singular).

A cross-linguistic study, therefore, holds promise of theoretical significance, beyond reflecting more fully the diversity of real-life language learning contexts. As the review of accounts of grammatical complexity will show (Chapter 2), conceptualizations of L2 morphological difficulty make assumptions about the very nature of grammatical learning, which may have been inherited from descriptions of the learning of English. Deliberately extending existing theoretical accounts to other TLs can subject theoretical accounts to additional scrutiny that may help with the pursuit of a transition account (Gregg, 1996) of SLA. The pressure to account for TL differences revealed by such comparisons can refine the accounts of morphological difficulty to a level that is sufficient for their ultimate integration into a coherent account of SLA.

The present proposal focuses on verbal morphology, owing to verbs' special status as the hub of sentential meaning—dictating argument structure, thematic roles, and case assignment. Verbal inflectional morphology is common denominator for a study with a cross-linguistic focus, since it deals with units of close to universal semantic meaningfulness, such as present and past tense, in contrast to the distinctions that are layered with semantic complexity, such as case. Verbal morphology has been the most

widely studied due to its role in syntactic processes: verbs are typically predicating (Gentner, 1982) and not referring; they are overall more involved in syntactic processes than nouns (Dressler, Stephany, Aksu-Koc, & Gillis, 2007, p. 68).<sup>3</sup> Perhaps owing to this, learning theories have been proposed with far greater enthusiasm for the acquisition of verbal, than nominal, inflectional morphology (e.g., Clahsen & Felser, 2006; Ullman, 2004) and tasked, in addition, with representing lexical and combinatorial processes (e.g., irregular and regular verbal forms, respectively). While I will be invoking data from the psycholinguistic literature on single- versus dual-mechanism processing of inflected words, I will only do so as far as it adds to insights to comparisons of different TLs and relates to learnability—ignoring the debates internal to this literature.

Chapter 1 summarizes the empirical facts on morphological development, from early morpheme order studies to recent corpus analyses. The review will include data from diverse TLs as much as possible, both in L1 and L2 development. As will become clear throughout the review, scholars disagree on what the facts to be explained are. This is not surprising in the absence of predictions guided by a learning theory. In chapter two, I will present an overview of approaches to morphological difficulty—focusing on the accounts that emphasize syntax-morphology interdependencies, followed by those based on general-cognitive mechanisms and processing principles. In particular, I will focus on spelling out the assumptions made by both accounts about the nature of the learning mechanism transitioning the learner from one stage to the next. In chapter three, I will argue that examining rates of growth for different target languages can push theories to

---

<sup>3</sup> Even though nouns are also influenced by contextual factors—for example, their case—their role in syntactic processes such as agreement is one of providing the “inputs” for the verb to agree with, where it is the verb that does the agreeing in response to the noun.



be more explicit in the learning mechanisms they posit. This may lead to more accurate descriptions of the data and to novel insights into the nature of learning in adults, potentially testing whether it proceeds in piecemeal fashion or whether elements in a complex system are acquired in a way that reflects their similarities at an abstract level. I will conclude Chapter 3 by presenting the research questions of the study and by providing descriptions of the relevant aspects of target-language grammars (German, Italian, and Czech). In Chapter 4, I will describe the research methods and the data sources, as well as the error taxonomy and its application to the data during data coding. Chapters 5 and 6 report the results: Chapter 5 concentrates on the results of comparisons among the target languages with respect to the proportions of different error types. Chapter 6 focuses on the production of inflection in L2 German, examining the data through the lens of interlanguage phonological processes. Chapter 7 concludes this dissertation by offering key takeaways and theoretical implications while noting the study's limitations.

## Chapter 1: Regularities in L1 and L2 Learning of Inflectional Morphology

There are well-documented sequences in children's L1 and adult L2 morphological development. Although the exact ordering of morphological features in L2 development have differed from study to study, there is general agreement about broad patterns in the data (Dulay & Burt, 1973; R. Ellis, 1994, 2015; Mitchell & Myles, 2004), sometimes referred to as “Long's Law” (e.g., Ellis, 2015). The sources of discrepancies among studies in the learning orders they propose are multiple and include differences in learners' L1s, as well as the differences in experimental task demands, which, in turn, stemming from different theoretical perspectives on what constitutes “learning” and “knowledge”. For example, approaches that endeavor to characterize the nature of L2 syntactic competence generate their evidence from sentence interpretation and grammaticality judgments, whereas a more applied, testing, or skills-based perspective would include accurate production or aspects of learner performance (e.g., speed).

This chapter will first characterize the regularities in child L1 acquisition, particularly highlighting any cross-linguistic and typological differences. Then it will review the work on the so-called “morpheme orders” identified in the learning of English as a second language. Finally, it will summarize findings from learners of TLs other than English, which have employed both production data and psycholinguistic measures.

### 1.1 Morphological development in L1

The existence of broad regularities in L2 acquisition—and the very desire to find them—parallels observations of developmental sequences in L1 acquisition. In L1 acquisition, morphological development is characterized by the presence of root

infinitives (RI)—seemingly non-finite forms that lack overt morphological marking and are produced where a finite form is required. RIs have been attested across typologically diverse languages (Rizzi, 1993/1994; Wexler, 1994). Notably, RIs, which are errors of omission, are more common than errors of commission (supplying wrong inflection). The gradual disappearance of RIs in development has led some researchers to conclude that the syntactic projections supporting finiteness mature over time, even though others have disagreed with RIs' characterization as infinitival in the first place (Phillips, 1995).

Whether truly non-finite or tacitly finite but lacking overt morphological markers, RIs lack surface morphology and vary cross-linguistically both in prevalence and the age at which they disappear from child productions (Phillips, 1995). The length of RI persistence in child speech has been linked to the relative complexity of the language's inflectional paradigm (Legate & Yang, 2007).

Transcending paradigm complexity, robust differences have also been attested along typological lines. Morphological systems are acquired earlier in agglutinative languages than in fusional languages, as early cross-linguistic comparisons of L1 acquisition showed (Slobin, 1985). This finding has been extended through more cross-linguistic comparisons (Laaha & Gillis, 2007), in which languages were not merely construed as representing distinct idealized *types* (e.g., “*agglutinating*”) but, following principles of quantitative typology (Hempel & Oppenheim, 1936, cited in Dressler, 2007, p. 3, 5), as possessing different *levels* of the typological property of interest—“agglutination”, or “inflection” (Dressler et al., 2007). For example, within the “inflecting” group of languages, the graded nature of inflection as a typological property was taken into account. French, German, and Dutch were not merely considered “weakly

inflecting”, and Greek, Croatian, and Russian “strongly inflecting”. Rather, the differences were treated as continuous properties, with French, for instance, being *less* inflecting than German, or Russian *less* inflecting than Greek. Within the same language type (e.g., weakly or strongly inflecting), morphological systems of languages whose inflectional paradigms were richer were acquired faster. Among the weakly inflecting languages, children acquiring German and Dutch outpaced children acquiring French; among the strongly inflecting languages, children acquiring Greek developed inflection at a faster rate than did the Russian and Croatian children (Stephany, Voeikova, Christofidou, Gagarina, Kovacevic, Palmovic, & Hrzica, 2007, p. 46).

Comparisons were also conducted between typological groups: agglutinating, weakly inflecting, and strongly inflecting. Both paradigmatic and syntagmatic richness were considered: paradigmatic richness refers to the number of structural choices made available by a language, whereas syntagmatic richness refers to the average length of morpheme sequences in words. For the development of inflection in the verbal domain, only paradigmatic morphological richness was predictive of rate of development. By contrast, in the nominal domain syntagmatic richness mattered as well (Xanthos, 2007, p. 64)—expressed as the average number of affixes per word. Spearman correlation values between paradigm richness attested in the input and children’s speed of development were quite high: for verbs (paradigmatic richness)—0.76,  $p = 0.003$ ; for nouns (paradigmatic)—0.93,  $p < 0.001$ , syntagmatic—0.77,  $p = 0.02$ . Notably, the roles of transparency, uniformity, and salience could not be consistently identified (Dressler et al., 2007, p. 70; Xanthos, p. 64). This suggests that their role may be one of a tiebreaker or more easily identifiable when languages within the same typological group are examined.

These results put observations about the role of language typology in acquisition from earlier studies on firm quantitative ground. In earlier work, the advantages of agglutination were noted and explained in a general-cognitive light as stemming from the transparency of form-meaning mappings (Peters, 1997, p. 181; Slobin, 1985, p. 1216). The more recent studies conducted by Dressler and colleagues (2007) propose that paradigm richness may also contribute to this developmental advantage, possibly by exerting communicative pressure on the child to pay attention to subtle differences in meaning between forms (p. 9). Both explanations may be valid, considering that communicative pressure alone does not necessarily predict a difference between agglutinating and strongly inflecting languages.

## 1.2 Morphological development in L2: English

The morpheme order studies yielded the observation that, often independent of instruction, learners tend to acquire the grammatical features of TL English in a similar order (Brown, 1973; Dulay & Burt, 1973; Larsen-Freeman, 1975; Pica, 1983). Despite two frequent criticisms levied against the early work on acquisition orders, the findings of morpheme order studies are still considered valid (Larsen-Freeman & Long, 1991).

The first criticism concerns their cross-sectional nature; the second is their reliance on arbitrary criteria to determine learning, such as a 90% accuracy rate. Addressing the first problem, a number of subsequent studies tracked individual learners longitudinally and yielded similar findings (e.g., Dyson, 2009; Lardiere, 1998) both across different L1s in the learning of English and across different L2s, including German and Swedish (Pienemann, 2005). Thus, acquisition orders are not merely an artifact of averaging across multiple learners. However, the potential for discrepancies between the

performance of any one learner picked at random and the acquisition orders captured in aggregate also exists. Insofar as studies oriented at describing the orders do not make commitments to a particular learning procedure, operating on specified inputs to produce a range of expected outputs, expecting their results to adequately capture developmental idiosyncrasies of single learners will be a recipe for disappointment. As such, these discrepancies may not speak to the presence or absence of orders in learning but underscore the need for principled criteria in deciding which ones of them are noise and which would invalidate the notion of orderly development.

Concerning the second criticism, different criteria have been proposed in the literature to make morpheme studies less vulnerable to tracking the noise in data that arises from learner performance. For example, the emergence criterion (Meisel, Clahsen, & Pienemann, 1981) relies on the first instances of reliable, contrastive use of a morpheme, ignoring the lingering optionality that may persist in learner productions for years. However, neither emergence nor the achievement of some level of accuracy indexes any learning phenomena interesting in its own right. Rather, it is a snapshot in time of the underlying movement driven by the operation of learning processes. The utility of both lies in their ability to reflect an underlying growth curve that is driven by the operation of learning mechanisms, which then produce distinct “orders”—or rankings of morphemes by accuracy—at whichever points one samples along this trajectory. Therefore, any discrepancies in orders that arise as a byproduct of one’s choice of the points to be sampled (90% accuracy, first emergence) are not altogether surprising, and neither invalidate nor prove the existence of a universal learning mechanism. Difference

in orders (or accuracy rankings) will then emerge because of differences in growth rates, much like lines that will cross if their slopes are different.

These caveats aside, a consensus seems to have emerged from studies relying on error rates in learner productions that morphological errors in L2 are more varied than those attested in child L1 and include erroneous inflections, not just omissions (Bruhn de Garavito, 2004; Jia & Fuse, 2007; Mezzano, 2003; Morales, 2014; Prevost & White, 2000). On the other hand, some of the findings have been invoked in the syntactic literature (e.g., White, 2003, p. 196) to make *comparative* claims about the prevalence of omission and substitution. In particular, their authors argue that learners predominantly omit inflection, but when they do inflect, they tend to inflect correctly (Grondin & White, 1996; Haznedar & Schwartz, 1997; Ionin & Wexler, 2002; Prevost & White, 2000b; White, 2002).

Some other developmental patterns in morphological development across a number of studies concern the relative difficulty of bound morphology compared to morphological features expressed through free morphemes (Dyson, 2009; Jia & Fuse, 2007; Lardiere, 1998a, b; Vainikka & Young-Scholten, 1996; Zobl & Liceras, 1994). This distinction is appealing because it transcends theoretical boundaries (UG, processing, general-cognitive accounts) by drawing on the salience of free morphemes as freestanding words in an adult's mind (Van Patten, 2004).

Looking at error rates on verbal inflectional morphology, what stands out is the overall low accuracy regardless of length of residence. For instance, Lardiere's (1998) subject Patty (LOR at first recording was 10 years, last recording—18 years), whose first languages were two varieties of Chinese (Hokkien, Mandarin), correctly marked tense

only about 34% of the time (through inflection or tense-expressing auxiliary). This rate stayed steady over the course of the 8.5 years that elapsed between the first and subsequent recordings.

This pattern holds even in studies that included younger learners, who cannot possibly be deemed to have entered a fossilized state. In a study conducted within the processability theory framework, Dyson (2009) analyzed the productions by two adolescent, beginner-level learners of English as an L2 who were native speakers of Chinese. The measurements were spread over the course of nine months, and both participants had received some English instruction in their home country prior to relocating to Australia. Only one of the two learners reached the emergence criterion on third person singular by the sixth—and last—measurement, achieved through the suppliance of three correct forms out of 51 available obligatory contexts. The other learner did not reach the criterion, supplying zero correct forms on four measurements, 2 (out of possible 46) on measurement 4, and 1 (out of 43) on the very last occasion. Tense marking did not fare much better: the first learner showed emergence of irregular past-tense marking at the second measurement but the regular *-ed* rule only at the 4<sup>th</sup> (based on one correct token and two overgeneralizations), continuing at that level at measurement 5 (one token only), and producing no tokens whatsoever at measurement six. The second learner demonstrated the emergence of irregular past at measurement four (with 10 tokens) but produced no tokens of regular *-ed* past over nine months.

Similarly low accuracy was reported in a study comparing early- and late-starters of English as an L2—with ages at arrival between 5 and 16, who were also native speakers of Mandarin (Jia & Fuse, 2007). The study started tracking participants three



months after arrival in the United States and continued over five years. By the 16<sup>th</sup> testing session administered at the end of five years of residence in the U.S., only three out of 10 participants had achieved 80% or higher accuracy on third person singular, four out of 10 on the irregular past tense, and none of the 10 on regular past tense. These data were obtained from spontaneous productions during interviews with the researchers. Over 90% of errors in both domains were errors of omission, not wrong inflection.

Ruling out a purely phonological explanation of these difficulties, it has been shown that learners with L1 backgrounds other than Chinese, including Korean (Johnson & Newport, 1989), Hungarian (DeKeyser, 2000), and Russian (DeKeyser, Alfi-Shabtay, & Ravid, 2010), among others, experience difficulty with English inflectional morphology. These difficulties have been demonstrated on a variety of tasks that are deemed to be less processing-intensive, such as grammaticality judgments, rather than spontaneous production.

Somewhat obscuring the regularities, L1 influences have been shown to be one source of discrepancies in the learning orders (Goldschneider & DeKeyser, 2001; Murakami & Alexopolou, 2015). For example, there are departures from the generalization that free morphemes are learned earlier than bound ones that can be attributed to L1 influences. The relationships between L1 and L2 are far from straightforward and do not boil down to simple transfer. For example, learners of English who are L1 speakers of Chinese master third person singular *-s* before regular past tense, even though neither feature is expressed morphologically in Chinese (Jia & Fuse, 2007; Luk & Shirai, 2009). By contrast, L1 speakers of Korean were among those learners for whom regular past-tense marking (*-ed*) had one of the highest target-like use percentages

(along with L1 speakers of Turkish, Japanese, Russian, German and French), even though Korean lacks this feature too. Speakers of Spanish, surprisingly, had one of the lowest TLU rankings for third person singular *-s* among all L1 groups, even though Spanish marks this feature combination (Murakami & Alexopolou, 2015).

Most importantly, even those learners whose L1s did have morphological features comparable to those of the TL, while showing higher accuracy than speakers of L1s lacking them, were well short of 100% accuracy even at the highest proficiency examination levels (Murakami & Alexopolou, 2015: CPE—Cambridge English: Proficiency, equivalent to the C2 level in the CEFR framework). Therefore, any account of the facilitating role of L1-L2 overlap has to consider that ceiling level. On the one hand, the absence of a morpheme in the L1 seems capable of depressing learners' accuracy. For example, progressive *-ing*, plural *-s*, and possessive *'s* are considered among the easiest based both on earlier morpheme order studies and on theoretical grounds (as belonging to lower-level syntactic projections), yet they exhibited the strongest influences from the presence or absence of comparable features in the L1. Conversely, third person singular *-s*—one of the hardest features of English as an L2—showed the least variation among the L1s examined. This points to an attenuating effect of the L1, not one erasing other dimensions of difficulty altogether.

**Conclusions.** The original formulation of acquisition orders explained their existence as the result of a universal learning mechanism. Arguably, such a mechanism is conceptually separate from the ingredients supplied to it by the L1 and the input. The operation of the mechanism on variable prior knowledge (L1 grammar and the weights of hypotheses about grammar it supplies) and variable input would result in noticeably

different accuracy rankings. Therefore, even the absence of uniform orders in the data need not mean that the operating learning mechanisms are any different.

Even demonstrable L1 influences on accuracy orders are not incompatible with the notion of general acquisition orders. Depending on the learning mechanism one assumes (e.g., MacWhinney, 1989; Yang, 2002; cf. Pinker, 1984), the L1 may merely change the weights of learners' prior hypotheses about the TL grammar, in a Bayesian sense. This would mean that hypotheses with lower prior probabilities would require more evidence from input to successfully influence grammar building, leading to slower acquisition. A learner whose L1 marks tense morphologically would accept that expectation as their prior hypothesis about the TL by default, which would then be strengthened or left without support by the input. By contrast, a learner with no knowledge of past-tense morphological marking from their L1 would have no reason to postulate this feature for the TL *a priori* and would take longer to learn it.

The discrepancies in orders revealed by different studies may never be resolved, nor is such a resolution a prerequisite for successful theory building. Without postulating a learning mechanism that produces close-enough approximations of the empirically observed patterns, not much will be gained from additional data. Large-scale examinations of the kind supplied by corpus studies illustrate the limits of putting description before theory, and scale alone does not always translate into more confidence in the results. On balance, without a general idea of what kind of a learning mechanism lurks behind the differences, it is hard to evaluate which of them may be informative and which ones are noise. Finally, it is also not clear whether the "naturalness" of any orders stems from "nature" or results from range restriction of sorts, reflecting prevalent patterns

in the development of learners of *English*, as opposed to any target language. More recent thinking in this area has shifted towards elucidating the *properties* of morphemes (ranging from perceptual to syntactic) that could account for the observed orders, rather than insisting that it is particular grammatical features that are acquired in a fixed sequence (e.g., DeKeyser, Alfi-Shabtay, Ravid, & Shi, 2017).

### 1.3 Morphological development in L2: other target languages

Work on TLs other than English has been sparse. Rather than originating as a research topic in its own right, the inclusion of other TLs has been a by-product of the multitude of theoretical approaches enlisted to explain the acquisition of grammar in L2, each largely concerned with their own agendas.

The acquisition of Spanish was studied from a Processability Theory (PT) perspective by Bonilla (2015). Applied to Spanish, PT predicts first the emergence of plural marking on lexical heads (*manzanas*, ‘apples’), followed by intraphrasal agreement within the determiner phrase (*las manzanas*, ‘the-fem-pl, cf. ‘apple-s’), followed by interphrasal agreement (agreement marking *-s*). This order was upheld in oral productions of 21 L2 learners of Spanish at beginner, intermediate, and advanced levels. In all learners, syntactic manifestations of the levels, in line with Pienemann (2004), emerged before the corresponding morphology. Indirectly, this also supports the notion of bound morphology being more “difficult”. With respect to agreement marking specifically, only five learners out of 21 reached stage four (interphrasal agreement). Proficiency levels are not detailed for individual learners: instead, the author reports that there were seven learners at each level, corresponding to approximately 180, 750, and 895 hours of instruction (beginner, intermediate, and advanced levels, respectively), and an additional year abroad for the

advanced level. While it is unknown what the proficiency levels were of the learners who achieved each stage of agreement marking, one can estimate that at least two of the advanced students did not reach the acquisition criterion for stage-four features. Since the study did not focus on types of errors, it is impossible to draw further conclusions about the rates of errors of different kinds.

In an analysis of L2 French that did separate erroneously supplied morphological marking from uninflected forms, Herschensohn (2001) provides a breakdown of errors by two intermediate-level learners, one of whom completed a six-month study abroad program. While the number of participants was very small, the study involved sizeable samples of learner discourse with a high number of obligatory contexts for the forms of interest. Over the course of three interviews spanning six months, the suppliance of correct inflectional morphology in obligatory contexts in the present (and past, in parentheses) tenses increased in both learners: 88% (45%) and 74% (10%)—at the first interview, 86% (56%) and 89% (88%)—at the second interview, and 96% (79%) and 98% (97%) at the last, third interview. The second (and more accurate) participant was the one who had studied abroad. Even though the accuracy is already rather high, it is even higher if one ignores errors that involved ellipsis. Without taking into account ellipsis errors, out of the remaining 60 errors, 38% involved the substitution of present-tense forms for past-tense forms, while the remaining two-thirds (62%) were inflectional errors in a narrow sense, broken down into infinitival forms (where finite was required) in less than half of those cases (16 out of 37 inflectional), slightly fewer yet (14/37)

involved applying incorrect morphological marking, and a handful of substitutions of forms marked for the wrong person-number features (7/37).<sup>4</sup>

Studies involving learners of German and French (Prévost & White, 2000) and Spanish (Bruhn de Garavito, 2003; Mezzano, 2003) are in agreement with these generally high accuracy rates. Although they employed different methods, learners of different ages, and proficiency is not equated across these three studies (three years of exposure for French, less than two for German), tentative observations can be drawn. Rates of non-suppliance of inflection were higher for French than German (cited in Morales, 2014, p. 91), even though the corresponding syntactic operation (verb movement) had been acquired by then. For French, non-suppliance reached 23-24%, while for German the figure was 10-16%. Not all finite forms that were supplied were correct, and incorrectly provided forms were more prevalent among learners of German (~12%) than among learners of French (4-5%).

By contrast, lack of finiteness marking in Spanish amounted to only 4% after just 24 hours of exposure in a beginner-level course (Mezzano, 2003), while erroneous agreement was also encountered—12% after 88 hours of classroom instruction. Person-marking forms were used interchangeably, with singular forms tending to be used in place of plural (Mezzano, 2003). Similarly low error rates on agreement were reported by Bruhn de Garavito (2003): only 10%, including both substitutions of inflected forms for one another, as well as infinitives in place of finite forms. A study of learners of Russian showed similarly high levels of accuracy on verbal inflection (~80%—Tkachenko & Chernigovskaya, 2010).

---

<sup>4</sup> Despite the low number of participants, the study had qualitative depth by virtue of its longitudinal design, which resulted in a high number of instances of the features studied.

This pattern of relatively high accuracy contrasts with the difficulties with inflection demonstrated by L2 learners of English (reviewed in the previous section) and echoes observations on aphasia across speakers of different first languages. Similar to learners of English as an L2, English-speaking aphasia patients have been reported to predominantly omit—rather than substitute—inflectional endings (Grodzinsky, 1984; Gorema, 1998). By contrast, speakers of languages in which uninflected forms result in non-words predominantly make substitution errors—as, for example Italian (Miceli, Mazzucchi, Menn, & Goodglass, 1983), Greek (Kehayia, 1990; Kehayia, Jarema, & Kadzielawa, 1990), or Hebrew and Arabic (Mimouni & Jarema, 1997), in which affixes are discontinuous and are inserted into a consonantal root. Therefore, the error patterns reported for English and constituting the “natural order” may be idiosyncratic to a poorly inflected language.<sup>5</sup>

At least on some accounts (see Chapter 2), the errors are linked to developing syntactic knowledge, whereas on others they are caused by retrieval or access failures. Thus, researchers have attempted to isolate the difficulty and only examine compositional processes at the word level, thereby separating it from any processing demands imposed by production or syntactic computations within a sentence. This approach has yielded a separate set of facts to be explained and integrated with the findings obtained using other paradigms.

---

<sup>5</sup> Even though the exact phonological realizations of morphemes differ from language to language, the “natural order” implies some measure of cross-linguistic relevance. Considering that in English a number of morphemes are realized identically (third person *-s*, possessive *'s*, and noun plural marker *-s*), any putative explanations of their relative difficulty would logically have to be based on grammatical meaning, at least in part.

**Decomposition in L2 processing of inflected forms.** Single-word morphological processing research has centered on characterizing the decomposition and whole-form storage of inflected words, and the changes in the relative reliance on them throughout L2 development. As such, learning orders of morphological features have not been of interest. Although cross-linguistic comparisons are as rare in this literature as in the research from other paradigms, the results of psycholinguistic studies that examined different TLs lend themselves more easily to comparisons, owing to the narrower variation in their methods.

On the one hand, many studies of English as a TL have failed to show morphological decomposition of inflected forms, as reflected by the magnitude of priming effects in masked priming tasks (Kirkici & Clahsen, 2013; Neubauer & Clahsen, 2009; Silva & Clahsen, 2008). A handful of studies that used paradigms that allow for *conscious* perception of the prime—such as cross-modal priming—did sometimes show decomposition in L2 English (Basnight-Brown, Chen, Hua, Kostic, & Feldman, 2007; Feldman et al., 2010). By contrast, a growing literature on other TLs includes studies that have demonstrated decomposition in learners of French (masked—Coughlin & Tremblay, 2015), early bilinguals in Finnish (Lehtonen & Laine, 2003), early (Lehtonen et al., 2006) and late bilinguals in Swedish (Portin et al., 2007), and learners of Spanish as early as at the intermediate level (masked—Foote, 2015; Presson, Sagarra, MacWhinney, & Kowalski, 2012) but not advanced (naming—Bowden, Gelfand, Sanz, & Ullman, 2010).

Two opposing predictions have been made with respect to developmental trajectories in the reliance on decomposition versus whole-form storage in the L2. On the one hand, dual models that posit separate neurocognitive mechanisms for the two kinds



of processes specify that learners move from whole-form storage to decomposition as their proficiency increases, even though decomposition may not be achieved fully (Clahsen, Felser, Neubauer, Sato, & Silva, 2010; Ullman, 2004). On the other hand, a different body of psycholinguistic studies (mainly employing morphologically richer languages) have claimed that the direction is the opposite: initially, learners decompose inflected forms but eventually move on to retrieving them whole as increased experience with them allows for the proceduralization of compositional rules (Portin, Lehtonen, & Laine, 2007). Of course, it is also possible that there are multiple shifts between the two throughout development, and these switches are just too granular to be detected by research programs that sample learners at discrete points on the developmental trajectory.

It is unclear how robust the psycholinguistic insights are where cross-linguistic differences are concerned, considering the overall low number of such studies and their inevitable reliance on learners at higher proficiency levels. For a priming study, one has to recruit learners who are familiar with the inflected forms of interest and have accumulated sufficient lexical knowledge to support 30-50 items per condition (including less-frequent items). However, ideas have been put forward regarding how the processing of inflected forms *may* differ in languages differing in typology and morphological complexity. For instance, it has been argued that in languages with richer paradigms the application of inflection is a graded, probabilistic process based on the recognition of arbitrary subclasses (Gor & Jackson, 2013).<sup>6</sup> Thus, producing or comprehending an

---

<sup>6</sup> Ultimately, it is a question best solved empirically: if inflection is indeed applied in a graded fashion in inflectionally rich languages, learners would be expected to make errors in the application of a default (usually “regular”) pattern to novel items—applying subrules and alternations where none are needed—“irregularizing”, so to speak.

inflected form entails the application of several processes involving both inflection and phonological alternations.

In my view, rather than positing a categorical difference between the richer and poorer languages, one may instead view inflection in both as a confluence of processes. The distinction between the core and the periphery is particularly called for in this context. While it may be obscured by alternations, the “core” process involved (in the TLs studied) is that of combining a stem (in whatever form is revealed after the phonological processes) and an affix. Morphophonological alternations can be seen as falling in the middle of a continuum from full suppletion to subregular patterns (with rule-like status) and regarded as the periphery. Viewed through a connectionist lens, the affixation pattern should be a more reliable cue for the learner, since it applies uniformly and categorically, whereas any subrules will, by definition, have lower frequencies and, therefore, lower cue reliability (Presson et al., 2012).

**Conclusions.** Decades of research on L2 grammatical development have produced some foundational insights into the course of learning of inflectional morphology, creating the overall perception of it as a lengthy uphill climb that is destined to stop short of the goal. Even though these developmental regularities have been deemed to reflect a single universal acquisition order (itself stemming from universal learning principles), they were largely generalized from data from English L2 learners. While studies on other target languages (TLs) also exist, they are less numerous and have been conducted in efforts to validate existing acquisition theories, as opposed to specifically focusing on what might be similar and different in the course of learning these different TLs.

Despite the lack of studies explicitly comparing different TLs, some differences appear to emerge across single-TL studies, however coarse cross-study comparisons may be (complicated by differences in the TLs, learners' L1s and ages, and contexts of learning). Learners of TLs other than English appear to achieve higher accuracy on inflectional morphology earlier in development. Thus, learners' difficulties with inflectional morphology in these morphologically richer TLs are not merely "scaled up" difficulties experienced by learners of English, and the accuracy they achieve would be radically underpredicted if one were to simply extrapolate from English learners' data. Psycholinguistic data are more mixed but suggest intriguing possibilities with respect to learners of morphologically richer languages relying less on whole-form storage of morphologically complex words than learners of English.

Most studies that have examined production accuracy did not concern themselves with the specific ways in which learners' utterances deviated from the target. However, the few cases that took this direction can spark new lines of inquiry that are theoretically meaningful. Even those studies that have interpreted different error types as theoretically relevant (e.g., Herschensohn, 2001) have done so ad hoc. Some potentially relevant distinctions that have emerged concern omission of inflectional endings, substitution of inflected forms for one another, and uses of non-finite forms in place of finite ones. Meanwhile, a focus on error types has the potential for formulating fine-grained predictions with respect to how learning paths may be different for different TLs.

In this dissertation, therefore, cross-linguistic data will be analyzed with the express purpose of drawing comparisons between learning processes as reflected by different error types. In Chapter 2, I will apply existing theoretical accounts of L2

learning to the task of explaining the data summarized above, particularly contrasting the predictions of syntactic and general-cognitive accounts with respect to the existence of any cross-linguistic differences, their nature, and implications for the learning mechanisms they assume.

## Chapter 2: Accounts of Morphological Development

Theoretical accounts of L2 learning difficulty followed tend to emphasize either issues of syntactic representation versus performance or general-cognitive processing properties, such as salience broadly construed. I will review each approach in turn, first summarizing the positions within the respective literatures, focusing, where possible, on cross-linguistic data, and conclude by laying out that approach's assumptions about the process of learning, as well as any extensions it permits to cross-linguistic comparisons. I will additionally highlight any outstanding issues within each approach that could benefit from cross-linguistic analysis and a detailed examination of learner error types.

### 2.1 Syntactic Competence

Several syntactic accounts have been proposed to explain L1 and L2 data, situating the acquisition and learning of morphology within the development of grammatical competence. These accounts take differing stances with respect to the nature of the knowledge learners start out with ("the initial state"), the role of Universal Grammar in the learning process, as well as the scope of its influence, and the interaction between the syntactic properties of the L1 and those of the L2. Due to its interest in linguistic representation over performance, the UG tradition has centered on linking patterns in learner performance (e.g., errors in the use of overt IM) to the syntactic representations underlying their behavior or on denying such links. The specifics of the morphology-syntax links have been a matter of intense internal debate in this literature, and the issue of their relative sequencing and, thus, causal relationship, has been theory-carrying.

At one extreme, learner difficulties with IM have been taken to reflect underlying representational deficits, which are either viewed as permanent (Clahsen, 1988) or gradually disappearing. At the other end of the spectrum, accounts such as the Missing Surface Inflection Hypothesis (Haznedar, 1997) have posited intact syntactic representations in learners and placed the source of the difficulty with IM solely at the stage of PF (phonetic form). That is, the difficulty lies in mapping the feature-marked forms (correctly generated by the syntax) to morphological forms of the specific lexical item. Effectively, this dissolves any ties between observed morphology and abstract syntax by absolving syntax from any responsibility for learner errors.

Positioned between these two extremes are theoretical accounts that do acknowledge the connection between morphology and syntactic representations and have concerned themselves with clarifying which one drives the other. First, according to early influential views, morphological forms in both L1 and L2 development give rise to the functional projections that enable them: e.g., in English it is the copulas and third-person singular *-s* that trigger the emergence of the AgrP (agreement phrase), complementizers—the emergence of the CP (complementizer phrase), and so on. This means that learners should first achieve accuracy on inflected forms and then start exhibiting facility with the syntactic phenomena associated with those respective projections (e.g., word order operations).

For instance, according to the Minimal Trees Hypothesis (Vainikka & Young-Scholten, 1996), in early grammars only lexical projections are present, while functional projections are missing altogether, and syntactic structure is not projected above the level of the VP (verb phrase). Exposure to morphologically complex forms through input then

allows the learner to gradually build those higher projections. This proposal was meant to account for opposing patterns in child L1 and adult L2 acquisition: while morphological features expressed through affixes emerge earlier than free morphology in child L1 acquisition, in L2 learning the pattern is reversed and it is bound morphemes that present a challenge. Another implementation of this idea (“morphology causes syntax”), the Valueless Features Hypothesis (Eubank, 1994) proposes that the syntactic projections themselves are present from the beginning but their syntactic features are “inert”—not specified as either strong or weak, and thus rely on exposure to morphologically inflected forms to be determined as such.

Zobl and Liceras (1994), in contrast, argue that all functional projections are present from the start and the difficulties arise due to the marked way of merging inflections with lexical heads in English, which requires the lowering of affixes from IP onto V, compared to the unmarked way requiring the raising of V to I (e.g., French). Even though this proposal treats all features at the same syntactic level as having the same difficulty, its merit may be in postulating an asymmetry in difficulty between movement operations and, by virtue of that, in allowing for cross-linguistic differences in L2 learning difficulty.

Building on the distinction between lexical and functional projections, Hawkins (2001) brings more nuance to this approach. In keeping with Vainikka and Young-Scholten (1996), he contends that the initial state for L2 may be comprised of lexical projections transferred from L1 (which may persist or rapidly restructure under the influence of L2 input), whereas functional projections emerge gradually. The gradual manner of development applies both to the emergence of projections in relation to one

another (IP before CP), as well as to sets of morphemes expressed at the level of the same functional projection. For example, head-complement features develop earlier (*Aux be*) than non-local binding relations (tense), with specifier-head relations (agreement of I with its specifier—the subject) being last. Allowing for this granularity among features belonging to the same level allows Hawkins to incorporate L1 influence into the model. He argues that L1 transfer will facilitate learning at the stage when it is relevant and the necessary projections are in place. He uses as an example the development of third person singular *-s* and past-tense *-ed* in Spanish and Japanese L1 learners. When both groups of learners have an underspecified IP that is only used as a landing site for moved auxiliary verbs, both groups perform poorly on third person singular, even though transfer would be expected for Spanish speakers. When learners begin to realize morphology in IP, however, Spanish L1 speakers are able to benefit from L1 transfer and become more accurate on third-person singular than on past tense *-ed*, whereas Japanese L1 learners remain equally inaccurate on both features.

Conversely, another family of accounts shows that syntax may be in place before the morphology that relies on it. Demonstrations to this effect have gone hand-in-hand with arguing not merely for syntax before morphology, but for a complete break between the two. For example, in one report an endstate learner (Lardiere 1998a, b, L1 Chinese) mastered nominative and accusative case assignment, adverb and negation placement, and non-null subjects near perfectly—all taken to be indicative of robust syntactic projections well above the VP, while tense and agreement were still impaired. Similar observations have been made about an L1-Turkish child learner of English (Haznedar & Schwartz, 1997) and a sample of L1-Russian child learners (Ionin & Wexler, 2010).



Notably, unbound morphology (such as auxiliaries and copulas) were supplied more accurately than were bound inflected forms. Studies conducted in the processability framework further exemplify this sequencing pattern, even though PT does not directly predict this difference (Bonilla, 2015; Dyson, 2009). Arguably, word-movement operations may be available to learners without involving much syntax per se: it is impossible to say with certainty whether the processes by which learners arrive at seemingly “nativelike” word orders or case, for example, are syntactic in the same way as a native speaker’s or whether these nativelike productions are the product of pattern-matching strategies of a general-cognitive kind (Bley-Vroman, 1997).

While syntactic accounts have captured an important generalization in the data (higher accuracy on syntactic than morphological phenomena), their explanations of it miss the mark in several ways. First, relegating morphological errors to the PF or the morpholexicon does not account for the regularities in learner errors so painstakingly documented. As argued by Franceschina (2001), if the existence of patterns in learner errors is used to argue against a *global* syntactic breakdown in the L2 (otherwise errors would be randomly distributed), the same reasoning should apply to the morphology and PF modules. If the L2 PF is deficient, it should be across the board, which is not the case, as the studies reviewed in Chapter 1 showed. In other words, why would the phonological module respect syntactic distinctions—such as the difference between case and tense, for instance (Lardiere, 1998)?

Claims of native-like competence (despite optionality in performance) in the endstate (Lardiere, 1998; 2006; White, 2003)—enabled by continued access to UG—exonerate syntax as the source of difficulty but also enable an inconvenient restatement

of the problem. From asking “why does L2 syntax not generate appropriately inflected forms?” one is left wondering why L2 morphology (or morpholexicon, or PF) sometimes does and at other times does not generate target-like inflected forms, largely following syntactic distinctions. The answer to this question matters from both theoretical and practical standpoints, and claiming that the knowledge is “there” but we cannot observe or measure it offers little by way of implications for instruction, testing, or materials design. Such reasoning is also inconsistent conceptually: any successful suppliance of morphology is attributed to the operation of target-like syntax, whereas any departures from the target are somehow not syntactic but morphological or PF-related. Without any means to independently determine the source of any given utterance, such explanations are not credible.

One promising feature of syntactic theories that can resolve this contradiction is their acknowledgment of morpholexical learning as a necessary addition to UG and syntactic knowledge. However, the specifics of such learning are poorly understood and have largely been left outside the scope of syntactic development theories in L2. Learning theories developed in the L1 acquisition literature have been more explicit and have integrated UG with input-driven distributional learning (Lidz & Gagliardi, 2015; Pinker, 1984; Yang, 2002). Their focus on learnability, rather than solely competence and performance, provides a way to account for the regularities in errors by linking them to aspects of the grammar that generated them without dismissing them as being solely reflective of performance. Extending this approach to L2 learning can help L2 syntactic theories navigate some of the contradictions described above.

Employing a different syntactic formalism, one rooted in lexical-functional grammar, processability theory (PT—Pienemann, 2005) has linked syntactic complexity to processing difficulty. It posits an acquisition order proceeding from features expressed at the level of lexical items (e.g., noun plurals in English), to the within-phrasal level (e.g., agreement between article and noun), and, finally, the level of interphrasal links (e.g., subject-verb agreement). Thus, difficulty increases with the need for feature coordination across phrases, compared to the difficulty of phenomena that are expressed locally on lexical heads. This approach has spawned a lot of cross-linguistic evidence but has focused on cross-validating the proposed acquisition orders on different TLs, without an interest in any differences in rates of development not accommodated or predicted by it.

Even though PT posits an explicit processing machinery, it employs the processor to recast syntactic phenomena in psychological terms, as opposed to detailing how it serves as a learning mechanism. The relation that PT posits between ease of processing and ease of acquisition is one of identity. The two need not be the same, as has been claimed for L1 (Pinker, 1984): for example, producing agreement across intervening material is memory-costly even for adults fully competent in the L1, as exemplified by agreement attraction errors such as *The key to the cabinets \*are on the table*. For a child, errors would be expected on such structures even after they have been “acquired”. In contrast, ease of acquisition is a function of the availability of relevant evidence in the input.

While it is easy to dismiss any production errors as caused by processing bottlenecks when focusing on one target language, a true test for a purely processing-

driven explanation is its ability to hold across multiple TLs. Learners of all TLs should be equally susceptible to processing difficulties and memory breakdowns. Any residual differences among TLs would be attributable to learnability and the differences in the TLs' surface realizations of morphology, such as the number of overtly expressed morphemes, number of homonymous morphemes, or the morphemes' perceptual properties. After all, why should it be more difficult for learners of English to produce an utterance such as *My child walks to school* with correct inflectional marking than it would be for a learner of German?

**Conclusions.** Despite discrepancies among the L2 morpheme order studies, a recurring theme across several research literatures has emerged concerning the higher difficulty of bound morphemes compared to free morphemes (e.g., auxiliaries) (e.g., Pienemann, 2005; Vainikka & Young-Scholten, 1996; Zobl & Liceras, 1994). True to their mission of characterizing linguistic competence, linguistic theories have put more weight on emphasizing what is common, rather than unique, to learning different TLs: the development (or preexisting presence) of syntactic projections, the interface between syntax and phonological form, or the differences and similarities between L1 and TL.

Extensions of syntactic theories to predicting the comparative ease (and difficulty) of learning for different TLs are less clear and fraught with contradictions. To the extent that the core syntactic architecture proposed by linguistic theory is shared among languages, so should be learning difficulty. Thus, third-person singular marking should cause comparable difficulty for learners of different TLs from a syntactic standpoint, even though it could be additionally complicated by the idiosyncrasies of different TLs' morpholexicons. On the other hand, by virtue of presupposing abstract syntactic

knowledge in learners, these theories leave room for synergies of the kind proposed in the L1 acquisition literature (Legate & Yang, 2007; Yang, 2002), which arise from the activation of abstract morphosyntactic features whenever morphemes are encountered. For example, encountering a verb overtly marked for second-person singular would contribute to the learning of third-person singular, by strengthening the “singular” feature. On this proposal, the parameters of learners’ grammars change in response to encountering inflected forms, not just the lexical entries of the verbs. This approach reconciles a common syntactic architecture shared by different TLs with room for cross-linguistic differences that would be due to the specifics of their morphemes’ realizations. On this view, TLs with inflectional paradigms in which more contrasts are marked overtly would have lower learning difficulty than those depriving learners of overt evidence in input. Theories of L2 syntax have not pursued this approach and have mainly placed learners’ underperformance on inflection at the level of morpholexicon—not morphosyntax.

Another point of uncertainty in a cross-linguistic extension is whether any learnability benefits afforded by the presence of multiple overtly marked morphemes would be offset by the increased need for *morpholexical* learning. For example, having six distinct inflected forms in the present tense—rather than just two (as in English, one of which is zero-marked)—may strengthen the [Tense+] grammar faster. However, it may also mean that for each lexical verb six inflected forms need to be learned (along with any potential phonological alternations), as opposed to just one.<sup>7</sup> All of these

---

<sup>7</sup> This does not imply that all six forms would be necessarily learned together or intentionally. Over time, however, all six would have to be integrated by the learner into the mental lexicon for successful production.

apparent extensions are highly speculative and have not, to my knowledge, been advanced from within these literatures themselves.

None of the L2 syntactic theories have integrated learnability into their accounts of how grammatical competence is acquired and performed in the domain of inflection. Such an integration would advance theory building by linking the two and explaining how imperfect performance may arise from a relatively sophisticated grammar. However, a syntactic lens has much to offer to a characterization of L2 learning, especially in developing a taxonomy of learner errors. The distinction between competence and performance, or a tripartite split into competence, performance, and learnability may provide a path to viewing certain error types as reflecting the acquisition of abstract principles of grammar (e.g., agreement), others as reflecting the accessibility of inflected forms during production, and yet another group as being indicative of the accumulation of morpholexical knowledge.

## 2.2 General-cognitive Approaches

The linguistic accounts reviewed in the previous section acknowledge the need for learning, in addition to any innate syntactic knowledge they posit, and have to be supplemented by proposals detailing such learning. The general-cognitive approach (GC) can supply this missing piece of the puzzle by specifying the factors influencing morpholexical learning, which has been outside the scope of linguistically oriented theories. One alternative approach to explaining learning sequences has been to study the properties of the morphemes in question and classify them along dimensions such as perceptual salience (e.g., syllabic status), homonymy, transparency of form-to-meaning mapping, and so on. Similar factors have been proposed for child L1 acquisition but as

part of broader learning theories rooted in the syntactic tradition (Pinker, 1984; Yang, 2002).

In SLA, an influential account positing a general cognitive explanation was advanced by Goldschneider and DeKeyser (2001). On this view, a feature's salience is a confluence of factors that include its syntactic function along with perceptual, semantic, and frequency characteristics. Together these variables explained as much as 70% of the variance in acquisition orders of English morphemes in ESL.

This approach differs from syntactic theories in its key assumptions about learning. First, by situating learning difficulty in individual morphemes one tacitly assumes that morphemes are both adequately segmented in the input and then learned one-by-one. This contrasts with the notion entertained by some syntactic theorists that learners entertain abstract hypotheses at a level above any individual elements—one of grammatical features and parameter settings for the TL grammar (cf. Legate & Yang, 2007). This is evident from the fact that only the frequency counts of individual morphemes matter for predicting a feature's learning trajectory, whereas any indirect evidence available in the input does not. Such indirect evidence is supplied by the oppositions entered into by individual morphemes: in the simplest case—the opposition between a morpheme itself and its absence. Furthermore, the effects of frequency are considered without reference to any typological patterns, unlike in the L1 acquisition literature, where the effects of frequency are deemed to vary depend on the learning mechanism one posits. For example, on Pinker's (1984) proposal the learning mechanism needs radically different frequencies for agglutinated versus fusional forms to achieve equivalent strength of knowledge. Even though it is acknowledged by the GC proposal

that some typological configurations may result in higher salience (e.g., agglutinative because of the transparency of each morpheme being mapped to only one feature), any facilitation resulting from it has not been elaborated or weighed against the other aspects of salience.

The second assumption is a summative, flat view of the effects of saliency factors without postulating any hierarchy among them. Because all predictors are treated equally, the model has descriptive but not explanatory power. If feature A scores high on variables 1, 2 but not 3, the predicted effect is identical to a low score on variable 1 coupled with high scores on 2 and 3. The inputs to morphological development included in the model capture notable regularities in the data and allow the formulation of fine-grained predictions of learning difficulty if one compares one morpheme to another.

Extending this model to comparisons between entire morphological systems (of different TLs) is not straightforward and would require adjustments that go beyond the original proposal, considering that it was not formulated for cross-linguistic use. Since difficulty is a property of individual morphemes on this view, TLs with a higher number of them in their inflectional paradigm should have an overall slower rate of learning. This is because even in the most beneficial scenario each additional morpheme (expressed overtly) has a non-zero difficulty score. Conversely, a TL with fewer overtly marked morphological contrasts would seem preferable.

Due to the flat structure within the set of predictors it proposes, the GC model formalized in 2000 and 2005 (Goldschneider & DeKeyser, 2000; DeKeyser, 2005) mathematically implies no difference between a TL with many morphemes of low individual difficulty and a TL with only a few high-difficulty ones (English). To build



into the model any synergies (such as the added transparency from agglutination), one would presumably lower difficulty scores for all the morphemes, but it is unclear by how much.<sup>8</sup> In a contrast with syntactic accounts, in assessing learning difficulty the GC model has placed less interest in the structural relationships among morphemes, including the possibility that difficulty may well lie in a morpheme's very uniqueness as the only *overtly* marked form in a paradigm, not just its low frequency per se.

Similar to syntactic accounts, the GC approach has not pursued examining error types, even though interesting possibilities arise from attempting to use GC principles to predict error differences cross-linguistically. On an extremely speculative reading of it, it appears consistent with predicting infinitival or unmarked forms in place of inflected forms: not using any morpheme would seem easier than a morpheme with non-zero difficulty. At the same time, substitutions of inflected forms for one another are also not antithetical to this proposal. However, on my reading, substitutions would have to be directed towards more salient morphemes, even though it is not clear which salience dimensions would be prioritized (e.g., morphological, phonological, or syntactic).

For example, if it is claimed that third-person singular (English: *-s*, German *-t*) is difficult because it is non-syllabic, bound, and semantically redundant, learners could be predicted to produce any number of alternative “easier” forms, in theory—ranging from uninflected forms to an competitor that is syllabic (e.g., *-en*) or more frequent. On the other hand, substitutions that do not go in the direction of increased salience would be

---

<sup>8</sup> In a similar vein, the GC account does not explain task-related discrepancies: why would the same morpheme with all the same properties be more easily comprehended or judged as grammatical than it is produced? Any accounts of such differences are of necessity ad hoc and external to the proposal.

harder to account for. Any uses of forms that do not exist, furthermore, would be the most difficult to accommodate.

**Conclusions.** The theoretical approaches to L2 grammatical difficulty summarized in this chapter have provided complementary perspectives on morphological learning but have yet to be integrated. Their integration, however, is not straightforward.

The two approaches (general-cognitive, syntactic) are not mutually exclusive: the need for distributional learning is undisputed in the syntactic literature, whereas general-cognitive approaches include a syntactic dimension of difficulty. While being broadly complementary, the syntactic and GC lines of research differ in their fundamental views of *what* it is that learners learn. On the GC view, it is individual morphemes (or form-meaning mappings); on the syntactic view, it is abstract features and parameter settings. Thus, when applied to the same data, syntactic and general cognitive approaches can jointly provide a new perspective on the nature of the learning mechanism in L2.

To achieve this, the path forward could be to select one domain of learner behavior that is meaningful to most parties, such as oral spontaneous production, or written production, and then spell out the patterns of performance that would be expected under each theoretical proposal. In doing so, a focus on different error types and cross-linguistic comparison would be instrumental to pushing each approach to maximum clarity about its predictions, which are currently hard to extrapolate.

Cross-linguistic comparisons can be informative for separating learning difficulty and processing difficulty, which is something that both approaches could benefit from. While language processing systems share largely the same working memory and production architectures (Levelt, 1989), their linguistic systems present vastly differing

amounts of learning data in their inflectional paradigms. Therefore, difficulties stemming from processing bottlenecks should be similar across TLs, assuming similar processing demands for any given structure. Conversely, differences across TLs would speak in favor of deeper learnability challenges associated with different TLs' systems.

Additional insight is also to be gained from considering learner errors as well. As pointed out in Chapter 1, detailed accounts of error types have not been pursued in previous research; nor have the theories outlined in this chapter made references to error types and what they would signify within the respective frameworks. When applied on top of a cross-linguistic approach, an additional analysis of errors would allow to test deeper and finer-grained hypotheses about the L2 learning processes. For example, if learners of Italian are more accurate at producing verbs inflected for the third person singular than are learners of English, does that mean they have acquired more robust representations of agreement, the exact lexical items tested, or is the inflectional marker more salient? Posing this question cross-linguistically would not be sufficient.

An error analysis in this context, however, could offer answers. One could speculate that errors related to the use of uninflected or infinitival forms in inflectionally rich TLs should match those observed in learners of English if the source of the difficulty lies in processing resource limitations. If, however, learners of inflectionally richer TLs make other kinds of errors more readily (e.g., substituting other inflected forms), a learnability benefit can be inferred for the richer TL. By contrast, different frequencies of errors related to the use of inexistent inflected forms could reflect the processing *costs* of richer inflectional paradigms—such as more competition in the mental lexicon during the retrieval of inflected forms.

The purpose of this dissertation is, therefore, to produce data of a kind not examined or generated before: data that are both cross-linguistic and granular at the level of error types. At a minimum, this chapter has shown that both theoretical approaches can use these kinds of data to refine their predictions and test some of the hypotheses that have been advanced about the sources of learners' problems with inflectional morphology. In addition, when both theoretical approaches are applied to the same data, their assumptions about the nature of learning can also be tested against each other.

## Chapter 3: Research Questions and Motivations

Chapter 1 showed that cross-linguistic L2 acquisition data are exceptionally sparse, and any comparisons have to be pieced together from disparate studies originating from a variety of SLA research fields, driven by different agendas and utilizing different methods. Despite this sparseness, some tentative observations can be drawn from the handful of studies that have investigated diverse target languages (TLs). As Chapter 2 demonstrated, neither syntactic nor general-cognitive accounts of L2 learning can explain these observations or formulate predictions as to what the data “should” look like if a cross-linguistic comparison were to be attempted. Nevertheless, cross-linguistic research, coupled with a detailed examination of learner errors, can offer insight into the learning of L2 morphology that cannot be achieved by either syntactic or general-cognitive theories based on single target-language data alone.

This chapter will first recapitulate key gaps in the research on L2 morphological learning, especially as they relate to predicted difficulties of the morphological systems of different TLs. It will connect these gaps to the research reported in this dissertation. Finally, the target languages selected for the current research will be described from the standpoint of their verbal inflectional morphology and the systems of verb classes.

### 3.1 Benefits of Paradigm Complexity

Data on the learning of TLs other than English (presented in Chapter 1) appear to suggest that learners achieve higher accuracy at producing inflectional morphology earlier in development than do learners of English. However, the obvious caveat here is that English and other TLs have not been compared head-to-head within the same studies,

to the best of my knowledge. This is not predicted by either syntactic or general-cognitive accounts.

According to the former, morpholexical learning should occur gradually and, as implied by its item-by-item nature, should be slower the more there is to learn. On my reading of the latter account (as formalized in DeKeyser, 2005 and Goldschneider & DeKeyser, 2000), the difficulty of learning an inflectional paradigm can be expressed as arising bottom-up from the individual difficulties of the morphemes in question. This implies that a greater total number of morphemes should result in slower learning overall.

However, a number of studies across different subfields of language science have indicated that a longer “to-do” list may not be detrimental to the task of learning. In L1 acquisition, similarly counterintuitive patterns of findings have been dealt with by positing competing grammars that coexist in a child’s mind (Yang, 2002). These grammars presumably include conflicting settings for a given parameter—for example, [+Tense] for a grammar resembling English or [-Tense] for a grammar resembling Chinese—and compete until one wins.

For example, a child acquiring English receives evidence in favor of the [+Tense] setting not only from past-tense forms, but also—indirectly—from present-tense forms that express agreement, which strengthens the tensed grammar until it wins the competition with the grammar without tense (Guasti, 2002; Legate & Yang, 2007). The rankings of languages based on the amount of evidence compatible with a tensed and a tenseless grammar were shown to mirror the rates of disappearance of bare infinitives in child language (Legate & Yang, 2007). Similar proposals, couched in terms of competing grammars, have also been advanced for L2 learning, however, without catching on

(Amaral & Roeper, 2014) or advancing hypotheses as to how grammar competition is resolved.

Further evidence suggesting that complexity, broadly construed, may be beneficial during learning have originated from the cognitive science and information theory. In artificial grammar learning (Thompson & Newport, 2007), pseudo-syntactic frames consisting of syllables (cf., English *am\_\_ing*) were learned better if the variability of the intervening syllables was higher. Maximally diverse filler syllables likely highlighted the stability of the frame bookending them. Extended to L2 morphology, this could mean that encountering lexical items inflected in more diverse ways may strengthen learners' lexical representations. This insight parallels observations from L1 acquisition, where the diversity of contexts in which parents used finiteness morphemes (e.g., *is*) predicted children's productive use not only of the morpheme in question nine months later, but also of other finiteness markers with the same feature specification (third-person singular *-s*) 15 months later (Rispoli & Hadley, 2012). Although here the "complexity" in these examples extends to the contexts surrounding the elements of interest, the value of such facilitatory effects lies in opening the door to speculation about any effects of *paradigmatic* complexity more narrowly.

Richer inflectional paradigms' benefit may lie in putting communicative pressure on the learner and inducing attention to morphological elements, forcing the learner to discover the semantic distinctions they encode (child L1 acquisition—Dressler et al., 2007). Offering an L2 parallel to this proposal, an artificial L2 learning experiment (Fedzechkina, Jaeger, & Newport, 2012) showed that the learning of differential object marking was facilitated (as reflected by higher accuracy scores) when the morphemes

signaling it were the only cues to thematic roles. Conversely, learning was impeded when the information encoded by DOM was recoverable from word order as well. These results generally echo the notion of semantic salience proposed under general cognitive accounts (e.g., VanPatten, 2004): if a morpheme is not critical for comprehension, the semantic dimension of its salience is lowered, and learning happens at a slower rate.

When applied to distinctions between morphemes in the same paradigm (such as present-tense indicative) across different languages, it is unclear whether communicative pressure would predict differential outcomes for individual morphemes within it. Since inflectionally rich languages also tend to be PRO-drop, which makes morphology more essential to comprehension, more controlled studies will be needed to separate the impact of paradigm richness *per se* from the confounding communicative pressure introduced by missing overt subjects.

On the connectionist account, the presence of multiple non-zero marked forms in the paradigm may increase the validity of inflectional morphemes as cues to meaning. In this literature, learning of inflectional paradigms is a process of strengthening form-meaning mappings (Kempe & MacWhinney, 1998). For example, Russian and German systems of noun inflections in the nominative and the accusative cases differ in the number of unique inflected forms (higher in Russian), the average uniqueness of inflections—or the ratio of unique inflected forms to the total number of possible forms (higher in German), as well as the validity of inflection as a cue to thematic roles (defined as the reliability with which agent/patient roles can be predicted from the case).

In Russian, case is expressed synthetically, whereas in German it is marked by the combination of the morphology of the article (and sometimes also an affix). Learners of



both languages completed a forced-choice picture identification task, where they had to select the picture representing the agent of a sentence they just heard. Controlling for familiarity with their respective L2s, learners of Russian made fewer errors than did learners of German, whose accuracy improved, however, with additional self-reported experience. This led the researchers to conclude that case marking is learned faster in Russian. In addition, a connectionist network simulation replicated the advantage of Russian in correctly selecting the second noun as the agent in a case-marked OVS sentence. This demonstrates the joint benefits of paradigm richness and cue reliability, which were conflated in this study.

### 3.2 Potential Trade-offs between Learning and Processing

There are two opposing psycholinguistic implications of paradigm complexity for online processing. On the one hand, current psycholinguistic models suggest that the net effect of neighborhood and cohort density should be negative rather than facilitative. For example, in models of spoken word recognition (e.g., Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Zwitserlood, 1989) increased cohort size or neighborhood size (Luce & Pisoni, 1998) result in slower recognition due to the competition of lexical items during selection. However, in models of speech production the presence of phonologically similar competitors facilitates word naming once the lemma has been selected (Lupker, 1982; Meyer & Schriefers, 1991).

When considering the psycholinguistic repercussions of complexity, one inevitably ventures into lexical territory and the interplay between phonology and morphology, or broadly—the array of tools that express morphological contrasts (e.g., root vowel changes and quasi-regularities associated with verb classes). Therefore,

inflectional paradigm complexity may differentially affect phenomena that are closer to the lexical or the combinatorial ends of the spectrum.

By contrast, when one examines *learning*, as opposed to processing, some psycholinguistic evidence points to denser lexical neighborhoods being beneficial. In L1 acquisition, root infinitives have been observed to be less frequent with verbs from dense neighborhoods than with those from sparse ones (Hoover, Storkel, & Rice, 2012). The authors interpreted their findings as indicating that verbs from more dense neighborhoods are represented in the mental lexicon with more phonological detail and are, therefore, more accessible in production. They also proposed that the existence of phonologically similar words pushes the learner to differentiate any given word's phonological form more finely. By extension, this would apply to inflected forms stored in the lexicon as well and afford learnability benefits in TLs that contain more such forms. Learnability evidence from artificial language learning studies also suggests that high input variability may facilitate regularization (Hudson Kam & Newport, 2007, 2009). Admittedly, regularization of this kind is a double-edged sword: whereas the learning of the most regular pattern is facilitated, input is essentially being coerced toward it, resulting in intake that does not match the input's distributional properties.

The question of tradeoffs, therefore, will rely on one's definition of learning. For example, if by learning we mean the disappearance of bare (unmarked) or infinitival forms from learners' productions, then exposure to varied inflected forms in input may be facilitative. If learning is equated with producing inflected forms correct down to the last letter, then this type of input may be adverse in so far as a greater number of forms in the mental lexicon leads to greater competition during retrieval. Thus, a distinction between

psycholinguistic processing and outcomes of the learning process can be studied by recourse to error types.

### 3.3 The Current Research

The most obvious gap in research that this dissertation seeks to address is the very lack of a set of cross-linguistic phenomena to explain. A bona fide cross-linguistic examination of morphological learning can provide theoretically relevant insights, and specifically:

- Refine syntactic and general-cognitive approaches by serving as a testing ground for their predictions, especially with respect to processing versus learnability;
- Test the assumptions about the learning process implicit in syntactic and general-cognitive theories against the same set of data, particularly any increases in learning difficulty that accompany increases (or decreases?) in the complexity of a TL's morphological system.

In addition, a close focus on learner productions and the ways in which they deviate from the target can lend further insights into learning:

- Link aspects of performance on the *same* task to different mental processes and levels of grammatical knowledge or its abstraction (cf. the common practice of using performance on different tasks to tap different types or levels of knowledge): e.g., morpholexical versus morphosyntactic learning; learning of abstract principles related to agreement versus flawless online production of the exact inflected form required;
- Characterize the learning of entire grammatical systems (of varying complexity) rather than of single morphemes.

Thus, the present research has combined a cross-linguistic focus with an interest in learner errors.

**Research questions and hypotheses.** To contribute to existing accounts of L2 morphological development and resolve some of their contradictions, the present research has carried out cross-linguistic comparisons of rates of learning in L2 learners. The goal of this dissertation is to explicitly contrast the repercussions of TL paradigm richness for the overall use of inflected versus non-inflected forms, as well as for the retrieval of correct forms during production. This research endeavors to provide insight into the following questions:

1. Are rates of learning of L2 inflectional morphology proportionate to the complexity of the TL morphological system, expressed as the number of distinctions made in the paradigm?
  - a. As the TL morphological richness increases, will learners become proportionately less accurate in written production at marking agreement through inflectional morphology on the verb?
  - b. Does morphological richness impact different kinds of knowledge differently?—Are there differences among TLs differing in complexity for errors involving more combinatorial, rule-like processes, compared to errors involving aspects of inflection involving lexical knowledge (and instantiated as knowledge about lexical items)?
2. Are there tradeoffs between the effects of morphological richness on the mastery of agreement as a concept and its influence on online retrieval (potentially negative)?

- a. Does the percentage of uninflected, substituted inflected forms, and inexistent forms (e.g., formed following a compositional rule correctly but failing to respect subregular morphophonological patterns) differ among languages of different morphological richness?
  - b. Will this ratio change at different speeds among the different TLs? Specifically, will completely uninflected (or seemingly infinitival) forms drop off more sharply (i.e., starting at lower proficiency levels) in richer languages than in poorer languages?
3. Can interlanguage phonological processes account for the pattern of errors in German?
    - a. Does accuracy differ between inflectional endings that are phonologically felicitous, compared to endings that are phonologically marked or typically dispreferred in the interlanguage?
    - b. When errors of substitution occur, do they flow in the direction from more phonologically complex to less phonologically complex realizations?

Due to the exploratory nature of this research, detailed hypotheses were impossible to formulate from the standpoint of either syntactic or general-cognitive accounts.

### 3.4 Target languages and their inflectional systems

To answer the research questions, TLs of varying paradigmatic complexity have been selected, for which publicly accessible learner corpora were available through the Merlin project, that includes learner productions in German, Czech, and Italian.

In previous research (Kempe & MacWhinney, 1998), complexity was conceptualized along three dimensions. First, of interest is the number of dimensions

expressed in a given paradigm: for example, such common ones as number, person, gender, and less commonly grammaticized distinctions, such as animacy. Second, the interaction of the number of dimensions with the number of levels each dimension can take (e.g., gender—masculine, feminine or masculine, feminine, neuter) can be expressed as their product, which reflects the total number of cells in the paradigm. For example, for German, which expresses the dimensions of person (with three levels) and number (two levels) on the verb, but not gender, that total is six. Third, the total of all cells (or forms) can be expressed relative to the number of cells containing *unique* elements: in German, there are three cells representing the combinations of persons with the plural number, but two out of the three combinations are marked by the same inflection on the verb, *-en*, yielding only two unique forms. This would produce a uniqueness ratio of  $0.67 = 2/3$ —one unique affix out of possible three.

In the present study, any references to “complexity” refer to the structural complexity of the TL—that is, the number of choices (distinct morphemes) contained in its morphological paradigm, and not *putative* difficulty for the learner. The descriptions of each TL’s complexity below will focus on the indicative mood, since the subjunctive or even the imperative are restricted both in use and in pedagogical emphasis. By the time the subjunctive is introduced in instruction, the development of inflection and agreement should be well underway. Following this logic, the passive voice was also excluded, since it is an advanced structure which tends to be expressed analytically.

The remainder of the chapter will first introduce the verbal class systems of each TL, moving to brief summaries of their tense systems (as relevant to the present

research), and finally presenting each tense paradigm and the number of morphological distinctions they express.

*Inflectional systems of the target languages: verb classes.* The three languages represent major language groups within the Indo-European family—Germanic, Romance, and Slavic—and express similar ranges of grammatical meaning, such as tense, aspect, agreement. All three use the Latin alphabet. In verbal conjugation, all three have a system of verb classes, which give rise to morphophonological sub-regularities. In the case of German, there is a regular/irregular distinction, which is the most visible in the formation of the preterite and the past participle (which is used to form the perfect, pluperfect, analytical future, and the passive voice), but is also evident in root vowel changes in the present. The vowel changes in the present tense stem from a distinct phonological process, Umlaut, and involve vowel harmony conditioned by the no longer surviving phonology of the affix. By contrast, the differences in vowels between the infinitive, preterite, and past participle go back to pre-Indo-European processes of Ablaut.

In the case of Italian, the three verb classes are signaled by thematic vowels, resembling the system of Latin; in addition, there is a small number of truly irregular verbs (e.g., have, go, be, etc.). Finally, Czech has an even more elaborate verbal class system: depending on the linguistic analysis used, as few as four and as many as six classes are identified (e.g., Janda & Townsend, 2002). Regardless of the merits of each analysis, it is fair to argue that the Czech system is the most intricate of the three. Since arbitrary verb classes are present in all three languages, it bears emphasizing that the recognition of class membership by learners is not the focus of the present research. One could argue that within the general domain of inflection these sub-regularities of root

transformations represent a more lexicalized process than the application of inflectional morphemes. For this reason, it is our view that the application of inflectional endings is and should be treated as a procedure separable from phonological root and suffix alternations at morpheme boundaries. This is why any inflected forms that only differ from each other by some element signaling class membership (e.g., Italian—thematic vowel; Czech—suffix) will not be counted as distinct as long as the inflectional ending stays constant regardless of class.

*Inflectional systems of the target languages: tenses.* Verbal inflection is expressed in a network of tenses, which are sometimes formed synthetically and sometimes analytically in the TLs. It is the synthetic forms that are the focus of the present study. All three languages have at least two synthetic tenses. In German, it is the present and the preterite; the perfect and pluperfect are expressed analytically through the combination of an auxiliary (forms of *have*, *be*) with the past participle which remains unchanged. Since *have* and *be* are the only components of this form that are conditioned by agreement and are likely to be highly overlearned, we will not consider these two tenses as contributing to paradigm richness in a major sense. Likewise, the analytical future in German is formed with the auxiliary *werden* (cf., English ‘*will*’) and the infinitive, and will not be factored into the total complexity picture. In Italian, there are four synthetic tenses—the present, the imperfect past, the absolute past, and the simple future. Similar to German, there is also a perfect and a pluperfect tense, and a future perfect, differing from each other only in the tense of the auxiliary. However, in contrast to German, it is not only the auxiliaries but also the participles that agree with the subject in gender and number, expressed through four distinct forms (masculine singular,



feminine singular, masculine plural, feminine plural)—but only with unaccusative verbs requiring the auxiliary *essere*. Since the participle forms are identical for the three tenses, we can count the four participle forms once. In Czech, there is a synthetic present, which has a present meaning for imperfective verbs and a future meaning for perfective ones; the past is expressed analytically through a combination of the auxiliary *be* and a past participle—agreeing with the subject in person/number and in gender/animacy/number, respectively. Therefore, maintaining consistency with the treatment of participles in Italian, we can add this tense to Czech’s overall complexity but limit it to the six distinct participle forms. The analytical future in Czech and German was not considered because it is formed as a combination of an auxiliary and the infinitive. In both languages, the grammatical present tense has the capacity to express future meanings (in Czech—obligatorily so for perfective verbs).

In the present tense, all three languages express six number-person combinations in the sense that these are non-zero marked. However, in German the first- and third-person plural forms are homophonous with the infinitive, whereas the second person plural form (informal) has the same inflectional ending as the third person singular but, for irregular verbs only, differs from it in the root vowel. Thus, if we de-emphasize root processes, the second person plural would not be counted as a distinct form, leaving a total of three different morphemes unambiguously different from the infinitival template. Another possibility is to acknowledge that the root differences would merit a separate count and even to count the two other plural forms once, resulting in a total count of up to five distinct forms. In Italian, all six person-number combinations are expressed by distinct morphemes, none of which resemble the infinitive morphology. Even though the

thematic vowels preceding inflectional morphemes vary by verb class, only six forms were counted as contributing to overall complexity (instead of eighteen—six endings multiplied by three possible thematic vowels). Finally, Czech expresses all six person-number contrasts through forms that are distinct from each other and the infinitive. The calculation of the number of forms contributing to complexity is made less straightforward by the interaction of inflectional endings and verb classes. In contrast to Italian, where verb classes mostly determine alternations in the root and do not influence the choice of inflectional morphemes, in Czech the endings themselves differ depending on class. For example, the first-person singular form can end in vowel + *-m* (for two classes), *-i*, or *-u*. All other person-number combinations are more uniform and end in a vowel suffix followed by an ending (same for all classes). On the most conservative calculation, we can ignore the vowel differences between two verbal classes (before *-m*), in parallel to the choice made for the Italian present tense, where we also ignored thematic vowel differences. This amounts to three forms of the first person singular, one of the second person singular (collapsed across all vowel suffixes), three possibilities for the third person singular, one each for first and second person plural (with different vowels preceding the ending), and at least two broadly distinct forms of the third person plural—one ending in *-ou* and the other in *-(vowel+j)i*. This is a total of 11 forms.

In the synthetic past tense, both German and Italian use some combination of the root, a suffix expressing tense, and person-number endings. In German, both the first- and third-person singular forms are zero-marked, on a strict morphological analysis. But because the suffix expressing tense, *-(e)te*, ends in *-e*, and the person-number endings are identical to those of the present tense, the appearance is created that the final vowel of the

tense suffix marks first-person singular. There is no such illusion for irregular verbs, which do not add a suffix but use the preterite form plus person-number ending (cf., steal—stole). For these irregular verbs, the first- and third-person singular forms are identical and not marked overtly. Therefore, in both regular and irregular cases these forms can either not be counted at all or counted as one. This would give a total count ranging from three distinct forms with overt morphology (second person singular, first/third person plural, second person plural) to four (first/third person singular, second person singular, first/third person plural, second person plural). Because of the presence of the tense-expressing suffix, even the first- and third-person plural forms do not resemble the infinitive as much as their present-tense counterparts, so counting them at least once seems fair.

In Italian, the synthetic imperfective tense forms follow the same general “recipe” as in German, represented in six distinct forms: a stem (root plus thematic vowel) is followed by a suffix (-v) and an ending representing the person-number combination. The inflectional endings in the imperfect tense are identical to those of the present indicative. In both Italian and German, the formation of past-tense forms essentially amounts to agglutination, raising an interesting contrast to English. Where past-tense marking is omitted in English, it is inherently uncertain whether tense or agreement is to blame. In German and Italian, by contrast, a dissociation is possible where either the tense- or agreement-bearing morphemes can be omitted or otherwise incorrectly rendered. In addition, the Italian absolute past adds six inflected forms to a learner’s plate, which are derived using a different set of person-number inflectional endings.

In the past tense in Czech, in addition to person and number, gender and animacy are also expressed. This form blurs the distinction between synthetic and analytical expression: even though there is an analytical component to it in the form of an auxiliary (*be*) expressing person-number, gender is expressed on the lexical verb. There are three distinct forms in the singular—for the masculine, feminine, and neuter genders; and three forms in the plural—one for masculine animate (plural), another for masculine inanimate or feminine plural, and a third for neuter plural, totaling six forms. It was thus deemed more reasonable to include this form in the consideration of overall complexity rather than rule it out the same way that German perfect tense was ruled out, for instance, in which *only* the auxiliary changes but not the participle. To maintain consistency with Italian and German, only the number of distinct forms of the participle (conditioned by the grammatical gender of the agreeing subject) was considered as a contributor to overall complexity, not the product of distinct auxiliary by distinct participle forms. Thus, there are four distinct forms of the participle.

Finally, the Italian simple future tense was also considered as adding to the overall paradigm complexity of Italian. This tense marks all six person-number contrasts, which, however, are expressed as cliticized (and reduced) forms of the verb *go*.

In addition to the structural descriptions of target languages just presented, the frequencies of inflected forms were also taken into account (see section *Corpus frequencies of inflected forms* below), to control for the possibility that forms inflected for certain feature combinations may be more frequent in some TLs than in others. Such differences may arise, for instance, due to Czech and Italian being PRO-drop languages. The data on German were obtained from the DeReKo corpus (*Das Deutsche*

*Referenzkorpus*) maintained by the Institute of German Language in Mannheim (Institut für Deutsche Sprache, 2017). Data on Italian were extracted from the general reference corpus of Italian, *Coris / Codis* (Corpus di Italiano Scritto—Rossini Favretti, Tamburini, & De Santis, 2002). Finally, the frequencies for Czech are reported based on the Czech National Corpus (Kren et al., 2015).

**Phonological realizations of inflectional morphemes.** Any comparisons centering on paradigm complexity are complicated by the fact that phonological realizations of morphemes differ as well. These differences in phonological realization can result in consonant clusters of different length and complexity, interacting with the different restrictions on permissible syllable structures in learners' L1s, as well as L1-independent markedness constraints on consonant clusters in the interlanguage. Since all TLs are suffixing (German and Italian strongly, Czech—weakly) (Dryer, 2013b), the combination of roots ending in consonants with inflections expressed as consonants can create clusters that can be disproportionately located in codas (unless the inflection is syllabic). This potential imbalance towards codas poses an additional burden because codas are both more constrained typologically than syllable onsets in the number and identity of the consonants they allow (e.g., Blevins, 1996; Goldsmith, 2011, p. 190; Selkirk, 1980) and pose a greater difficulty for L2 learners (Anderson, 1987; Eckman, 1986; Sato, 1984) than do onsets.

The properties of the target languages' syllable structures, as well as those of learners' first languages, are summarized in Table 1. Syllable complexity, according to the World Atlas of Languages (Madieson, 2013), is a broad classification that is nevertheless intuitive: at the simplest level, there are languages that only allow the (C)V

syllable type (with an optional consonant onset); moderately complex languages permit CCV and CVC syllables—with CCVC being the most complex combination allowed—but pose strict restrictions on what the second consonant in the onset can be; complex syllable structures are those which permit three or more consonants in the onset and/or do not restrict the nature of the second consonant as narrowly as in the moderate group, while allowing two or more consonants in the coda.

Even though not all languages of interest are represented in the Syllable Structure map of WALS, its classification is still included in the table as a starting point. The learner metadata provided by the Merlin corpus (Wiesniewski et al., 2013) do not distinguish between European and Brazilian Portuguese, nor do they specify the varieties of Arabic spoken by learners.

Table 1 *Syllable structure and permissible coda clusters in target and first languages*

Language	Syllable Complexity (WALS classification)	Allowed Codas
German	Complex <sup>1</sup>	Maximum onset is CCC (C1 has to be /s/ or /ʃ/; maximum coda—sonorant in nucleus plus CCC with CC in appendix outside syllable (coronal obstruents). <sup>15</sup>
Czech	Complex <sup>4</sup>	Both Moravian and Bohemian: CCCC maximum in onset, CCC in coda (rare in practice).
Italian	Moderately complex <sup>3</sup>	Codas with monophthong nuclei only: r l m n; fricatives or stops only as a result of germination.
Russian	Complex <sup>1</sup>	Maximum onset: CCCC; CCCC possible in coda if C1 is a liquid, CCC more common <sup>10</sup> . Fewer combinatorial restrictions <sup>9</sup> .
Polish	Complex <sup>1</sup>	Up to six consecutive consonants: onsets of maximum CCCC, codas of CCCCC. <sup>12</sup>
Slovak	Complex <sup>5</sup>	Maximum onset: CCCC; maximum coda: CCC (only 4 attested); 53 possible two-consonant clusters. <sup>5</sup> /l/ and /r/ can form nucleus (are syllabic). <sup>6</sup>

French	Complex <sup>1</sup>	Onset optional, maximum is CCC; coda of CCC possible but very restricted, CC more common. <sup>13</sup>
Spanish	Moderately complex <sup>1</sup>	CC onsets only. Coda optional: any single consonant except palatals; /s/ most frequent <sup>2</sup>
Portuguese	Moderately complex	Few syllable-final consonants allowed, restricted. European variety allows more clusters than Brazilian. <sup>8</sup>
Hungarian	Complex <sup>1</sup>	CCC onsets possible but rare, CC more common—both found in foreign words; CCC coda possible. <sup>16</sup>
Arabic*	Variable	Modern Standard Arabic: maximum onset is C; coda CC. <sup>14</sup> Highly dialect-specific: Moroccan allows complex clusters, particularly in onsets. <sup>14</sup> Egyptian: Complex <sup>1</sup>
Chinese	Mandarin, Cantonese: Moderately complex	Mandarin: in coda only glides, nasals, or /r/ as a result of affixation; no obstruents.
Turkish	Moderately complex <sup>1</sup>	Generally no CC onsets. CC codas restricted: sonorant + obstruent; voiceless fricative + oral plosive; /ks/ <sup>7</sup>

*Notes.* Sources: 1. Maddieson (2013). 2. Núñez-Cedeño (2007). 3. Hall (1944). 4. Šimáčková, Podlipský, Chládková (2012). 5. Gregová (2011). 6. Pouplier & Beňuš (2011). 7. Kornfilt (2013). 8. Parkinson (2009). 9. Davidson & Roon (2008). 10. Halle (1959). 11. Chew (2003). 12. Sadowska (2012). 13. Dell (1995). 14. Hamdi, Ghazali, & Barkat-Defradas (2005). 15. Grijzenhout & Joppen (1998).

Overwhelmingly, the first languages of learners sampled in the corpus belong at least to the moderately complex type in their syllable structures. It is not clear to what extent the differences in syllable complexity impacts written production. Considering that task and register effects on consonant cluster simplification tend to favor more deliberate, monitored speech (Dickerson, 1974; Gatbonton, 1975), it seems that writing an essay for a proficiency exam would exert similar kinds of pressure and discourage cluster simplification. Furthermore, neither consonant cluster simplification nor similar hypotheses, such as the prosodic transfer hypothesis (Goad, White, & Steele, 2003), would explain oversuppliance of inflection. Nor do they explain any asymmetries in accuracy for morphemes with the same phonological realization, such as English plural

vs. third person singular *-s*. Although limited, there is evidence that second language learners are more likely to simplify clusters formed by inflectional morphemes, rather than those occurring at the ends of monomorphemic words, in contrast to native speakers (Bayley, 1996; Saunders, 1987; Wolfram & Hatfield, 1984), which means that the simplification is sensitive to morphological phenomena and is not exhaustively explained by phonology.

**Corpus frequencies of inflected forms.** To rule out the possibility that some inflected forms are more frequent in some target language than in others, a corpus analysis was conducted. The availability of high-profile, representative reference corpora varied for the target languages, and written corpora were the most easily available. Therefore, the following analysis will primarily focus on written corpus data. In each case, the most authoritative corpus sources with the most relevant annotations were prioritized.

Table 2 *Corpus frequencies of inflected forms in Czech (spoken, written)*

Form	Written Frequency (ipm), ranking	Spoken Frequency (ipm), ranking
Present Tense		
1 Pers. Sg.	9,809.93, 3rd	28,334.69, 2nd
2 Pers. Sg.	1,558.75	10,935.59, 3rd
3 Pers. Sg.	33,192, 1st	38,158.94, 1st
1 Pers. Pl.	4,649.37	7,074.83
2 Pers. Pl.	2,682.46	1,879.84
3 Pers. Pl.	10,077.94, 2nd	8,155.04
Future Tense		
1 Pers. Sg.	247.97, 3rd	1,016.71, 2nd
2 Pers. Sg.	81.09	541.68



3 Pers. Sg.	1,547.35, 1st	2,472.14, 1st
1 Pers Pl.	202.95	594.49, 3rd
2 Pers. Pl.	144.93	494.55
3 Pers. Pl.	482.98, 2nd	560.07

---

*Note.* Frequencies are listed relativized, in items per million, calculated as raw frequency (number of search results) divided by corpus size and multiplied by 1,000,000.

For Czech, both spoken and written corpora were available. SYN2015 is a written representative corpus of 100 million tokens. It is compiled from traditional—as opposed to web-crawled—sources, such as fiction, non-fiction, and periodicals dating from 2010-2014. It is lemmatized, morphologically tagged, syntactically annotated, and published within the Czech National Corpus framework (Křen et al., 2016). ORAL is a reference corpus of informal spoken Czech, containing over 500 hours of conversations between friends and family recorded between 2002 and 2011, or over five million tokens (Kopřivová et al., 2017). It is lemmatized and annotated for the same morphosyntactic features as the SYN2015 written corpus.

**Present tense.** Both in written and spoken language use, the most frequent form used was the third-person singular (Table 2). The second-most frequent inflected form in the written corpus was third-person plural, followed by first-person singular. In the spoken corpus, the first-person singular was ranked second—higher than in the written corpus, whereas the second-person singular took third place—ranking higher than in written use. These results likely reflect the sampling of speech situations/genres in the

spoken corpus, which included spontaneous exchanges between two interlocutors and provided many contexts for the use of second-person singular forms ('informal' *you*).

**Future tense.** Third-person singular continued to be the most frequent form in both written and spoken Czech (Table 2), followed by third-person plural and first-person singular in written Czech and by first-person singular and first-person plural in spoken Czech. These rankings of the future-tense forms were exactly the same as in the present tense in the written corpus; they deviated only by one position in the spoken corpus, where the third-most frequently used form was now the first-person plural—and not the second-person singular, as in the present.

The present- and future-tense counts were added together for the purpose of representing them in the joint table comparing frequencies in all target languages (Table 5).

For Italian, the corpus that came the closest to desired annotation depth was PAISÀ, a web corpus of 250 million tokens, compiled from texts around 2010 (Lyding, Stemle, Borghetti, Brunello, Castagnolli, Dell'Orletta, Dittmann, Lenci, & Pirelli, 2014). It is richly annotated for parts of speech with morphosyntactic properties, as well as dependency relations. It is not balanced or representative, but due to the nature of the text genres included (blogs, Wikipedia entries) its register can be considered somewhat less formal than that of a written corpus (e.g., based on fiction and periodicals) and spoken conversation. The counts reported in Table 3 are combined for all tenses.

Table 3 *Written frequencies of inflected forms in a web corpus of Italian*

Form	Frequency (ipm), ranking
1 Pers. Sg.	1375.41, 3rd
2 Pers. Sg.	319.04
3 Pers. Sg.	33270.48, 1st
1 Pers. Pl.	804.47
2 Pers. Pl.	197.95
3 Pers. Pl.	9699.24, 2nd

*Note.* Frequencies are listed relativized, in items per million, and calculated as raw frequency (number of search results) divided by corpus size, multiplied by 1,000,000.

The DWDS corpus of German, created at the Berlin-Brandenburg Academy of Sciences, was used to look up the frequencies of inflected forms. To concentrate on the most recent usage patterns, we focused on the sub-corpus comprising texts from the 2000s, DWDS-Kernkorpus 21, a differentiated, but not balanced, collection of texts from newspapers, fiction, journalism, and scientific research literature (Geyken, 2007). The corpus contains 15 462 297 tokens from 12 186 documents and was automatically tagged for parts of speech but not morphological features.

Table 4 *Frequencies of German inflected forms in a written corpus*

Form	Frequency (ipm), ranking
1 Pers. Sg.	3446.19, 3rd
2 Pers. Sg.	758.23
3 Pers. Sg.	5432.18, 1st
1 Pers Pl.	932.78
2 Pers. Pl.	16.08
3 Pers. Pl.	5106.55, 2nd

*Note.* Frequencies are listed relativized, in items per million, calculated as raw frequency (number of search results) divided by corpus size, times 1,000,000.

The absence of morphosyntactic annotation necessitated additional strategies for obtaining valid data. The iterative process of refining the search criteria is detailed below for the first person singular; only its results are reported for the other person-number combinations, unless there were additional changes to the search criteria prompted by challenges particular to any given form.

As a starting point of the search, I combined syntax searching for parts of speech (finite thematic verbs, finite auxiliary verbs, and finite modal verbs) with regular expressions referring to letters corresponding to German verbal affixes (e.g., *-e*, *-st*) in word-final positions. This process yielded results that could be further improved: for example, this search returned false positives, such as third-person singular preterite and subjunctive, which are indistinguishable from first-person singular forms based on their surface attributes alone.

An inspection of the first 100 results from this search revealed that only 56 out of 100 tokens were, in fact, first-person singular forms; one form was a participle, and the others were third-person preterite and subjunctive forms falsely identified as first-person

singular based on their surface attributes. Based on the low accuracy of the search, the criteria were augmented to specify that the verb also had to co-occur within the personal pronoun (*ich*, 'I'). To determine the appropriate window within a sentence to specify, we counted the number of times within the same sample of 100 results that the subject and verb were *n* words apart. Even though there were extreme cases where the verb and its subject were 9, 10, or 13 words apart, in the vast majority of cases (71%) the subject and verb were directly adjacent (that is, one word apart), and in a further 17% of them the distance ranged between two and four words. A window of six words covered 94% of the sample and was deemed appropriate. We specified the final search criteria to look for occurrences of words ending in *-e* and tagged as finite thematic verbs within eight positions of the pronoun "ich" in either direction (left or right). The eight-position window was chosen to make the search more permissive and allow for contingencies such as the inclusion of an adverb or a preposition phrase between the subject and the verb.

This reduced the number of hits from 109,023 (with false positives) to 21,489. A sample of 100 hits revealed 14 incorrectly included forms, improving the accuracy of the search from 56% to 86%. These incorrectly returned cases involved duplicates that were produced when the same verb token was counted twice—first with its true subject and a second time with another instance of "ich" in a following clause, for example the sentence *I like to cook, but I prefer to eat out when I am tired* would produce the following hits: *I like*, *I prefer* and *I am*—matching the true purpose of our investigation, but also *like [to cook, but] I* and *prefer [to eat out when] I*. In German all such cases involve a comma, so the solution was to search for all such instances separately and then

subtract their count (3787) from the total of 21,489. This yielded a count of 17,702, or 1144.85 items per million.

In addition, I searched for forms of modal and auxiliary verbs that do not end in *e*, such as *kann*, *muss*, *soll*, *darf*, *weiß*, specifying the same criteria for co-occurrence with an overt subject pronoun as for finite thematic verbs. This was necessary only for the first person singular, because all the other person-number combinations in German are expressed through identical, non-zero inflectional endings on both thematic and modal verbs.

The same process was followed to obtain counts of second-person singular forms (ending in *-st* or *-ßt*). In this case, misclassification issues were due to the overlap of second- and third-person singular surface forms for stems ending in *-s*: for example, forms such as *erweist* or *schießt* (from *erweis-en* and *schießen*, respectively). However, this was helped by restricting the search to forms occurring in the vicinity of the second-person singular pronoun (“du”), within the same span as that described for first-person singular above. No separate search was necessary for forms of modal verbs, since their endings match those of thematic verbs.

For the third person singular, it was not possible to restrict potential subjects in the same way as for the first and the second person: subjects can be expressed both through pronouns and noun phrases, but in the absence of morphosyntactic annotations on the nouns the search cannot be restricted to singular nouns only. Thus, as an approximation, we searched only for tokens of finite verbs ending in *-t* occurring within the same window as a closed class of pronouns—personal pronouns in the nominative case (specified as an exhaustive list) and certain indefinite pronouns (e.g., ‘somebody’),

‘nobody’, ‘one’—*jemand, niemand, man*). In addition, the same search procedure was applied to return instances of modal verbs that do not end in *-t*: the same list of pronominal subjects was specified within the same span of words from the verb (eight). This method, admittedly, underestimates the frequency of third-person singular forms in written discourse. Since it is the relative prevalence of inflected form, not their absolute frequencies, that is at the heart of this analysis, this was deemed an appropriate compromise.

For the first person plural, which is homonymous with third- and second-person plural (formal), we specified as criteria the co-occurrence with the pronoun ‘we’, but also with the phrases ‘I and \_’, ‘\_ and I’. In the case of ‘I and \_’, we expanded the search region following it (preceding the verb) from eight to ten positions, to allow for two-word determiner-noun phrases, such as “I and *my sister*”, or “I and the neighbor”. The number of results with this type of subject was very low, 15, and was corrected manually to exclude wrong matches that captured fragments of adjacent clauses.

The search for forms of the second person plural was complicated by the fact that the second person plural pronoun is homophonous with the feminine possessive, *ihr* (‘her’), and the dative of the feminine pronoun *sie* (‘she’). This meant that applying the criteria used for the other forms—specifying the inflectional ending *-t* in the vicinity of a pronoun—returned many false positives, including strings where *ihr* was preceded by a preposition or modified a noun which was the true subject of a third-person singular verb. With third person singular verbs also ending in *-t*, the difference between the third person singular and second person plural is often in the presence or absence of root vowel processes: for example, compared to the citation form of the verb ‘to read’, *lesen*, the

third person singular form is marked both by the ending *-t* and by the vowel change from *e* to *ie*, *liest* ('[he/she/it] reads'), whereas the second person plural only has the *-t* ending (*ihr lest*). It was impossible to narrow down the criteria sufficiently through the search syntax alone, but for a handful of modal verbs that have this vowel pattern it was feasible to specify exclusion criteria and rule out their third person forms, specified individually. From the results produced by this process, two samples of 100 tokens each were examined—one with the search results for the SV word order and the other for the VS word order. Within each sample, the number of true second-person plural forms was tallied and divided by 100 to produce an accuracy estimate for this search method. The total number of hits returned by the search was then multiplied by this percentage to produce an adjusted estimate of the form's occurrence. The accuracy of this search for the SV word order was 11%, or 11 target forms out of 100 results produced; for the VS word order, the figure was 5%.

For the third person plural, the procedure was the same as for the third person singular but with the personal pronoun 'they' (*sie*) and the plural definite article (*die*) specified as desired neighbors within the same eight-word window (in the case of *sie*) or within nine words, in the case of *die*, to accommodate the following noun. Just as was the case for the third person singular, this method underestimates the true rate of occurrence of third person plural forms, because it does not account for subjects that are bare plural nouns (without a definite article) or subjects that are expressed as coordination constructions of two singular nouns (*der/das\_ und der/das\_*), except when one of the coordinated nouns is a feminine noun with a definite article, which is homonymous with the plural definite article.



Samples of 100 results were examined for accuracy for the SV and VS word-order searches. In the SV sample, 58 tokens were true matches. A further 33 tokens were not, in fact, third person plural forms but matched the search criteria due to appearing adjacent to the pronouns specified; these would all be ruled out by the corrective search procedure specified above, which searches for sequences containing a comma; out of these 33, only three would be wrongly excluded by this procedure. Thus, the percentage of correct matches would be 88%, or  $[58 + (33-3)]/100$ . The remaining nine incorrect search results matched the criteria superficially but were not true third person plural forms: for example, they involved verbs ending in *-en* appearing close to an accusative use of *sie* (feminine) or with a subject that included a plural noun followed by a coordination construction with *I*, leading to a first person plural interpretation. There were no additional restrictions that we could specify to rule out such cases.

In the VS sample, 33% of cases outright were ones included correctly. A further 50% were incorrectly returned forms captured by the permissiveness of the search, all of which would be ruled out by the procedure excluding sequences with commas. Only one token would have been falsely excluded based on this search. Finally, 16 tokens represented cases that would not be affected by excluding sequences with commas: mostly, they were first-person plural verbs followed by a direct object with the definite article *die* (accusative of feminine and plural). Combined, these results represented an 83% accuracy of the search results.

Separately, the same search was conducted for the form *sind* ('are'), since it would evade the search for forms ending in *-en*. The results were then added to the counts of forms ending in *-en*, producing the numbers reported in Table 4.

In all three languages, the first four positions in the frequency ranking of inflected forms was identical (Table 5), with the third person singular being the most frequent, followed by the third person plural and the first person singular. The least frequent form in German and Italian was the second person plural, whereas in Czech it was second to last, outranking the second person singular.

Table 5 *Comparison of rank orders of inflected forms in written German, Italian, and Czech*

Ranking of forms	Target Language		
	German	Italian	Czech
1, most frequent	3 <sup>rd</sup> person singular (she)	3 <sup>rd</sup> person singular (she)	3 <sup>rd</sup> person singular (she)
2	3 <sup>rd</sup> person plural (they)	3 <sup>rd</sup> person plural (they)	3 <sup>rd</sup> person plural (they)
3	1 <sup>st</sup> person singular (I)	1 <sup>st</sup> person singular (I)	1 <sup>st</sup> person singular (I)
4	1 <sup>st</sup> person plural (we)	1 <sup>st</sup> person plural (we)	1 <sup>st</sup> person plural (we)
5	2 <sup>nd</sup> person singular (you)	2 <sup>nd</sup> person singular (you)	2 <sup>nd</sup> person plural (you Pl.)
6, least frequent	2 <sup>nd</sup> person plural (you Pl.)	2 <sup>nd</sup> person plural (you Pl.)	2 <sup>nd</sup> person singular (you)

These data indicate that there are little to no material differences in the frequency of occurrence of the inflected verbal forms among the target languages. Any discrepancies in learner success on these forms cannot be explained by differences in input properties—at least, in its written usage.

Table 6 *Summary of key differences between the morphological systems of target languages*

Attribute	Target Language		
	German	Italian	Czech
Synthetic tenses	2	4	1
Analytic tenses	3, pers.-num. agreement on auxiliary	3, pers.-num. agreement on auxiliary, participle (gender/number)	2, pers.-num. agreement on auxiliary, participle (gender/number/animacy)
Gender	not expressed	participles	participles

Frequency of inflected forms	5 <sup>th</sup> : you-Sg.; 6 <sup>th</sup> : you-Pl.	5 <sup>th</sup> : you-Sg.; 6 <sup>th</sup> : you-Pl.	5 <sup>th</sup> : you-Pl; 6 <sup>th</sup> : you-Sg.
Verb classes	2, effect on participle endings	3 (thematic vowels), no effect on endings	4-6, effect on endings <sup>9</sup>

**Conclusions.** Integrating these observations on paradigm complexity across the multiple tenses, mood, and voice combinations (Table 6), one notes the general ranking from German at the least complex end (with minimal complications in the form of regular/irregular differences), to Italian (complex with additional complications in the form of regular and multiple irregulars), to Czech at the most complex pole—with the most numerous distinctions and with the highest number of perturbations caused by lexical class groupings and the alternations they condition. Even though there are a few nuances in operationalizing complexity, which result in slightly different counts, it is the relative complexity of the three TLs that matters for the present study.

On a morpheme-by-morpheme view of learning espoused by general cognitive accounts, accuracy on inflection should be highest (or achieved earliest) in German, followed by Italian and Czech, potentially differing in the relative ordering of Italian and Czech learners in the present versus the past tense. By contrast, if forms interact in development, and richness is beneficial (or at least not detrimental) to learnability, production accuracy would be positively related to the number of distinct morphological forms in the target language.

## Chapter 4: Methods—Corpus Study of Written Learner Productions

The present chapter will describe the methods employed in the study, including data sources, elicitation tasks employed, and the cleaning and coding of data.

Additionally, background information is provided about the distribution of learners' L1s among the target languages, with the goal of ruling out potential differences that could bias the results. The procedures involved in the actual analysis of the data are presented in detail in Chapter 5, considering the iterative nature of the decisions made at each step and their close connection to the results.

### 4.1 Data Source and Learner Backgrounds

Written production data by learners of German, Italian and Czech were obtained from the Merlin corpus (Wiesniewski et al., 2013), which spans the levels from A1 to C1 of the CEFR proficiency framework and contains essays written as part of foreign language proficiency exams in the respective target languages. The written essay responses were collected by the Merlin project and subsequently rerated by its staff utilizing the CEFR guidelines. The test-takers were speakers of a variety of L1s, and the implications of this diversity for the results of this study will be addressed in the following sub-section (Learner L1s). Despite this variation, L1 influence alone does not exhaustively explain grammatical difficulties in the second language, as was shown in Chapter 1.

The texts in German and Italian were the writing sections of proficiency exams developed and administered by the provider *telc*; the Czech texts came from proficiency tests administered by the test center at the Institute of Language and Preparatory Studies

at Charles University in Prague. Both providers are members in the Association of Language Testers in Europe and are audited by this organization (Texts and test institutions, n.d.).

Consistent with the CEFR approach to testing, the writing tasks were designed to be representative of everyday language use and to have a communicative impetus in the form of a prompt. For example, a task in the informal register included responding to a friend's invitation to visit and asking what kind of birthday present their child would prefer. To elicit language use in the formal register, for example, the learner would respond to a mock-up of a job ad and inquire about a few topics specified in the prompt. Performance elicited by such tasks can be expected to be monitored but also relatively spontaneous, due to the task's communicative relevance and lack of focus on accuracy alone.

Since the essays that provided the data had been collected in the language testing context, the intended proficiency levels of the tests did not always match the proficiency levels of learners as rated by test raters (Table 7). For example, a test-taker of a B1 proficiency exam in Czech may have been rated higher (e.g., B1+) or lower (e.g., A2) than the stated level. All references to learner proficiency in this paper denote proficiency as rated by raters, not the proficiency levels targeted in the tests.

Table 7 *Merlin corpus statistics: Number of texts re-rated at each CEFR level*

Target Language		CEFR Level							
		A1	A2	A2+	B1	B1+	B2	B2+	C1
German	Texts	57	199	107	217	115	219	73	42
	Sentences	280	1334	977	2322	1583	3274	1071	603

Czech	Texts	1	76	112	90	75	72	9	4
	Sentences	5	787	1728	1668	1116	1103	124	73
Italian	Texts	29	289	92	341	53	2	0	0
	Sentences	180	2475	901	4314	707	20	0	0

Because data were extremely sparse for Czech and Italian at the A1 level and Italian at the B2 level (as indicated by the low counts of learner texts and sentences), only data from levels A2 through B1+ were included in the study. This was done to maximize the comparability of the data and to explore the full range of variation associated with linguistic typology at each level.

Learner texts in the corpus came tokenized, lemmatized, and automatically tagged for parts of speech, morphological features, and syntactic dependencies. In addition, manual annotations were supplied by the Merlin project for “target hypothesis”—minimally corrected versions of language produced by the learners. For example, a learner’s sentence such as “Yesterday I \*walk home” would receive the target hypothesis annotation of “Yesterday I walked home”.

**Learner L1s.** Learners’ first languages varied among the target languages. To pinpoint the sources of this variation and to determine whether the L1s were distributed in ways that could jeopardize the analysis of learners’ accuracy, a chi-square analysis of independence was conducted. To this end, L1s were aggregated into seven categories (Figure 1), and their counts were cross-tabulated against the TLs. The groupings were meant to reflect any relatedness of the L1s or any commonalities in their morphological systems.

The first and, perhaps, least interesting group, was comprised of learner texts for which L1 information was either not reported (i.e., not collected) or reported as “Other”.

Secondly, a number of Romance languages were represented (Spanish, Italian, Portuguese, French) in the sample, particularly among learners of Italian and German. Counts of learners who were L1 speakers of Slavic languages were tabulated together and included Russian (more common among learners of Czech and German), Polish (more common among learners of German and Italian), Czech (only a handful of learners of German and Italian), and Slovak (one learner of Czech). Next, English and Chinese were grouped together owing to their relatively poor inflectional morphology. This decision was motivated by the low number of learners with Chinese as their L1 (10), which would have resulted in expected cell counts less than five for each TL if Chinese had been kept as a separate category. The grouping with English resulted in higher expected counts and allowed the use of the Chi-squared test of independence on the cross-tabulation of L1s. Also grouped together, despite being unrelated, were Turkish and Hungarian, on the grounds of both being agglutinating languages with rich morphological systems and neither being related to any of the TLs or the other L1s. Finally, Arabic formed one of the two single-language categories, owing to its unique morphological properties—discontinuous, interlocking roots and patterns; the other single-language category was German. On the one hand, its morphology was considered richer than English (or Chinese), making their grouping undesirable; on the other, it was not meaningfully related to any of the remaining L1s.

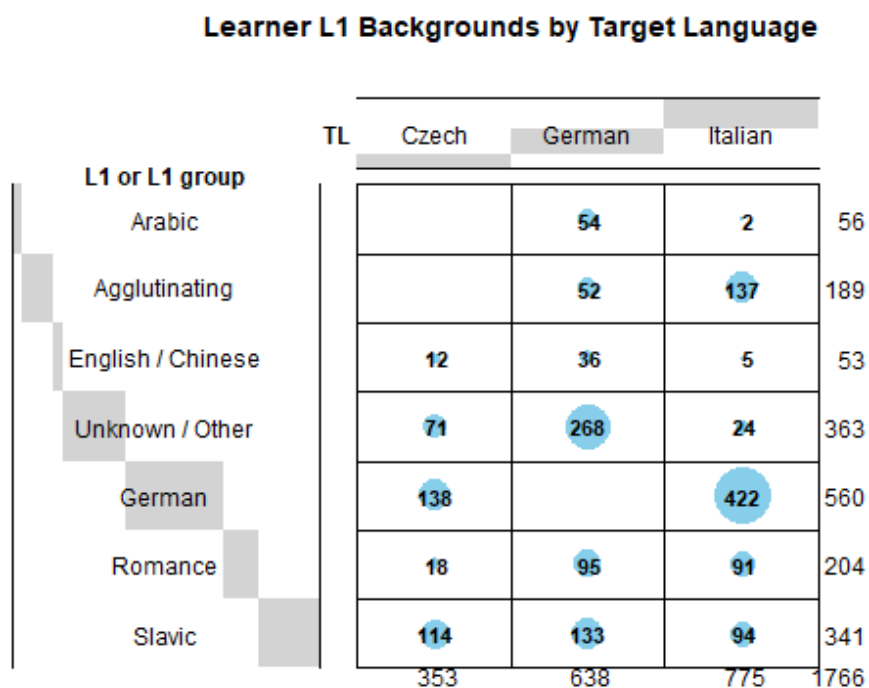


Figure 1. Learner L1 backgrounds by Target L2, aggregated across all proficiency levels.

The resulting contingency table of target languages and L1 languages and language groups was submitted to a chi-square test of independence, conducted in R software using the *chisq.test* function in its base *stats* package. The chi-square test revealed a significant association between TL and L1 group:  $\chi^2(12) = 858.3, p < .001$ . To examine the sources of this association, Pearson residuals were obtained for each cell and plotted in proportion to their magnitude (Figure 2, left panel) using the *corrplot* package (Wei & Simko, 2017). In addition, for each residual a percentage score of its influence on the total Chi square statistic was calculated, according to the formula: squared residual / chi square statistic \* 100%. The results of this procedure are represented in Figure 2 (right panel). Notably, German as the L1 was only a represented



among learners of Italian and Czech and was logically impossible as a cell value for German as the TL.

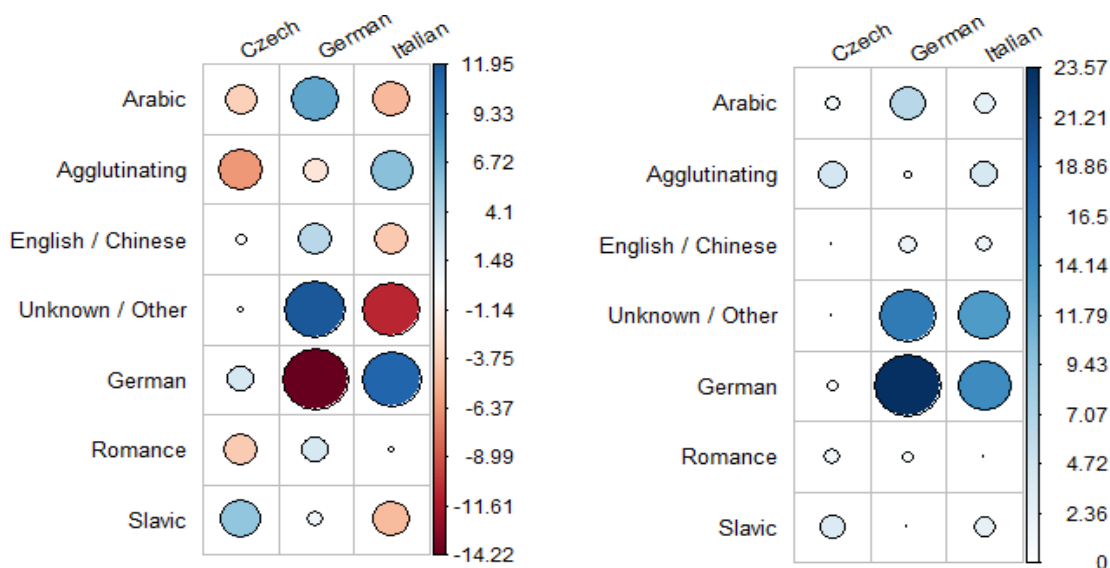


Figure 2. L1-TL contingency table Chi-square test residuals (left) and their % contribution to total statistic (right).

Italian had an overrepresentation of learners whose L1s were agglutinating and whose L1 was German, compared to what would be expected based on chance alone; it also had fewer than expected speakers of Arabic, English or Chinese, and Slavic languages as an L1, as well as those learners for whom L1 information was absent. German, by contrast, had an overrepresentation of these exact groups—Arabic, English / Chinese, and Unknown L1 backgrounds. Czech had a higher representation of learners with Slavic L1 backgrounds and, less so, German; underrepresented were learners whose L1s were agglutinating or Romance languages, as well as Arabic.

Next, an inspection of the percentage contribution of these differences to the total chi-square statistic revealed that the chi-square test was driven primarily by the absence of German L1 among learners of German, as well as the high proportion of learners of German for whom L1 information was not reported. The patterns in the distributions of

L1s among the TLs were examined in closer detail because of their potential ability to invalidate any findings of differences among TLs. First, vastly different proportions of L1s closely related to the TL could mean that the TLs with higher concentrations of such learners would be at an advantage overall or with respect to particular error types. Second, equally problematic would have been a situation where the TLs would have vastly differing proportions of learners with backgrounds in some of the “richer” L1s, as opposed to some of the inflectionally “poorer” L1s.

With respect to the first possibility, none of the L1s in the sample were closely related to German. By contrast, among learners of Czech, having an L1 background in one of the Slavic languages was more common than expected. Finally, among learners of Italian L1 knowledge of a Romance language was represented at a frequency consistent with the expected count. Romance languages were, in fact, more common among learners of German than Italian. The contributions of the differences in numbers of Romance and Slavic languages as L1s to the total chi-square score were not sizeable (Figure 2, right panel). One potential effect of these differences could be an advantage for the speakers of the related L1s on those aspects of inflected forms that relate to their phonological makeup, due to the presence of shared roots or inflectional endings (between the L1 and the TL). On such an account, therefore, learners of German would commit more phonological errors, owing to their not speaking L1s closely related to it; by contrast, learners of Italian and Czech—each with a group of speakers of related L1s—would be more familiar with some of the roots or inflectional endings and would confuse their phonological makeup less frequently.

With respect to the second possibility (involving differences in proportions of L1 speakers of inflectionally “richer” or “poorer” languages among the TLs), the only group of L1s that could be considered noticeably “poorer” than the rest was English and Chinese. Overall, there were few learners with these L1s, and they were concentrated in the groups of learners of Czech and German, representing 3% (12 out of 353) and 5.6% (36 out of 638) of these groups, respectively. The residual values showed that English and Chinese were overrepresented among learners of German and underrepresented among learners of Italian (Figure 2, left), but both contributed only minimally to the total chi-square statistic (Figure 2, right).

Thus, a pattern of findings where German learners show a higher rate of phonological errors than learners of Italian and Czech could be indicative of being biased by the distribution of L1s in the sample. In this case, the absence of learners with L1s closely related to German would mean that they couldn’t capitalize on any phonological similarities—in contrast to those learners of Czech and Italian who were speakers of Slavic and Romance L1s, respectively. Similarly, a pattern of errors where German learners demonstrate a higher prevalence of bare, uninflected forms than learners of Italian could have as its source the learners with English and Chinese as their L1s in the German corpus than in the Italian corpus. This difference, nevertheless, would not be expected to affect the use of infinitival forms instead of finite ones, owing to the fact that in all TLs infinitival forms are morphologically marked and, therefore, supplying this marking would not reflect L1 transfer.

## 4.2 Procedure

The Merlin corpus was queried through its online interface (ANNIS—Krause & Zeldes, 2016) to extract errors of types that were broadly relevant to the present research. The frequencies for each error type were obtained for each target language at each CEFR level separately. The following search terms were used: EA\_type = “G\_Agr” for errors annotated in the corpus as agreement errors; G\_Inflex\_Inexist\_type=“verb” for errors annotated as uses of inexistent inflected forms of verbs; G\_Verb\_compl\_type = “ch” for errors involving the wrong form of a complement or auxiliary. In addition, a search was conducted for the “sentence” annotation to return the number of sentences in each sub-corpus, with the goal of using the sentence counts to relativize the observed number of errors for each TL-proficiency combination when plotting the data.

Second, the errors were coded by speakers of the target languages according to the scheme described in the following section (*Error categories and their significance*). The procedure for it was piloted on the German sub-corpora A1 and A2, which finalized the categories of errors to be used for all TLs. For Czech and Italian data, coding was carried out by target-language experts, who were either native speakers of these languages or had academic training in them which was coupled with experience residing in a community where the target language was the primary language. This step also entailed data cleaning, such as removal of duplicates or instances that reflected a lack of lexical or orthographic knowledge rather than the knowledge of inflectional morphology of the TL.

Finally, the data were modeled using the Poisson regression approach. Data were analyzed broadly following the steps outlined in Hilbe (2016) for modeling count data,

proceeding from assessing the dispersion of the data to choosing a final best-fitting Poisson regression model, examining its coefficients, interactions and pairwise contrasts. The resulting models were then validated on out-of-sample data. The specifics of these steps and the conclusions drawn from them will be presented in Chapter 5.

### 4.3 Error Categories and Their Significance

This section links the source data annotations to the categories of analysis developed for this study, based on theoretically driven interpretations of the processes that could have generated the errors. The error categories are presented in Table 8 with examples from each TL illustrating them. The present section focuses on a conceptual description of the coding scheme and the rationale behind treating certain groups of errors separately. The practical application of the coding scheme and the decisions made in less clear-cut situations are the subject of the following section (*Cleaning and Coding of Data*).

Three error annotations in the Merlin corpus were relevant to this research and served as the foundation for the tailored classification described in the remainder of this section: agreement errors, inexistent inflection errors, and verb complement errors. The “agreement” annotation in the corpus captured substitutions of one inflected form for another, regardless of how the two forms related to each other. By contrast, in the present study the directionality of these substitutions was deemed meaningful. Substituting an infinitive or bare (uninflected) form where a finite, inflected one is required, can speak to the strength of the “Tense” feature in the underlying grammar, according to some accounts proposed for child acquisition of English as an L1 (e.g., Yang, 2002). However, substitutions between two finite inflected forms would not warrant such an interpretation,

since both forms (the target form and the one substituted for it) could be considered reflecting a learner grammar that expresses Tense. One could speculate that substitutions of this kind may be traceable to lexical selection—on the assumption that inflected forms are stored unanalyzed in the mental lexicon—or to the application of a wrong compositional rule. To honor the different conclusions that can be drawn from these errors, tokens annotated as “agreement” errors in the corpus were separated into substitution, infinitive, and bare form errors in the study.

Going in the opposite direction, learners also sometimes use finite forms in contexts where non-finite ones are required: for example, on verb complements (with a modal auxiliary, or in analytical tense-aspect and mood forms). For comparison, in English this would amount to saying *I want to \*reads that book* or *She has \*reads that book*. In the Merlin corpus, such instances were annotated as “Verb complement” errors, and this tag applied without differentiation with respect to the type of complement required and also included incorrect auxiliaries—again, without taking into account the direction of the error or what specifically made it wrong. For instance, uses of the infinitive in lieu of a past participle, or selection of wrong auxiliary were tagged as “Verb complement” errors. In this study, however, overuse of inflection formed a separate category and covered the use of finite forms or participles in contexts requiring an infinitive. By contrast, uses of the infinitive in lieu of a participle were captured by the “infinitive” category described above, whereas auxiliary selection errors were dropped from the analysis altogether.

The third error annotation (“inexistent inflection”) in the corpus encompassed any departures from target-like orthography, as well as any inaccuracies in the segmental

composition of the form, whether they concerned the root, any suffixes, or endings alike, as long as they resulted in an inexistent form. This annotation was applied to errors vastly differing in their proximity to the target form. On the one hand, some forms were mostly on the right track and could be traced to a misapplication of rules and patterns present in the TL grammar. For example, the past participle form \**gekaufen* (used in place of *gekauft*) does not exist in German, yet it follows the familiar inflectional template of past participle formation for irregular verbs (*ge-* plus *-en*) and is easily interpretable both with respect to the lexical verb and its intended grammatical meaning. On the other hand, other errors deviated further from the target and may have stemmed from a failure to recall the phonological composition of the root or ending. For instance, this could occur as a result of phonological processes such as consonant cluster simplification or the vagueness of lexical representations themselves. In this study, therefore, the “Inexistent inflection” tag was separated into errors of verb class (misapplication of inflectional patterns associated with a verb class), root—involving wrong application or non-application of a root alternation process, bare—dropped inflectional ending, and phonological—errors with incorrectly specified segments of the root or ending (unless associated with another verb class or existing root alternation process). Admittedly, phonological errors partially overlap with orthographic errors. More detailed reasoning behind separating the two is presented in the following section (*Cleaning and Coding of Data*).

Table 8 *Error types adopted in the coding scheme, with examples from each TL*

German	Italian	Czech
<b>Substitution</b>		
A2, English: Ich *gratuiert max zum er geburtstag hat.	A2, Hungarian: *Vogliamo ( <i>vogliamo</i> ) mangiare cibo culiniare.	A2, German: Kdy *bude ( <i>budeš</i> ) na nadraží?
B1, French: Wir *werde das bei dir feiern.		
<b>Infinitive</b>		
A2, Arabic: und hute ich *treffen mit michael bei eine Kafftref in stadt um zwei Uhr mittag.	A2, German: Da lunedì a venerdì ho *frequentare ( <i>frequentato</i> ) un corso italiano.	A2, German: Bych *těšit ( <i>těšila se</i> ) mě velmi.
B1, L1—Other: Wenn man Deutsch lernet, dann *bekommen man einfache ein Job.		
<b>Overuse</b>		
A2, English: Am nächste Wochenende möchte ich bie dir *komme.	A2, German: Vuole *apprendere ( <i>apprendere</i> ) un po' d' italiano.	A2, L1—Other: V kolik hodin bude to *začíná ( <i>začne</i> )?
B1, Spanish: Wurde ich Freizeit *habe?	B1, Hungarian: Vorrei ti *aiuto ( <i>aiutarti</i> ).	
<b>Root</b>		
A2, L1 not reported: Kannst du bei mir *hiffen ( <i>helfen</i> ).	A2, German: E cosa *potriamo ( <i>possiamo</i> ) fare la sera?	A2, Russian: [...] *světí ( <i>svítí</i> ) slůnce.
B1, L1 not reported: Deshalb möchte ich mich *bewarben ( <i>bewerben</i> ).		
B1, L1 not reported: Vielleicht, du *könnst ( <i>kannst</i> ) zum meinen Geburtstag kommen.		
<b>Class</b>		
A2, L1 not reported: [...] dann hat im krankenhaus *geblibt ( <i>geblieben</i> ).	A2, Turkish: I thailandese *cuoca molto bene.	A2, German: *Museš ( <i>Musíš mě</i> ) meho navstítit dopoledne.
B1, L1 not reported: Ich habe dein Brief *bekommet ( <i>bekommen</i> ).	B1, Hungarian: Mi piace *lavorere ( <i>lavorare</i> ) in gruppo con altri.	
<b>Bare</b>		
A2, Spanish: Wenn du deine ausbildung Fertig *gemach ( <i>gemacht</i> ) Haste.	B1, Hungarian: ma io non sono potuto *giocar ( <i>giocare</i> ) a tennis.	A2, German: Co to *stuj ( <i>stoj</i> )?



B1, Spanish: Ich \*arbeit (*arbeite*) gans  
toll im Team [..]

#### Phonological

A2, Arabic: Ich \*Früch (*freue*) mich  
Für anne [..]

A2, German: O che cosa \*possimo  
(*possiamo*) fare la sera?

A2, English: Hraja (Hrají) si na  
pláži a vypadají

B1, Russian: Wieleicht \*möchtes  
(*möchtest*) du die Reise machen?

*Note.* Errors relevant to the present study are marked with asterisks. Corrected forms of the verbs are listed in parentheses.

## 4.4 Cleaning and Coding of Data

After error tokens were extracted from the Merlin corpus (see *Procedure* for the search syntax used), data were inspected and cleaned. For German data, this was done by the researcher; Czech and Italian data were cleaned by the target language experts who also conducted the coding. First, some cases were excluded either because they were duplicates or because the departure from the target form was deemed not to involve inflection or a process broadly associated with inflection. Table 9 summarizes types of cases that were excluded and provides illustrative examples, whereas the remainder of this section will present each in turn, in narrative form. Finally, some common decision points that occurred during coding will be presented for each error category. As a rule of thumb, both data exclusion and data coding decisions were motivated by the following principles: minimizing interpretation of learner intent; focusing on inflectional endings before anything else; excluding errors of selection and semantics (e.g., appropriate lexical elements; auxiliaries; tense and mood).

**Exclusions.** The types of departures from target-like use presented in this section were not considered errors, as long as they were *the only* departures from target-likeness

for a given token. When they co-occurred with another error that was relevant to this study, they were ignored (but noted), and the verb token was classified on the basis of the other, relevant, error. For example, an instance classified as a misspelling (German: *Ich \*arbite*) was not counted among *any* error category if its agreement was correct.

However, an instance of *Ich \*arbiten* was classified as an infinitive use error with an additional misspelling notation, due to the verb's being apparently non-finite (or plural), whereas *Ich \*arbitet* was considered a substitution error (based on the inflectional ending associated with the wrong person-number combination).

The category "Ambiguous" captured those tokens for which it was impossible to determine with certainty what the learner was trying to convey. For example, in German some tokens were ambiguous with respect to their part of speech, or multiple verbs were strung together without it being obvious how they related to each other. Rather than engaging in mind-reading, the data coders stayed as firmly as possible within the realm of what was uttered and, when conflicting interpretations presented themselves, preferred to exclude the token altogether.

Next, not considered errors were misspellings or typos ("Typo", Table 9). This group included instances such as single consonants where double were required, and vice versa; missing or superfluous diacritics (including on vowels); missing or superfluous umlaut (in German), as long as the presence or absence of the umlaut was not a legitimate root alternation process. For example, the token of *bisüiche* (instead of *besuche*) was considered correct because in the indicative mood there is no alternation involving *u* and *ü* among thematic verbs (Table 9). By contrast, the modal verb *müssen* does have a vowel alternation (with *u*, in some person-number combinations), and, therefore, instances of

*müss* were classified as errors (superfluous root process). Tokens involving the letters *e* and *i*, and their combinations (*ei*, *ie*) and transpositions, were treated as correct, as long as the identity of the word could still be determined and unless the change of the letters was part of an existing alternation. For example, *\*arbite* (instead of *arbeite*) was treated as a correct token, because there is no *ei-i* root process in the present tense of the indicative. By contrast, *\*hilfen* instead of *helfen* was treated as an error classified as a superfluous root alternation (because *e/i* is a legitimate root alternation in the indicative present tense and in the lexical family sharing the root). Similarly, the *a/ä* alternation is represented among thematic but not modal verbs, meaning that an instance such as *\*fähre* (instead of *fahre*) was classified as an error of superfluous root change (because some person-number combinations of this verb include *ä*), whereas a form such as *\*känn* (instead of *kann*), by contrast, was classified as a non-existent root, since no person-number combinations of the verb exist with the vowel *ä*.

The next group of exclusions involved stylistic and pragmatic errors. In Czech, uses of Common (vernacular) Czech forms were excluded and not considered errors, taking into account the fact that learners' histories and experiences in TL speech communities were unknown. Pragmatic errors concerned the use of 'formal' and 'informal' forms of address—for example, uses of verb forms agreeing with *du* ('informal' *you*) when a formal one agreeing with *Sie* ('formal' *you*) was called for, and vice versa. Stylistic errors included any non-targetlike uses that appeared to be situationally determined or involved the knowledge of the "educated" usage prescribed for the TL. For example, in German there is the so-called "introductory *es*" which can function as a false subject (as opposed to referring to a singular entity in the third person):

in these cases, the verb agrees with the “true” subject that follows, not the “es” pronoun—*Es kamen-Pl. 18 Besucher* (not: *Es \*kam-Sg. 18 Besucher*). However, if learners used “es” as the target of agreement, such tokens were excluded. Other special cases of agreement in German involve complex predicates with nominals, which can create confusing situations depending on the grammatical number of both the subject and the predicate nominal. These cases were also considered to be well outside the scope of “mainstream” agreement in L2 learning and were, therefore, excluded.

Exclusions related to selection included choices of wrong lexical items but also the choice of a wrong auxiliary, which in Italian and German, in part, depends on the knowledge of detailed semantic properties of verbs. For example, forms requiring a choice between *haben* and *sein* in German are the perfect and pluperfect tenses, and between *essere* and *avere* in Italian—*passato prossimo* or *trapassato prossimo*. In a similar vein, uniquely to Italian and Czech, cases involving wrong gender marking on the participle (in analytic tenses) were excluded. This was because accurate agreement production in these cases requires lexical knowledge of the grammatical gender of the nouns they agree with as much as a command of inflection proper. Non-targetlike uses of verb tense and mood forms were also not considered errors, as long as they were correctly marked for agreement: appropriately choosing tense and mood was deemed a problem that is more semantic in origin.

Table 9 *Examples of data excluded during data cleaning*

Reason for exclusion	Examples
Ambiguous	G, A2, Turkish: wann *ist deine Kinder *sein ( <i>not clear</i> ).
Typo	G, A2, Arabic: Ich *bisüche ( <i>besuche</i> ) Wir nechst Woche. It, A2, German: Mentre I miei genitori *preferiscono ( <i>preferiscono</i> ) che lei aspetterà ancora.[..].
Auxiliary selection	G, A2, L1 Unknown: [...] dann *hat ( <i>ist</i> ) im krankenhaus *geblibt. It, A2, German: Non *ha ( <i>è</i> ) cambiato tanto solo qualcosa.
Gender	It, A2, German: perche mio capo ha *fissata ( <i>fissato</i> ) una riunione [...] Cz, A2, L1 Unknown: Jaká hezká *byla ( <i>bylo</i> ) počasí!
Analytical form	G, A2, Russian: Sie *ist ( <i>no verb needed</i> ) *wonnen in Haus. Cz, A2+, German: Ve kolik hodin *budeš přijet ( <i>přijedeš</i> ) na nádraží?
Superfluous inflected element	G, A2+, L1 Unknown: [Es ist] sehr gut zu horen, dass du die Prüfung bestanden *gemacht ( <i>no participle needed, missing auxiliary</i> ).
Missing or wrong elements (reflexive, conditional particles, verb prefixes)	It, A2, Polish: Io *voreii *invitare ( <i>invitarti</i> ) *ti a cena. Cz, A2+, German: *Bychom ( <i>bych</i> ) se těšila, [...] Cz, A2, German: *Platim ( <i>Zaplatím</i> ) se když sedjeme.
Tense	It, A2, German: [...] non ci *abbiamo *visto ( <i>vediamo</i> ) Io sono ritornata. Cz, A2+, German: *Budu ( <i>byla</i> ) bych celý vikendu.
Lexical	G, A2, L1 Unknown: Ich *nahme Fahrkarte Bus gekauft. It, A2, French: Come lo *savete *stiamo ( <i>siamo</i> ) partiti in Italia [...]
Stylistic	G, A2+, L1 Unknown: Meine ganze Familie *leben ( <i>lebt, singular</i> ) in Serbien. G, A2, Hungarian: [...] wann es *ist ( <i>sind</i> ) zwei Söhne, [...] Cz, A2+, L1 Unknown: [...] jestli máte parkoviště (protože *jedem / <i>jedeme/ autem</i> ).

*Note.* Errors are denoted by asterisks (\*); target forms are listed in parentheses directly following the errors.

Also excluded were cases when superfluous elements were added, such as unnecessary auxiliaries or semi-lexical verbs (e.g., ‘give’, ‘do’) that appeared to have an intended aspectual meaning. These cases were considered “correct” as long as agreement was marked appropriately on one of the elements, and the lexical verb in the infinitive was not counted among infinitive errors. This was because with the asynchronous medium of written essays it is impossible to draw the line between any self-corrections and complex (but non-targetlike) predicates.

Unique to Czech was the exclusion of non-targetlike uses that involved wrong, missing, or superfluous prefixes (that can carry aspectual meaning in Czech), as well as wrong, missing, or superfluous reflexive and conditional particles. In German and Italian, close analogs of this error type were omissions of the reflexive pronoun. In addition, in Italian there were also uses of the full infinitive followed by a personal pronoun instead of an infinitive followed by a *postclitic* reflexive pronoun: *\*invitare ti* instead of *invitarti*. These instances were excluded from the analysis.

**Coding choices.** The category of substitution errors was restricted to substitutions within the same mood and tense. For instance, in German the imperative of many verbs in the second person singular corresponds to their stem (e.g., *gib!* *geh!*), but when such forms were encountered in syntactic contexts requiring a form of the indicative, they were classified as bare (*[er] gib-~~t~~*, *[es] geh-~~t~~*) as long as an overt subject was present—not as substitutions of a wrong mood form.

For infinitive errors, the coding of Italian and Czech data was mostly straightforward, as long as an overt subject or discursive information made it unambiguously clear what agreement was required. In German, the infinitive and 1<sup>st</sup> and

3<sup>rd</sup> person plural forms are homonymous. When faced with wrong agreement involving forms ending in *-en*, I considered it parsimonious to interpret them as infinitives and not plural forms, unless discursive evidence suggested otherwise. If some instances of forms ending in *-en* were, in fact, plurals and not infinitives, one would expect the counts of infinitive errors in German to be inflated at the expense of substitution errors, which would be artificially depressed, compared to Italian and Czech. However, as Chapter 5 shows, substitution error counts did not significantly differ among the TLs, whereas the other error types did—strongly suggesting that those differences did not occur as a result of any underestimation of substitution errors.

The category of bare forms included instances of fully omitted inflectional endings. Any suffixes or thematic vowels that preceded the inflectional ending (with reference to the target form) could be either retained or also omitted for the token to be considered in this category. For example, in Italian this category included such cases as those where the final *-e* of an infinitive was truncated but the stem was otherwise preserved: *mangiar* instead of *mangiare*. *Mangia*, however, was classified either as a substitution or overuse of inflection (if used instead of an infinitive).

Inflection overuse errors encompassed the uses of finite forms where non-finite ones were required, as, for example, participles in analytical tense or voice forms, or infinitives in analytical tenses or in complex predicates (with modal or semi-lexical verbs). In addition, the uses of participles instead of infinitives were also categorized as inflection overuse. This was done because in all three TLs the formation of the past participle necessitates considerable morphological transformations of an extent similar to inflecting a finite form for person-number features. Thus, excluding the uses of these

complex forms where an infinitival one suffices would have underestimated learners' facility with inflection, despite its overextension.

Root errors included cases with either superfluous or missing root transformations. The boundaries around this category were drawn broadly: besides root vowel alternations (e.g., vowel change *e-ie* in German or *o-uo* in Italian), it included errors on irregular verbs and verbs whose paradigms involve suppletion, even though those are not productive morphophonological processes of the contemporary TL grammar. For example, German irregular verbs *haben* and *sein* both involve considerable departures from the segmental makeup of the infinitive in some forms: e.g., *sein* (infinitive)—*bin* (1 pers. sg.), *ist* (3 pers. sg.), *sind* (1, 3 pers. pl.), *war* (1 pers. sg. preterite), *gewesen* (past participle). Whether root transformations were missing or superfluous was decided relative to the infinitive form: for example, the use of *\*habst* (instead of *hast*) was considered an instance of a missing root transformation, since the learner retained the segment of present in the infinitive (*b*) too faithfully. Most decision points in this category concerned differentiating these errors from typos and misspellings, as was discussed in the previous section (Exclusions).

Verb class errors were instances of verbs being inflected according to an inappropriate inflectional template associated with a different verb class. In Italian, verb class membership is discernible in the infinitive and is signaled by thematic vowels following the root. In German, there are no overt cues to verb class membership on the infinitive: instead, verb class is reflected in the choice of affixes on the past participle and sometimes correlates with root vowel changes. In Czech, verb class is identifiable from either the infinitive or the third-person plural of the non-past tense by the suffix or the



final segments of the root (Janda, p. 34). In addition, for some irregular verbs that have suppletive forms in their paradigms (e.g., German *sein*), any attempts by learners to “regularize” them were classified as verb class errors.

In German, the challenge with appropriately defining the category of verb class errors was in the overlap between the morphology of some participles and the morphology of infinitives. While most German verbs form the participle with both a prefix and a suffix, the prefix (*ge-*) is the most reliable way to identify past participles, because the suffixes (*-t* for regular verbs and *-en* for irregular) are homophonous with person-number endings. However, some verbs do not take on the prefix: verbs that already have a prefix only add the *-en* or *-t* suffix, depending on their membership in the irregular or regular class (*bekomm-en*, *begegn-et*); regular verbs ending in the productive suffix *-ieren* do not take the prefix either and take on the suffix (*-t*). This means that for verbs that do not take on the prefix, errors in the suffix are ambiguous between trying to derive the past participle using the wrong verb class template and between infinitive, 3<sup>rd</sup> person singular, and 1/3 person plural. In these cases, the deciding factor was the presence or absence of an auxiliary verb that could set an unambiguous expectation for the use of a particular tense. Without an auxiliary, it was parsimonious to interpret such instances as present-tense forms with wrong person-number features, as opposed to positing a missing auxiliary accompanied by an incorrectly inflected participle.

Finally, phonological errors included instances of non-existent forms, sometimes comprised of recognizable elements of the TL system and other times of ones that are absent in it altogether (e.g., *antwortes*, *möchtes*, *möcht*, *fruch*). The boundaries around this category were defined primarily in terms of exclusion, aiming to house errors not

already captured by the remaining error categories. In particular, non-existent errors were distinguished from root errors in that root errors involved existing root alternation processes applied incorrectly and not merely *any* inaccuracy in the root—which would fall under the non-existent category.

It is impossible to determine with any certainty what process would generate such instances and what exact lexical representations may underlie them. Chapter 6 explores one such possibility—namely, that, at least for German data, some errors can be accounted for by phonological processes such as consonant cluster simplification. However, for the purposes of the main analysis (Chapter 5), this category made it possible to isolate such cases with minimal speculation about their origins and status in the learners' grammars.

## Chapter 5: Results—Cross-linguistic Differences in Inflection Error Frequency

The present chapter will describe the data analytic process, starting with the structure of the data and the variables used in modeling it. Second, it will lay out the procedures of model specification and selection, including alternative models tested and their fit, and the testing of key assumptions involved in modeling count data. Third, the results of the best-fitting model will be presented, including the follow-up analyses conducted on significant interaction terms. Where applicable, developmental differences will be pointed out. However, any references to proficiency-related changes are to be taken with caution, since the data are cross-sectional. Finally, the chapter will explore the power of the model it proposes to predict previously unseen data, rather than achieving the best fit to the data on whose basis it was specified to begin with.

As laid out in Chapter 4, errors of different types may have different origins and may potentially reflect different knowledge states and properties of learners' grammars. In addition, the relative prevalence of some error types over others may call into question the assumptions of theories of learning. In Yang's proposal, which was formulated based on the presence versus absence of inflectional marking in English, it was forms with and without overt inflectional marking that were informative for the learner (with respect to the value of the Tense parameter in the target grammar). Accordingly, the presence of verbs with and without overtly marked inflection was taken as evidence of what a learner's parameter value might be at any given time. Considering that the TLs in the present study allow far fewer cases of verbs with no overt inflectional marking (e.g., the German imperative is a rare exception), infinitive forms (used incorrectly in place of

finite forms) were interpreted in a similar manner to bare forms—as suggesting a weak underlying knowledge of tense and inflection.

The process of coding learner production data was completed by target language (TL) experts (described in Chapter 4) and yielded counts of errors of seven types at each proficiency level within each TL. Thus, the data had the structure represented in Table 10. Despite data from learners at levels A1 and B2 being available in the Merlin corpus, only data from proficiency levels A2 through B1+ were used because they were available for all three TLs. At the A1 level, the corpus contained only one text by a learner of Czech, and level B2 contained only two texts by learners of Italian.

Translated into the language of statistical models, the research questions posed in Chapter 4 can be expressed as assessing the significance of variables predicting the number of verb inflection errors and the comparative fit of models including or omitting them. The predictors are: target language, proficiency level (cross-sectional), and error type; their attributes are summarized in Table 11.

Table 10 *Structure of the Data*

Target Language	CEFR	Error Type	Count	N of texts in corpus
German	A2	Substitution	N	N
German	A2	Infinitive		
German	A2	Overuse		
German	A2	Root		
German	A2	Class		
German	A2	Bare		
German	A2	Phonological		
German	A2+...	...		
German	...B1+	...		
Italian	A2...	...		
Italian	...B1+	...		

Czech	A2...	...
Czech	...B1+	...

---

*Note.* Levels of variables have been omitted to convey the overall structure of the dataset. The actual numbers of levels for each variable are listed in Table 2.

The analyses reported in the present chapter were conducted using R statistical software (R Core Team, 2019), specifically the *glm* function, which is part of the base package. It estimates generalized linear models by selecting one level of factor variables as the baseline and providing regression coefficients for the remaining levels of a factor against which the other regression coefficients are to be judged. To facilitate the interpretation of the regression coefficients, the levels to serve as the baseline were selected as follows. For the target language variable, German was selected as the baseline: its system of verb inflectional endings has the fewest distinct endings, and thus its paradigm can be considered the poorest among the languages studied. Therefore, regression coefficients for Italian and Czech can be interpreted as reflecting the relative benefits or penalties to accuracy due to their higher inflectional richness. Second, for the CEFR variable, the A2 proficiency level was retained as the baseline, meaning that the coefficients for levels A2+ or B1, for example, can be interpreted as multipliers adjusting the predicted number of errors compared to A2. Finally, for the error type variable substitution was selected as the baseline category because substitutions were the most frequent error category among all target languages.

Table 11 *Variables used in the analysis*

Variable name	Role	Type	Number of Levels (baseline listed first)
Frequency	Response	Count	N/A
Number of texts	Offset	Log of count	N/A
Target language	Predictor	Factor	3: German, Czech, Italian
CEFR level	Predictor	Factor	4: A2, A2+, B1, B1+
Error type	Predictor	Factor	6: substitution, infinitive, overuse, root, class, bare, other phonological

## 5.1 Regression Model Specification and Model Selection

The first goal of modeling was to assess the suitability of a Poisson model to the data and determine whether any adjustments would be needed to the final models. One of the basic assumptions in applying Poisson modeling to count data is that the variance is equal to the mean, implying that as the mean number of events increases, so does the variance. However, count data are frequently overdispersed, where their variance exceeds the mean, rendering the application of the Poisson distribution invalid due to distorted standard errors of model estimates (and thus, a distorted picture of their statistical significance). Values of the dispersion statistic above 1 are considered overdispersed and warranting other modeling approaches that include adjustments to the standard errors, such as the application of quasi-Poisson models, the use of robust standard errors, or the estimation of a negative binomial model. However, sometimes overdispersion is only apparent and signals that potentially relevant predictors or interactions have not been included in the model, which is why it is necessary to test a variety of model specifications.

Dispersion was assessed using the function  $P\_disp$  in the COUNT package for R (Hilbe, 2016) and calculated as the sum of squared Pearson Chi2 residuals divided by residual degrees of freedom (Hilbe, 2014, p. 79).

The modeling process started with fitting a series of Poisson regression models using the *glm* function in R statistical software (R Core Team, 2019). All models discussed below were estimated with an *offset*, which is used in count data modeling to adjust for differences in the sizes of the corpora. In the present case, since the corpora included different numbers of learner texts, the error counts would be higher the larger the size of the corpus. The offset was a log-transformed count of texts in each corpus: for example, for German A2 the offset would equal the logarithm of the number of texts at that level ( $\log(199)$ ).

Table 12 summarizes the specifications of the models tested during this step, along with their respective fit and dispersion metrics. The simplest model with all three predictors of interest was one with only their linear sum:  $\text{Count} \sim \text{target language} + \text{CEFR} + \text{error type}$ . Despite all predictors being significant, the model had poor fit, as indicated by the dispersion statistic of over 4. Progressively adding more interaction terms reduced the apparent overdispersion to 0.97 (thus achieving the desirable value of 1), in the case of the model with three pairwise interactions. A model with a three-way interaction between target language, proficiency, and error type was estimated but ultimately abandoned due to overfitting: it had no residual degrees of freedom and did not fit the data significantly better than model 4 (Table 12)— $\chi^2 = 43.65, p = .18$ .

Table 12 *Summary of Poisson Model Dispersion and Model Fit Values*

Model	AIC	Chi <sup>2</sup> , Dispersion
Target language + CEFR + error type	707.33	335.14, 4.65
Target language * error type	474.48	97.43, 1.62
Target language * error type + CEFR * error type	475.60	63.17, 1.50
Target language * error type + CEFR * error type + target language * CEFR	456.29	34.85, 0.97

Following the estimation of model 4, its residuals were plotted against the predicted values of error frequency (Figure 3). They displayed no discernible pattern in the distribution across the range of predicted values.

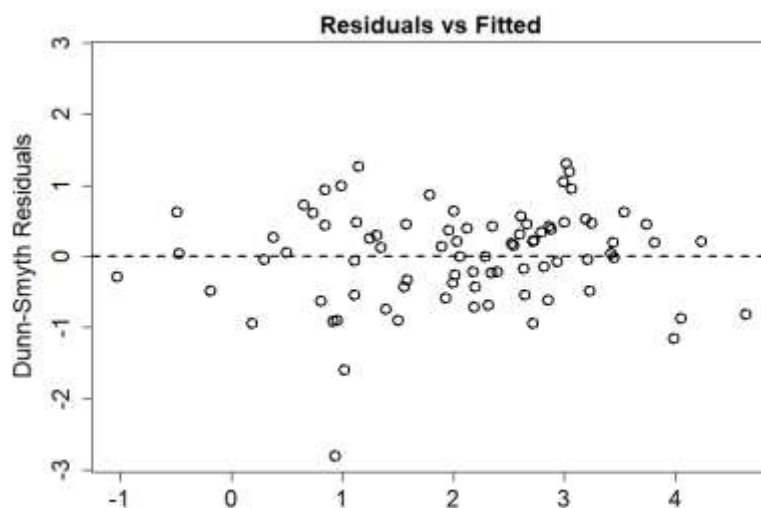


Figure 3. Model residuals plotted against fitted values for the model predicting error counts from: target language, CEFR, error type, and interactions between—TL\*error type; TL\*CEFR; CEFR\*error type.



The results of this step indicated that there was no significant overdispersion and, therefore, no need for the use of negative binomial modeling. Thus, model 4 served as the basis for a deeper exploration of the roles of the predictors and for pinpointing the sources of differences among the target languages, proficiency levels, and error types.

## 5.2 Regression Model Results

The results of the final, best-fitting model are presented in Table 13. Model coefficients for factor variables in R are provided for each of the factor levels separately and are interpreted as multiplier differences between the respective factor levels and the baseline level. In this case, the baseline level was the combination of Substitution (for error type), German (for target language), and A2 for proficiency. The regression coefficient (exponentiated) for “Target language: Czech”, therefore, would represent the ratio difference in the occurrence of substitution errors between German and Czech at the A2 proficiency level.

The regression analysis indicates that all three predictor variables—target language, error type, and proficiency, had levels that either significantly differed from the baseline or were part of interactions that did so. Summarizing the results from these disparate factor levels into a bird’s-eye picture (Table 14) using the *joint\_tests* function in the *emmeans* package (Lenth, 2019), which tests and compiles all the interaction contrasts in a model, we note that target language and error type are significant across the range of their individual levels, as well as their interaction with each other. Proficiency, however, only interacts with target language (and not error type) and is not itself significant. A visual inspection of the plot of the proficiency-error type interaction (averaging across target languages) agrees with this conclusion: all CEFR levels exhibit

the same pattern—highest rate of substitution errors, middle range comprised of infinitive, overuse, root, and class errors, lowest point representing bare form errors, and somewhat higher rate on phonological errors.

Table 13 *Regression model results predicting error rates in German, Italian, and Czech*

Parameter	Exponentiated Estimate	<i>p</i> value
<b>Intercept</b>	0.29	< .001
TL: Italian	0.83	.20
TL: Czech	0.95	.80
CEFR: A2+	1.11	.54
<b>CEFR: B1</b>	0.67	.01
<b>CEFR: B1+</b>	0.50	.002
Error: infinitive	0.78	.16
<b>Error: overuse</b>	0.54	.001
<b>Error: root</b>	0.23	< .001
<b>Error: class</b>	0.05	< .001
<b>Error: bare</b>	0.14	< .001
<b>Error: phonological</b>	0.31	< .001
<b>Interaction terms: target language by error type</b>		
<b>It * infinitive</b>	0.06	< .001
<b>Cz * infinitive</b>	0.61	.03
<b>It * overuse</b>	0.38	< .001
<b>Cz * overuse</b>	0.22	< .001
<b>It * root</b>	0.55	.02
<b>Cz * root</b>	0.33	.004
<b>It * class</b>	2.75	.003
<b>Cz * class</b>	6.54	< .001
<b>It * bare</b>	0.12	< .001
<b>Cz * bare</b>	0.48	.04
It * phonological	0.63	.07
<b>Cz * phonological</b>	2.16	.002
<b>Interaction terms: Error type by proficiency</b>		
Infinitive * A2+	0.93	.78
Infinitive * B1	0.76	.27
Infinitive * B1+	1.34	.34
Overuse * A2+	0.92	.78

Overuse * B1	0.88	.60
Overuse * B1+	1.40	.32
Root * A2+	0.87	.72
Root * B1	1.53	.14
Root * B1+	1.89	.11
Class * A2+	1.61	.13
<b>Class * B1</b>	<i>1.96</i>	<i>.02</i>
Class * B1+	1.47	.36
Bare * A2+	1.76	.19
<b>Bare * B1</b>	<i>2.67</i>	<i>.01</i>
Bare * B1+	1.96	.20
Phonological * A2+	0.71	.22
Phonological * B1	0.76	.29
Phonological * B1+	1.33	.37
<b>Interaction terms: target language by proficiency</b>		
It * A2+	0.81	.39
<b>Cz * A2+</b>	<i>1.57</i>	<i>.03</i>
<b>It * B1</b>	<i>1.90</i>	<i>&lt; .001</i>
Cz * B1	1.29	.25
It * B1+	1.64	.08
Cz * B1+	1.34	.25

---

Next, the interactions were examined in detail. The overall significance of interaction terms was tested using the *drop1* function, which sequentially compares the fit of the full model to the fit of models after their terms are dropped one at a time. This procedure uses the AIC and the likelihood ratio test (LRT) as the indicators of goodness of fit. The likelihood ratio test is the difference in model deviance between the full model and the model without the predictor. Significant differences in these indicators between two models reflect the overall importance of a variable, whereas non-significant tests of the difference suggest that a variable may be omitted to achieve a parsimonious account. The results of this analysis are summarized in Table 15.

Table 14 *Significance of model contrasts integrated by variable and interaction*

Model term	Df	F ratio	<i>p</i>
Target language (TL)	2	25.27	<.0001
Proficiency (CEFR)	3	2.21	0.08
Error type	6	72.14	<.0001
TL x CEFR	6	4.97	<.0001
TL x Error type	12	14.41	<.0001
CEFR x Error type	1	1.52	0.07

The interaction between target language and error type, when dropped from the model, significantly increased model deviance by 234.58 ( $p < .0001$ )—suggesting that different types of errors were characteristic for the target languages (across proficiency levels). Similarly, dropping the interaction between target language and proficiency level increased model deviance by 31.31 ( $p < .0001$ )—indicating that the effects of proficiency were not uniform for all target languages across the range of error types. However, omitting the proficiency-by-error type interaction resulted in a 28.5 increase in model deviance, which was not statistically significant ( $p = .05$ ). This suggests that for reasons of parsimony one may assume that different types of errors behaved similarly along the proficiency trajectory.

Table 15 *Contributions of interaction terms to model fit (assessed by single term deletions)*

Term deleted	Df	Deviance	AIC	Likelihood Ratio Test	<i>p</i>
None (full model)		43.56	456.29		
TL * error type	12	278.14	666.87	234.58	< .001
CEFR * error type	18	72.07	448.80	28.51	.05
TL * CEFR	6	74.87	475.60	31.31	< .001

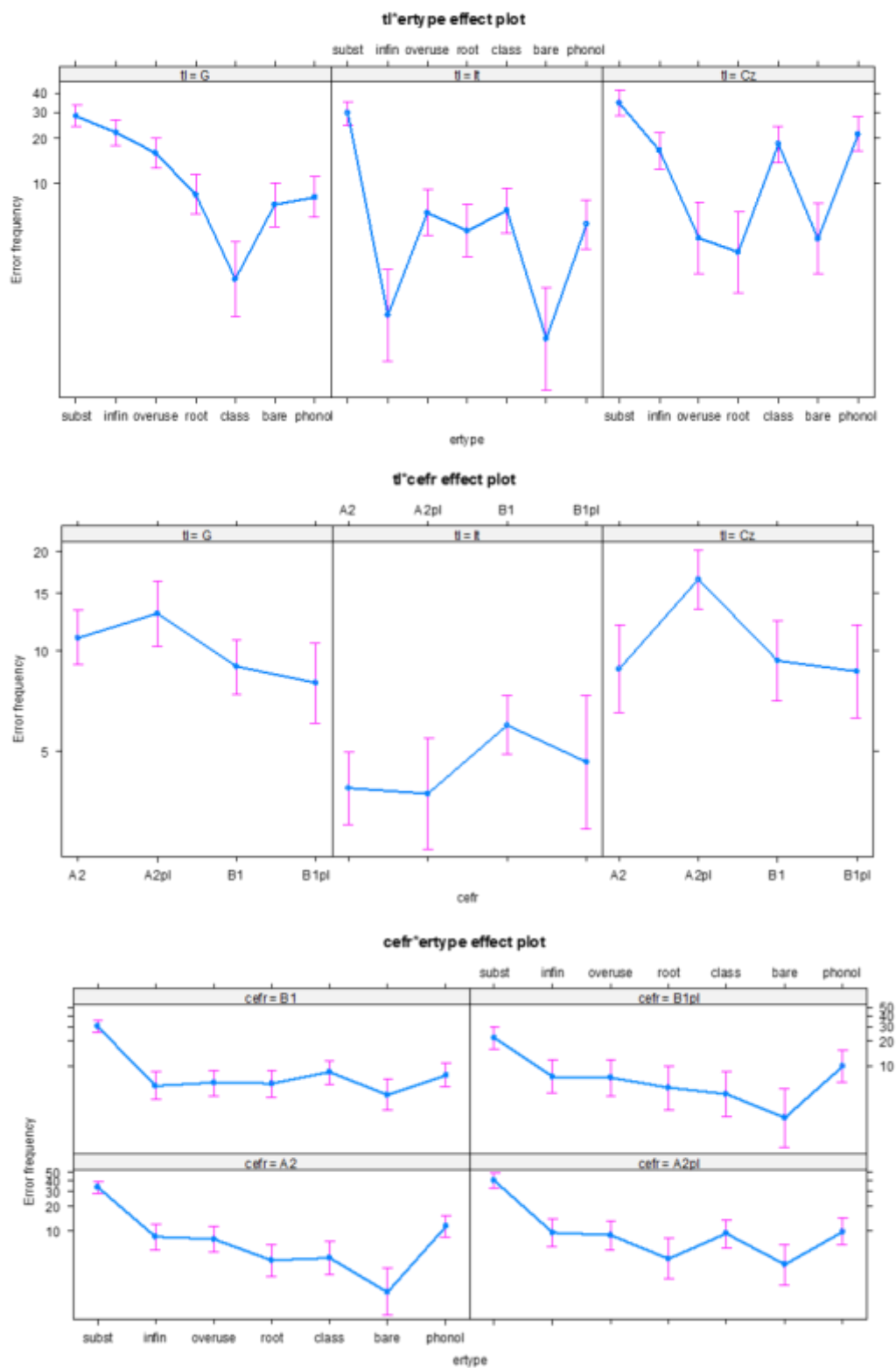


Figure 4. Plots of aggregated model effects. Top panel: target language by error type interaction; middle panel: target language by proficiency interaction; bottom panel: proficiency by error type interaction.

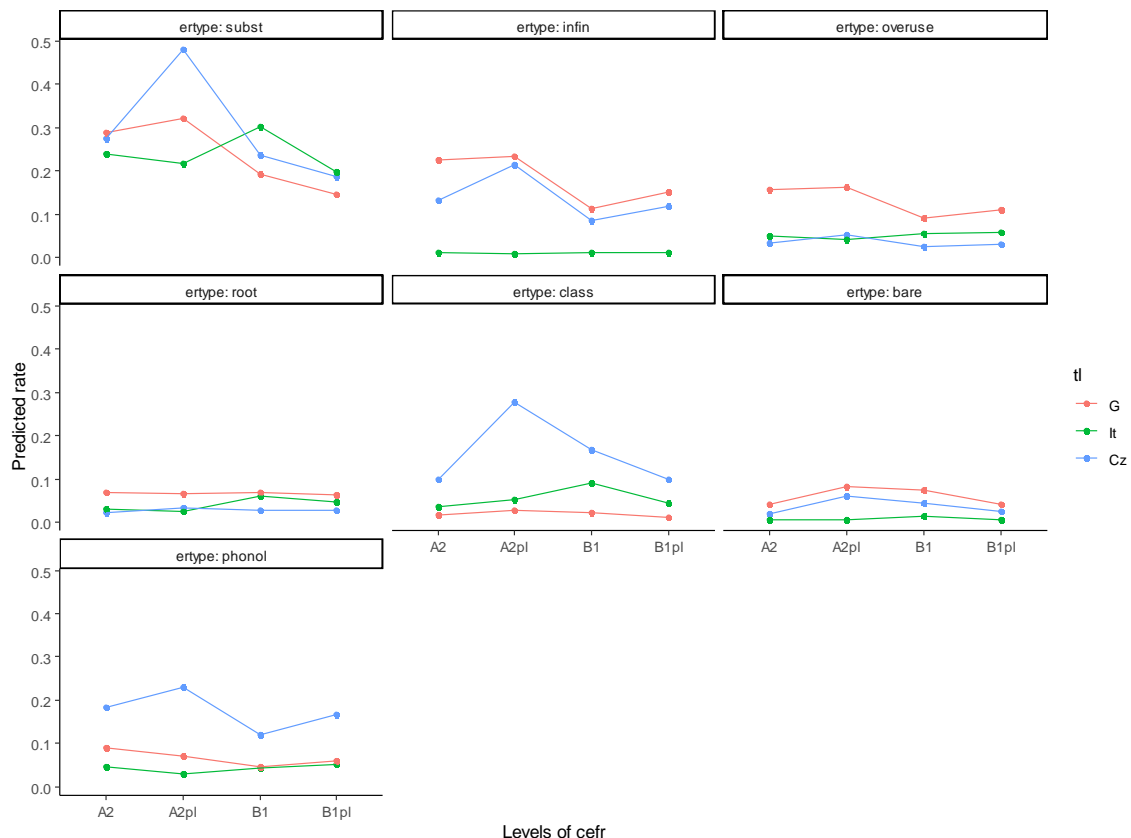


Figure 5. Model-predicted error rates by type for German, Italian, and Czech at proficiency levels A2 through B1+.

To tease apart which specific differences between the combinations of factor levels are driving the significance of the predictors, follow-up analyses were conducted using the *emmeans* package in R (Lenth, 2019) designed for pairwise comparisons. For models with a link function (in this case, Poisson with a log link), the package conducts comparisons on the log scale and back-transforms the results to the scale of the response variable. Bonferroni adjustments to p-values were made to account for multiple comparisons.

As a first step, these pairwise comparisons will be reported within each CEFR level, with a focus on the cross-linguistic differences in the relative frequency of error

types, providing insight into how the TLs compare at each proficiency level. Second, error prevalence will be compared within each TL (for each proficiency level separately) to capture patterns in learners' over- and underproduction of errors of different types. Finally, holding error type constant, any differences across proficiency levels will be highlighted, despite proficiency being a less prominent variable according to the regression model.

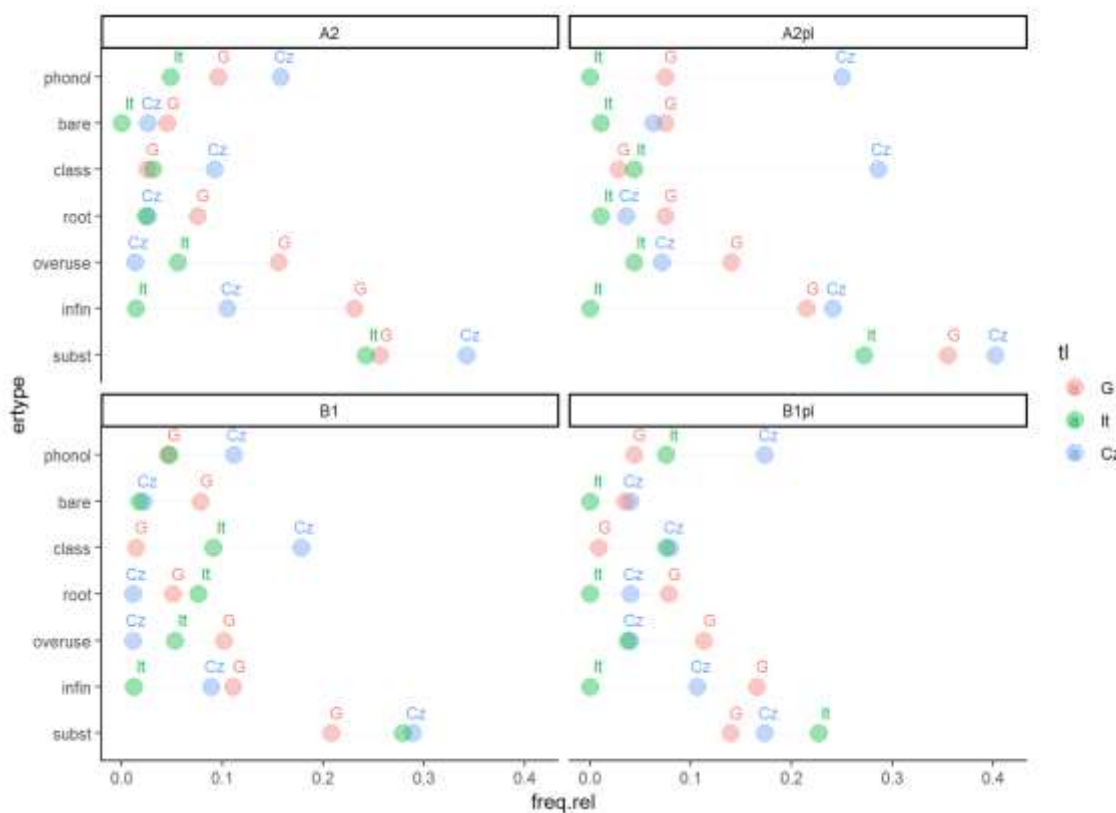


Figure 6. Relativized (per number of texts) observed frequencies of error types in German, Italian, and Czech across proficiency levels A2-B1+.

**Pairwise comparisons of error type frequencies in TLs.** This section summarizes in narrative form the results of pairwise comparisons presented in Table 16 and illustrated in Figures 4, 5, and 6.

At the A2 proficiency level, the target languages did not significantly differ in the prevalence of substitution errors (Table 16; Figure 4, top panel; Figure 5, “erctype: subst”;

Figure 6), indicating that using inflected forms interchangeably may be a common learner strategy not affected by typology. This lack of differences among TLs may cast doubt upon the classification of some errors in German: since the German infinitive is homophonous with the present-tense plural forms (first and third persons ending in *-en*), only forms that did *not* end in *-en* and are unambiguously finite were classified as substitution errors. This leaves open the possibility that the count of substitution errors in German might thus be underestimated—if some of the forms ending in *-en* were, in fact, finite forms, the counts of infinitive errors would have been artificially boosted at the expense of substitution errors. In that case, however, German would be expected to differ in the prevalence of both error types from both Italian and Czech: on substitution errors, German should have shown lower rates, and on infinitive errors—higher rates; all the while, no differences between Czech and Italian need to be predicted. In reality, the prevalence of infinitive errors was significantly higher in German than in Italian but not in Czech; Italian, in turn, had a lower rate of infinitival forms than Czech. This pattern cannot be explained by the classification of German homophonous forms alone: if infinitive errors had, in fact, received an artificial “boost” from this classification decision, then their count would have to be higher than that of both Italian and Czech. On errors of overuse of inflected forms, German had higher rates than both Italian and Czech, but the latter two did not significantly differ from each other. On errors involving the over- or non-application of root transformation processes, there were no significant differences among the TLs. Errors involving the application of a wrong verb class pattern were less prevalent in both German and Italian than in Czech but not significantly different between German and Italian. When it came to the use of bare forms with no



overt inflectional marking, German had a significantly higher prevalence than Italian, but not Czech, and Czech and Italian, in turn, did not differ from each other. Finally, with respect to phonological errors German did not significantly differ from either Italian or Czech, but Italian had a lower rate compared to Czech.

Table 16 *Summary of pairwise comparisons of error rates between target languages by type*

Error type	Proficiency Level (CEFR)							
	A2		A2+		B1		B1+	
Substitution	G = It = Cz		G ~ It, G ~ Cz	Cz > It	G = It = Cz		G = It = Cz	
Infinitive	G > It, G ~ Cz	Cz > It	G > It, G ~ Cz	Cz > It	G > It, G ~ Cz	Cz > It	G > It, G ~ Cz	Cz > It
Overuse	G > It, G > Cz	Cz ~ It	G > It, G ~ Cz	Cz ~ It	G ~ It, G > Cz	Cz ~ It	G ~ It, G ~ Cz	Cz ~ It
Root	G = It = Cz		G = It = Cz		G = It = Cz		G = It = Cz	
Class	G ~ It, G < Cz	Cz > It	G ~ It, G < Cz	Cz > It	G < It, G < Cz	Cz ~ It	G ~ It, G < Cz	Cz ~ It
Bare	G > It, G ~ Cz	Cz ~ It	G > It, G ~ Cz	Cz > It	G > It, G ~ Cz	Cz ~ It	G > It, G ~ Cz	Cz ~ It
Phonological	G ~ It, G ~ Cz	Cz > It	G ~ It, G < Cz	Cz > It	G ~ It, G < Cz	Cz > It	G ~ It, G < Cz	Cz ~ It

*Note.* The > symbol marks significant pairwise comparisons where the language

on the left had a higher incidence of an error type than the language on the right. The ~ symbol denotes pairwise contrasts that were not statistically significant. The equal sign (=) means that all possible pairwise combinations were not significant: G = It = Cz means that German did not significantly differ from Italian nor from Czech, and Italian and Czech did not significantly differ either.

At the A2+ level, substitution error rates did not differ significantly between German and either Italian or Czech. However, between the latter two it was Italian that had the lower rate. Errors of infinitive use were significantly higher in German than Italian but did not significantly differ between German and Czech, whereas Italian showed a significantly lower rate of them than Czech. Overuse of inflection was significantly more prevalent in German than Italian but not significantly higher in German than in Czech, despite trending in that direction. In turn, Italian and Czech did not significantly differ in the prevalence of inflection overuse. Continuing the trend

observed at the A2 level, at the A2+ level there were no significant differences in the rates of root transformation errors. Errors related to verb class were, once again, not significantly different between German and Italian but were lower in frequency in German than in Czech. In a comparison between Italian and Czech, Czech had the higher rate. The prevalence of bare forms without inflectional marking was significantly higher in German than in Italian but not Czech. Of the latter two, Czech had the higher rate. Errors in the phonological composition of the root or ending did not significantly differ in rate between German in Italian, but did so between German and Czech, with German having the lower rate. In turn, the rate in Czech was significantly higher than that in Italian.

At the B1 proficiency level, substitution errors were, once again, equally likely in all three TLs. Repeating the pattern observed at the lower proficiency levels, infinitive use errors were more prevalent in German than in Italian but did not significantly differ between German and Czech, whereas Czech had a higher rate than Italian. With respect to overuse of inflection, German did not significantly differ from Italian but showed a higher rate than Czech, which, in turn, was not different from Italian. Root errors did not differ in prevalence among the TLs, extending the trend observed since the A2 level. Errors of verb class were less prevalent in German than in Czech, consistent with the patterns at the lower levels. In a contrast to the A2+ level, however, German also had a lower rate of verb class errors than Italian (at A2+ the two did not significantly differ), and Italian did not significantly differ from Czech. The use of bare forms was significantly higher in German than in Italian but not Czech—mirroring the patterns at levels A2 and A2+. Between Czech and Italian, the rate of bare forms was comparable

with no statistically significant difference. Finally, with respect to phonological errors, German and Italian did not significantly differ from each other, but each had a lower incidence than Czech.

At the B1+ CEFR level, rates of substitution errors did not differ among the target languages, following the robust pattern that continued from the A2 level. Infinitive errors were significantly more prevalent in German than in Italian but did not differ measurably between German and Czech. However, Italian showed a lower rate of these errors than Czech. On overuse of inflected forms, German and Italian did not differ significantly, but German and Czech did, with the rate being higher in German, while Czech and Italian showed no significant difference. With respect to root transformation errors, the target languages did not differ. Errors of verb class were more prevalent in Czech than German but did not significantly differ between German and Italian. In turn, Italian and Czech did not differ from each other either. Use of bare uninflected forms was higher in German than Italian but not different between German and Italian, and nor was it different between Italian and Czech. Finally, phonological errors occurred at similar rates in German and Italian but were more prevalent in German compared to Czech, which had a higher incidence of them than Italian.

*Summary.* Across proficiency levels, the most stable pattern emerging from these pairwise contrasts is one where German shows no considerable advantage in accuracy over Italian and Czech on error types related to the acquisition of rules governing the use of finite forms, the unacceptability of bare uninflected forms, and appropriately constraining the use of finite forms. While Italian outperforms German on these types, Czech fares at least as well as German. Compared to Italian, the disadvantage of German

on bare form use and use of infinitives continues throughout all proficiency levels (A2 to B1+), whereas in the realm of inflection overuse the gap eventually closes at the B1 level, and German does not significantly differ from Italian anymore. The only area where German appears to have an advantage is errors of verb class and phonological errors. On verb class errors, German has lower predicted rates than Czech across all proficiency levels and lower predicted rates than Italian at the B1 level. On phonological errors, German outperforms Czech (as evidenced by *lower* predicted error rates) starting at the A2+ level and continuing through to B1+.

The contrast between German and Italian is informative with respect to the role of paradigm richness in learning: Italian has more distinct inflections and more verb classes and demonstrates that a rich paradigm to learn is not necessarily detrimental to accuracy; it can be beneficial for learning when not to use infinitive or bare forms and not to overuse finite forms when non-finite forms are required. The comparisons between German and Czech, on the one hand, and Czech and Italian, on the other, add nuance to this generalization: if paradigm richness comes additionally obscured by morphophonological alternations overlaid on top of it, the richness itself may be less effective in nudging the learner to developing appropriately constrained inflectional morphology. This is evident in the higher incidence in Czech of errors broadly associated with phonology, including the application of inflectional templates associated with the wrong verb class and the use of inexistent phonemic strings (both in the root and ending).

The clustering of error types that behave similarly according to TL typology is itself thought-provoking. On the one hand, one observes the close association of phonological and verb class errors, which pattern together (lower in German and Italian,

higher in Czech; Figures 8, 9). Their occurrence in TLs at different rates is an instance where the learning difficulty is in proportion to what the morphological system of the TL provides (Figure 8): a system that includes more verb classes and phonological processes inevitably presents more opportunities for learners to get them wrong. Hence, German has fewer and Czech more. Such proportionality would be consistent with the type of learning presupposed by general-cognitive accounts—essentially, serial accumulation of knowledge about individual lexical items. The need for such accumulation comes from the low salience of individual elements, which makes it impossible to learn any patterns from strokes of insight.

On the other hand, infinitive, bare form, and—to some extent—overuse of inflection errors form their own cluster (Figure 9). These errors of supplying inflection do not increase in lockstep with the complexity of the target system (Figure 8): the incidence of infinitive and bare form errors in Czech on par with German defies proportionality. When choosing an inflected form appropriate in a particular context, learners of Czech are faced with more options presented by their TL's morphological system than are learners of German, which lowers the probability of supplying a correct one. This lack of proportion between error rates and the complexity of the system to be learned is consistent with accounts of grammatical development that postulate that the disparate inflections that learners encounter are not just learned at their face value but also strengthen the underlying syntactic features. Thereby experiencing one inflected form ultimately benefits *all* inflected forms.

However, this account does not explain the differences between Czech and Italian, which are consistently directional and favor learners of Italian. Throughout the

entire proficiency span examined, learners of Czech show a higher incidence of infinitive use errors than learners of Italian, suggesting that, for morphological systems of equivalent complexity, there is a disadvantage brought on by phonological complexity, which nevertheless fails to outweigh the advantage of either of the “richer” systems over a “poorer” one (German).

Moreover, whatever the modulating effects of morphophonology may be, their extent does not appear to affect all domains: errors of bare form use, for example, did not differ between Italian and Czech in a significant way, with the exception of the A2+ proficiency level. This suggests that the presence of distinct inflected forms in the input, even ones that are less transparent or consistent (due to alternations), may be enough to facilitate the presence of *any* inflections in learners’ productions. Whether these inflections will ultimately be ones whose phonological forms are misstated, ones that mark the infinitive, or ones that are associated with a different verb class—is language-specific.

**Comparisons of error type prevalence within target languages.** This section seeks to pose a slightly different question to the last. Rather than asking whether, at a given proficiency level, learners of one target language are more likely to make errors of type X, it is interesting to explore the “mixture” of error types in each target language, characterizing the distribution of all error types relative to each other and then comparing these overall configurations among the languages. This analysis has the capacity to uncover learners’ favored and dispreferred ways of handling the uncertainty over choosing appropriate inflected forms in their language. The rank orders of error types within each target language at each proficiency level are summarized in Table 17. Since

the rank orders are remarkably preserved across the span of proficiency levels, data were also averaged across CEFR for each language to facilitate comparisons and to make the overall patterns stand out more (presented in Figures 7, 8 and Table 18). This averaging is additionally warranted by the general absence (with a couple of exceptions) of significant proficiency-related coefficients and interaction terms in the model (Table 13; also see subsequent section for a detailed account of proficiency-related changes).

Exploring Figure 7, one notices that the row corresponding to substitution errors looks remarkably similar for all three TLs: substitution errors were significantly more likely than all other error types, with the exception of infinitive errors in German, which were not significantly different in frequency from substitution errors. This points to the tendency of learners of all three TLs to use inflected forms interchangeably. Other than substitution errors, German was characterized by the prevalence of infinitive errors and the overuse of inflected forms—each of which outnumbered four out of the remaining six error types: root errors, verb class errors, use of bare uninflected forms, and phonological errors. The lowest-ranked error type in German was verb class (significantly lower than any of the six remaining error types). Root, bare, and phonological errors formed the middle of the pack and were all equally prevalent when compared to each other, all outnumbered verb class errors, and all were less prevalent than substitution, infinitive, and inflection overuse errors.

In Czech, substitution errors outnumbered all others and were followed in the “ranking” by infinitive errors, errors of verb class, and phonological errors. While the presence of infinitive errors among these runner-ups is similar to German, the high ranking of phonological and verb class errors stands in contrast to the German pattern.

All three of these types were more prevalent than overuse of inflection, root errors, and bare form errors. Unlike German, which had a strongly dispreferred error type (that was less prevalent than all of the remaining errors), in Czech the status as the most dispreferred error type was shared by three error categories—bare forms, root errors, and overuse of inflection. These three did not significantly differ from each other and were each significantly less prevalent than substitution, infinitive, verb class, and phonological errors.

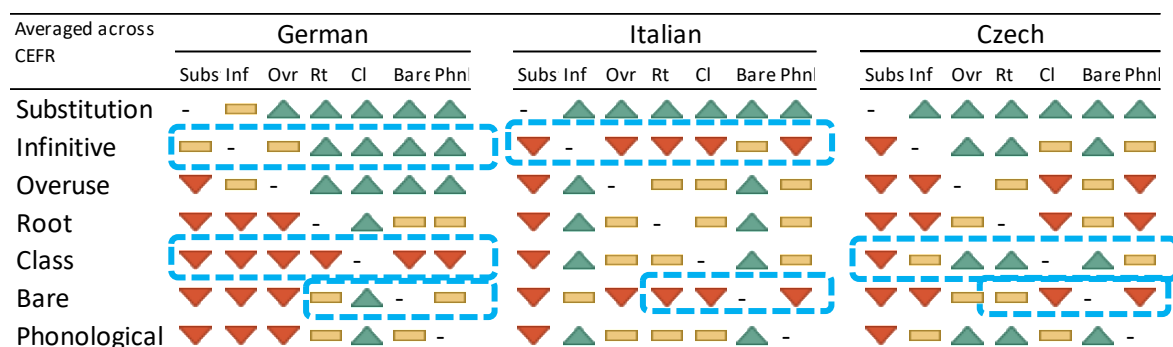


Figure 7. Summary of pairwise comparisons among error type rates within each TL, averaged across all proficiency levels.

*Note.* Comparisons are listed in rows: on the left, categories of errors are listed that serve as the basis of comparison in that row; symbols depict whether the row category had a higher (▲), lower (▼), or not significantly different (—) predicted incidence than the categories listed across the top row. For example, for errors of inflection overuse in German the pattern is (left to right): lower rate than substitution errors (▼); not significantly different rate from infinitive errors (—); higher rate than root, class, bare form, and phonological errors (▲▲▲▲).

Table 17 Rank orders of error types by target language and proficiency level

	German	Italian	Czech
<b>A2</b>			
1 most common	Substitution	Substitution	Substitution
2	Infinitive	Overuse	Infinitive
3	Overuse	Root	Phonological
4	Root	Class	Class



5	Phonological	Phonological	Overuse
6			Root
7 least common	Bare	Infinitive	Bare
	Class	Bare	
<b>A2+</b>			
1 most common	Substitution	Substitution	Substitution
2	Infinitive	Overuse	Infinitive
3	Overuse	Root	Phonological
4	Root	Class	Class
5	Phonological	Phonological	Overuse
6	Bare	Infinitive	Root
7 least common	Class	Bare	Bare
<b>B1</b>			
1 most common	Substitution	Substitution	Substitution
2	Infinitive	Overuse	Infinitive
3	Overuse	Root	Phonological
4	Root	Class	Class
5	Phonological	Phonological	Overuse
6	Bare	Infinitive	Root
7 least common	Class	Bare	Bare
<b>B1+</b>			
1 most common	Substitution	Substitution	Substitution
2	Infinitive	Overuse	Phonological
3	Overuse	Root	Infinitive
4	Root	Infinitive	Class
5	Bare	Class	Overuse
6	Phonological	Bare	Root
7 least common	Class	Phonological	Bare

*Note.* Error types were ranked based on the number of significant pairwise comparisons they participate in. For example, an error type that compared higher than three other error types is ranked higher than one that scored higher than two error types; one that tested significantly lower than another error type in one pairwise comparison would be ranked lower than an error type that did not show any significant differences to

the remaining error categories. Lack of significant differences between error types is indicated by the merging of cells.

Table 18 *Rank orders of error types by target language, averaged across all proficiency levels*

	German	Italian	Czech
<b>All proficiency levels</b>			
1 most common	Substitution	Substitution	Substitution
2	Infinitive	Overuse	Infinitive
3	Overuse	Root	Class
4	Root	Class	Phonological
5	Bare	Phonological	Overuse
6	Phonological	Infinitive	Root
7 least common	Class	Bare	Bare

In Italian, barring substitution errors (which dominated all other error types), the remaining types were equally uncommon: inflection overuse, root, class, and phonological errors were all equally likely amongst themselves and each outnumbered the two least-preferred error categories—infinitive and bare form errors. This dispreference for bare form errors is something that learners of Italian share with learners of Czech, along with the prevalence of verb class and phonological errors in the middle tier of errors. Shared with learners of German is the high ranking of inflection overuse errors among learners of Italian, but the ordering of bare form errors as tied for least-preferred in Italian contrasts with the position of these errors in the middle tier among learners of German. Conversely, the middle-tier ranking of verb class errors in Italian contrasts with their lowermost status in German.

	German	Italian	Czech
1 most common	Substitution	Substitution	Substitution
2	Infinitive	Overuse	Infinitive
3	Overuse	Root	Class
4	Root	Class	Phonological
5	Bare	Phonological	Overuse
6	Phonological	Infinitive	Root
7 least common	Class	Bare	Bare

Figure 8. Cross-over pattern in morphosyntactic and morpholexical errors depending on target-language complexity.

The differences in the patterns of pairwise comparisons detailed above can be linked to how elaborate the morphological systems are (Figure 8). In German, with its two major classes of verbs (“strong” and “weak”) there is less opportunity to apply the wrong inflection templates than in Italian and Czech. In Italian and Czech, with more distinct inflectional endings, some of which are also made up of longer strings of segments, there is more opportunity for learners to be uncertain about the exact phonemic composition of the inflectional endings and roots. In German, the higher or equal prevalence of bare form errors on three out of six contrasts stands out as well, in contrast to the dispreference for bare forms in Italian and Czech. This is consistent with the notion that more saturated morphological paradigms may be conducive to learning about the illegality of uninflected forms.

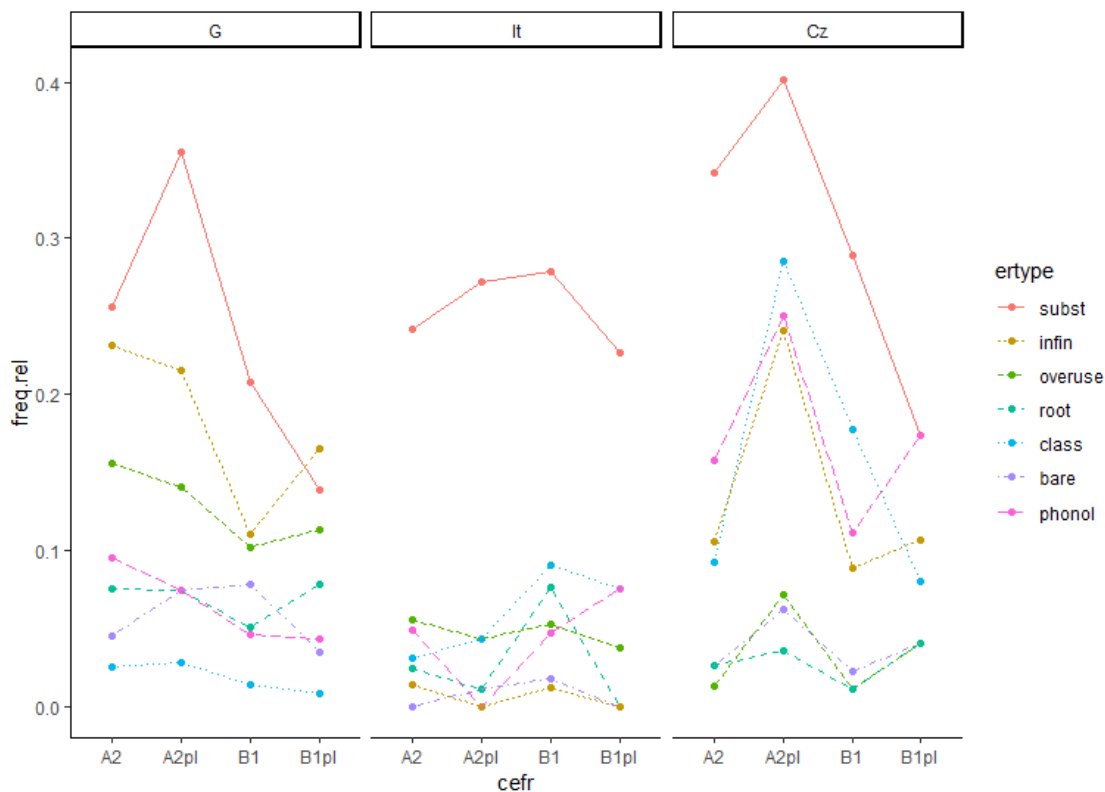


Figure 9. Error rates by type and target language across CEFR proficiency levels

Next, I will examine developmental nuances by zooming in on the cross-sectional changes in the rates of error types for each language separately.

**Role of proficiency.** As the model coefficients foreshadow (Table 13), there were very few instances in which it was possible to capture differences between proficiency levels in the incidence of error types. The most meaningful comparisons in this case are those holding constant target language and error type, while comparing the incidence of each error type between adjacent proficiency levels. The follow-up analysis of pairwise comparisons did not reveal any differences between CEFR levels for German nor Italian, whereas in Czech substitution errors declined between levels A2+ to B1+ by a factor of

2.58 ( $p = .02$ ) and, not statistically significantly, between levels A2+ and B1 by a factor of 2.02 ( $p = .09$ ).

Despite the overall paucity of significant pairwise contrasts, the significant interaction terms for proficiency and error type, on the one hand, and proficiency and target language, on the other, merit a closer look. Examining first the interaction between proficiency and error type, one sees that the only significant combinations of them in the model are: B1 and bare; B1 and class. The interpretation of these parameters in the model is that of a multiplier applied to the baseline category (substitution errors in German at the A2 level), along with the other multipliers (exponentiated regression coefficients)—for proficiency and error type. Thus, the model-predicted rate for errors of verb class in German at the B1 level would be calculated as: intercept (rate of substitution errors in German at A2) \* coefficient for CEFR=B1 \* coefficient for Error type=class \* coefficient for interaction (B1\*class).

The coefficients for the proficiency level of B1 and error types “bare” and “class” were all negative in the model, indicating lower expected prevalence of substitution errors at B1 (than A2, in German); lower expected prevalence of “bare” and “class” errors at A2 (than substitution errors, in German). However, the interaction terms B1\*bare and B1\*class were positive, running counter to the lower-level coefficients. This suggests that both error types improved *less* between the levels of A2 and B1 than would be expected. From this we conclude that in German errors associated with wrong verb class and uninflected bare forms are more persistent developmentally than would be expected from their frequencies alone or from the declines in other error types over the same proficiency span.

Finally, the overall significance of the interaction between proficiency and target language (Tables 14 and 15) appears to stem from the significant differences of the combinations “Czech, A2+” and “Italian, B1” from the reference level, as revealed by the significant model coefficients for these level combinations in the model (Table 13). Again, the significance of the interaction terms is interpreted in the context of the coefficients being additive (on the log scale) and multiplicative on the linear scale. For example, the model-predicted rate for substitution errors in Czech at the A2+ level is pieced together from: the intercept (rate of substitution errors in German at A2) \* coefficient for CEFR=A2+ \* coefficient for Target language=Czech \* coefficient for interaction (Czech \* A2+).

Neither Italian nor Czech had significant coefficients in the model, indicating that their rates of substitution errors at A2 were not significantly different from German, although numerically lower (0.83 and 0.95 times the German rate, respectively). Nor was the A2+ coefficient significant, meaning that rates of substitution errors in German did not noticeably change between A2 and A2+, although there was a trend for them to increase by a factor of 1.11. The coefficient for B1 was significant, with an expected decrease in substitution errors in German of 0.67 times. The interaction multiplier of 1.57 for Czech A2+ means a sharper increase in substitution errors between the levels of A2 and A2+ in Czech than would be predicted either from the pattern in German (over the same proficiency span) or from the pattern of non-significant Czech-German differences (at A2 for substitution). Similarly, the interaction multiplier of 1.90 for Italian at B1 negates the trends towards an overall lower rate of substitution errors in Italian (at A2, compared to German) and towards substitution errors decreasing between A2 and B1 (for

German). Instead, we see a model-predicted two-fold increase in their incidence over what would be expected from those individual trends.

These results are hard to interpret not only due to the cross-sectional nature of the data but also due to the way that proficiency is operationalized in the CEFR framework. Not only is grammatical accuracy part of the construct of proficiency, creating potential for circularity, but so are lexical and pragmatic aspects of language use, which can potentially overshadow accuracy when a holistic rating is produced. Therefore, the absence of significant changes from one proficiency level to the next could reflect the noisiness of proficiency measurement, or the particulars of the proficiency construct definition in the CEFR approach, or a true state of affairs in which it is the typology of the target language that predisposes learners to producing errors of different types at different rates that may change little throughout development.

### 5.3 Cross-Validation

A separate cross-validation analysis was conducted to assess the generalizability of the model proposed in the previous section to previously unseen data. This analysis was performed using the holdout method, in which the splitting of the data and the evaluation of the model occur only in one run (instead of over multiple runs that are subsequently averaged). Data were randomly split and assigned either to the training set (encompassing 59 out of 84 observations) or the testing set with 25 observations. Each observation corresponded to an error count for a particular combination of target language, proficiency, and error type (as presented in Table 10).

First, the mean of error frequency was calculated based on the training set, and the actual values of error frequency in the training data were compared to that mean, yielding

a set of prediction error metrics—mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). These metrics served as the basis against which to compare the performance of the regression model—both when applied to training data as well as when applied to testing data. Second, the Poisson regression model described in the previous section was estimated on the training data, and MSE, RMSE, and MAE were calculated from comparing actual frequency values to model-fitted ones. These values were compared to those obtained from a model with the mean alone. It is helpful to calculate the ratios of model-based MSE, RMSE, and MAE to their mean-based counterparts: values less than 1 would then indicate improved prediction accuracy, and the extent of their difference from 1 would quantify that improvement. Finally, the steps just outlined were repeated on the test data.

Examining the performance of the regression model on the training data, one sees that it considerably improved prediction accuracy (Table 19): depending on the metric in question, the prediction error associated with the model ranged between 0.3% and 6% of the mean-based error—corresponding to a 94 – 91 % reduction in prediction error. When fitted to the testing data, however, the model increased prediction error, compared to using the mean of error frequency: the increase varied by metric from 25% (MAE) to 81% for MSE. Due to this overfitting, successive models were tested further with progressively more model terms removed. All of them also fared poorly and resulted in an increase of error when applied to test data. However, the amount of the increase in prediction error became lower with more terms removed (Table 19).



Table 19 *Prediction accuracy of regression models when tested on unseen test data*

Model	Prediction Accuracy		
	Mean Squared Error	Root Mean Squared Error	Mean Absolute error
<b>3 Interactions: TL*error type + TL*proficiency + proficiency*error type</b>			
<u>Training set</u>			
Mean of frequency	307.46	17.53	12.16
Model (3 interactions)	0.82	0.91	0.72
Ratio (model error / mean-based error)	0.003	0.05	0.06
<u>Testing set</u>			
Mean of frequency	143.85	11.99	9.37
Model (3 interactions)	259.84	16.12	11.76
Ratio (model error / mean-based error)	1.81	1.34	1.25
<b>2 Interactions: TL*error type + TL*proficiency</b>			
<u>Training set</u>			
Mean of frequency	198.16	14.08	10.30
Model (2 interactions)	5.81	2.41	1.84
Ratio (model error / mean-based error)	0.03	0.17	0.18
<u>Testing set</u>			
Mean of frequency	404.16	20.10	12.71
Model (2 interactions)	582.87	24.14	15.46
Ratio (model error / mean-based error)	1.44	1.20	1.22
<b>1 interaction: TL *Error type</b>			
<u>Training set</u>			
Average frequency	184.45	13.58	9.93
Model (1 interaction)	10.35	3.22	2.45
Ratio (model error / mean-based error)	0.06	0.24	0.25
<u>Testing set</u>			
Average frequency	438.86	20.95	2.45
Model (1 interaction)	604.02	24.58	14.95
Ratio (model error / mean-based error)	1.38	1.17	1.12

The issue of overfitting may reflect both the adequacy of the model itself and the ratio of the number of observations to the number of predictors. Especially considering that proficiency was mostly not predictive of error rates, it may be the case that it may be

best operationalized as a variable with fewer levels than in the present study. Moving forward, explorations of this topic should include more data or condense the number of categories investigated. This can be achieved by combining the levels of the error type variable and by collapsing CEFR distinctions, each of which would reduce the number of factors and their combinations in the model. It would have fallen outside of the scope of the present study to test alternative ways of leveling the variables—which could all potentially have consequences for the theoretical implications of the findings and the interpretation of results. Identifying that natural “break” in the data along the proficiency dimension would, no doubt, be enough to merit an entire study of its own. Instead, I opted for setting up the validation maximally close to the original regression models estimated in the previous section.

## Chapter 6: Results—Production of Verbal Inflection in German: Phonological Environments

Research on the production of inflectional morphology by L2 learners repeatedly shows difficulties with producing inflectional endings, with learners either omitting them altogether or modifying them in some way. Recently it has been enriched by the consideration of phonological factors, which influence how morphological marking is realized overtly. This has led to claims that at least some of the observed problems with producing inflectional morphology are due to phonological processes. Some of these factors have included the differences between L1 and L2 syllable structure (e.g., Goad, White, & Steele, 2003) and the application of target-language or first-language (Beebe, 1980; Itakura, 2002; Lee, 2000; Schmidt, 1977; Yu, 2004) sociolinguistic variation patterns by second language learners. Considering that phonological processes are sensitive to grammatical constraints in native speakers (Labov, 1968, 1989; Neu, 1980; Wolfram, 1976), variation in learners' phonological production can indirectly expose the status of different morphological and syntactic environments in learners' interlanguages (Young-Scholten, 1997). Thus, addressing phonological processes, such as consonant cluster simplification, deletion, and assimilation, can strengthen the conclusions of studies on learners' production of morphology, especially where oral samples are concerned, since they would be particularly prone to phonological influences.

In the literature on suprasegmental processes in L2 learning, one distinguishes between universal developmental tendencies and transfer of L1 phonotactic properties (Major, 2001), both of which operate in learners. Among the universal influences,

markedness (Eckman, 1991) and sonority have been proposed to describe the relative prevalence of syllable types across languages and the differences in their learning difficulty. For example, simple codas are both more prevalent in the world's languages and preferred by learners over codas with more consonants, regardless of L1 syllable structure (Tarone, 1982).

Sonority (Cairns & Feinstein, 1982), in turn, complements the notion of markedness by elaborating the preferred and dispreferred sequences of segments. Namely, a syllable's nucleus is typically formed by elements that are the highest on the sonority hierarchy—vowels, and only in exceptional cases by nasals or liquids. Leading up to the nucleus, segmental units are generally expected to rise in sonority, which captures the preference of many languages for onset clusters such as stop + nasal over nasal + stop. By contrast, this preference is reversed in codas, which are required to sequence the segments in the order of *falling* sonority. However, the specific ways in which learners deal with marked material, such as clusters, seem to be influenced by the L1, with some learners preferring epenthesis (e.g., Abrahamsson, 1999), others deletion, and yet others feature assimilation.

Most research on the interactions between phonology and morphology has been conducted on L2 English. English verbs provide fertile ground for such research: consonant clusters are created when past-tense morphology (realized as /t, d/) is applied to stems that may already be consonant-heavy due to the permissiveness of English codas. The findings have been conflicting: in native speakers; deletion of /t, d/ is more likely where these phonemes do not carry grammatical meaning—that is, when they are part of a monomorphemic cluster, rather than when they express past tense (Labov,

1989). By contrast, L2 learners have been reported to show the opposite pattern, being more likely to delete /t, d/ in past tense clusters (Bayley, 1996) and /s/ in the third person singular than on plural nouns (Saunders, 1987), especially when length of residence in a target language community has been short (Wolfram, 1985; Wolfram & Hatfield, 1984). This suggests that early in development deletion may coincide with lack of acquisition of past tense marking. However, as learners acquire patterns of sociolinguistic variation on top of the morphosyntax of their target language, they may also start deleting more, not less, converging with native speaker tendencies (Hansen, 2001, 2005).

These divergent findings stem from differences in first languages among these studies, as well as from sampling from different points along proficiency trajectories, reflecting different levels of mastery of the grammatical features expressed by consonantal morphemes (e.g., tense, agreement, plural marking in English). Another source of variation in the findings lies in the properties of the English inflectional system, which expresses inflectional endings through obstruents and, therefore, conflates learner difficulties stemming from the presence of inflection and those stemming from the consonantal nature of it.

Cross-linguistic evidence becomes indispensable in this situation: by examining languages that realize inflection through different phonological means, one can separate the effects of incomplete grammatical acquisition, universal processes in phonological development, and the gradual learning of sociolinguistically conditioned variation approaching its use by the target language community. Examining learners' success on the same feature cross-linguistically essentially is equivalent to experimentally manipulating those properties of morphemes that have been found to contribute to

learning difficulty. For example, if, as on Goldschneider and DeKeyser's (2001) account, morphemes expressed as full syllables are acquired more easily, it may be fruitful to contrast learners' accuracy on the same grammatical feature in a target language X that expresses it syllabically with the accuracy of learners of target language Y that expresses that feature non-syllabically. This approach complements comparisons conducted within a single language—when, for example, the learning difficulty of non-syllabic third-person *-s* is compared to that of the syllabic *the*. In such comparisons, accounting for the semantic and syntactic differences of the morphological features studied is far from straightforward.

This chapter provides such cross-linguistic evidence by testing the effects of variables first identified based on English as a target language in the context of German. Focusing on the production of inflectional endings on verbs and the phonological environments that promote its variability, such an examination stands to offer insights not only on the learning of German, but also to enter into a dialogue with the English data. German as a target language offers a number of advantages for clarifying the contributions of phonological and phonotactic factors to morphological accuracy. In its present-tense paradigm German includes, among other inflectional endings, the same phonological marker as one of the English past-tense allomorphs, /t/. This takes away the confounding influence of the mastery of tense and aspect semantics from the examination of morphological marking using /t/. In addition, the agreement paradigm of German contains inflectional endings spanning opposite poles on the sonority continuum, from the first person *-e* through the plural (first and third persons) *-en* to the consonantal cluster *-st* (second person singular). Another advantage of German is that it has no

phonological process comparable to /t, d/ deletion in English. Therefore, if the underlying difficulty in producing inflectional endings is grammatical in nature and reflects the gradual acquisition of agreement, these morphemes should show similar degrees of accuracy—similar both among each other and to the difficulty of English *-s* and *-ed*. The operation of syllable structure constraints, by contrast, would be captured in higher accuracy of production on the morphemes expressed as more sonorous strings (*-e*, *-en*) than less sonorous ones (*-t*, *-st*).

While most of the research has been conducted on oral productions, investigating learner behavior through other tasks provides more ways to tease apart the effects of variables such as monitoring—engaged to different degrees in tasks varying in formality and style (Adamson & Regan, 1991; Tarone, 1982), perception (McAllister, 1997; Segui, Frauenfelder, & Hallé, 2001), and motoric output constraints (Flege, 1995; Leather & James, 1996). While this is by no means impossible when examining oral production data alone, broadening the scope of this research to include written data can bring to light the role of familiar factors in novel combinations.

The present chapter examines written production data by learners of German as a second or foreign language. The data come from essays written as part of language proficiency assessments. This places them even closer to the monitored end of the spectrum. Though not meta-linguistic, the task's written mode and, in particular, the purpose of demonstrating proficiency in a target language, would likely involve considerable attention to form and tap both declarative and procedural knowledge. It is also likely that oral and written communication engage linguistic competence and performance differently by virtue of their differing time requirements and opportunities

for revision. In addition, the written channel of communication removes some of the motor difficulty associated with spoken production. Thus, examining data from written production, in addition to the traditional spoken channel, can offer a way to study the old “ingredient” variables—such as monitoring, degree of meta-linguistic control, style shifts, or motor facility—combined in novel ways.

## 6.1 Methods

**Corpus.** From the online corpus of learner productions written for proficiency exams (Merlin), I extracted texts rated by the Merlin project staff at the CEFR A1 level. This returned 55 learner texts, written by test takers of the A1 CEFR exam, with the exception of nine at the A2 and four at the B1 exam levels. In addition to these, 20 texts were randomly selected from those written for the A1 examination but rated as A2 level overall. This was done to increase the sample and include a broader range of learner ability and performance, while keeping task characteristics constant. Care was taken to select samples rated at A2 overall but not uniformly for all aspects of performance. For at least some aspects (such as linguistic range, grammatical accuracy, or vocabulary control) these additional texts were rated A1.

The decision to limit the examination to mostly the A1 and a few instances of the A2 levels stemmed from a number of considerations. First, only limited metadata were available for learners in the corpus, and length of study, type of instruction, or length of residence in target-language communities were unknown. Since proficiency overlaps with time on task, focusing on the lowest proficiency levels locks in this variation in learner backgrounds before it fans out even further. If reaching the A1 level requires highly variable—and, for our participants, unknown—lengths of study, then the paths of



the learners who achieved A2 or B1 would be even more divergent. Second, this chapter investigates the effects of phonology indirectly, mediated by writing and orthographic skills, and growing literacy skills in the L2 would make this connection even more fragile.

Within the texts of this subcorpus a search was conducted for all instances of the “sentence” annotation. Within each sentence, all verbs and participles were manually highlighted and coded for a number of variables. Each verb constitutes one token for the purposes of the analyses reported below.

**Learners.** The first language of learners in the sample varied (Table 20) or was not reported for 24 out of 78 learners in the sample, accounting for 107 out of 479 tokens. All reported L1s allow consonant clusters in the syllable coda. Unless otherwise noted, the analyses reported below are based on all L1s, including unknown ones.

Table 20 *First language backgrounds of learners in the sample*

L1	N of Tokens	Count of speakers		
		Total	Rated A1	Rated A2
Arabic	53	8	8	-
English	88	11	4	7
French	7	1	1	-
Hungarian	3	1	1	-
Polish	14	2	-	2
Portuguese	61	10	6	4
Russian	30	7	7	-
Spanish	69	7	3	4
Turkish	47	7	4	3
not reported	107	21	21	-
<b>Total</b>	<b>479</b>	<b>75</b>	<b>55</b>	<b>20</b>

**Coding and independent variables.** First, each token was coded for correctness—one variable reflected correctness in a strict sense and considered roots and spelling; the other variable focused on the correctness of the inflectional ending itself. Considering that the focus of this chapter is on phonotactic influences on the production of inflectional endings, it is this second outcome variable that is of primary interest. Unless otherwise noted, the analyses reported pertain to the correctness of the ending.

Tokens that contained orthographic errors, such as: using a single vs. double consonant (*hofs* instead of *hoffs*; *hoerren* instead of  *hoeren*); missing the *h* after an *e* signifying a long vowel (*get* instead of *geht*); missing an apostrophe (“*wie gehts*” instead of *wie geht’s*); using *i* to denote the diphthong /ay/ (*ei*)—were considered correct on the inflectional ending variable, as long as the identity of the lexical item could still be determined with confidence. If the root change or spelling error resulted in homonymy with a different word, the case was omitted. For example, *Wie gut es dir?* was dropped, even though *gut* was most likely a misspelled instance of *geht*. By contrast, a token such as *Ich \*brache deine Hilfe* (correct: *brauche*) was retained, considering that *brache* is not a real existing word; such instances were coded as correct on the inflectional ending variable but as incorrect on the “strict” correctness variable.

Each token was classified with respect to the type of predicate, taking into account its syntactic role as an auxiliary, standalone verb, or complement, but also its belonging to broad lexical groups of closed-class (or functional) and open-class (or “content”, thematic verbs).<sup>9</sup> There were four classifications that relied on these two

---

<sup>9</sup> Thematic verbs are contrasted in the syntactic literature with functional verbs and sometimes colloquially referred to as “lexical”. In this chapter, “thematic” is used preferentially to avoid implying that functional verbs somehow are not part of the lexicon or do not have lexical entries of their own.

considerations in different ways. On the first classification, only lexical group membership was taken into account, and predicates were classified as either “thematic” or “functional”. This is consistent with cognitive accounts of L2 learning emphasizing the roles of frequency (functional verbs) and semanticity (thematic verbs). On the second classification, syntactic function was taken into account: thematic verbs appearing alone were classified as forming a simple predicate; functional verbs—auxiliaries (used in analytic tenses) and modal verbs—were coded as auxiliaries; and non-finite verbs were coded as complements (participles in perfect tense and infinitives used for the analytical future and with modal verbs). On the third, hybrid, classification, lexical class membership was combined with the way a verb was used syntactically in context. For example, uses of closed-class *haben* and *sein* with a complement (e.g., as part of analytic tense forms), were coded as auxiliaries, similar to the second classification; but their use with their primary lexical meanings (possession and existence) was classified as thematic (“Auxiliary thematic”). Finally, the fourth classification was the most detailed: it separated functional verbs into modal verbs, auxiliaries, and copulas, and further divided complements into infinitives and participles. The differences are summarized in Table 21.

Table 21 *Classification schemes for predicate type*

Scheme	N of levels	Levels						
Lexical	2	Functional				Thematic		
Syntactic	3	Auxiliary				Simple	Complement	
Hybrid	4	Auxiliary		Auxiliary thematic		Simple	Complement	
Detailed	7	Modal	Auxiliary	Copula	Auxiliary thematic	Simple	Infinitive	Participle

Then, each token was coded for the inflectional ending needed in its context and for the ending actually supplied by the learner. This means that simple, one-verb predicates used without an overt subject could not be analyzed for their appropriateness. For example, *Zu Hause gratuliert zur bestandenen Pruefung* was excluded. By contrast, where there was enough context to establish what a target-like completion should have been, tokens were included. For example, a sentence such as *Wie geht dir?* was included despite missing the required expletive subject: it can be identified as a frequent conversation formula and includes dative case on the experiencer, which rules the experiencer out as a possible subject and indicates that “es” was implied but omitted.

Tokens were annotated for tense: once narrowly and a second time broadly. For simple one-verb predicates the two annotations were identical, with “present” and “preterite” being the only two possibilities. However, for auxiliaries and complements, they differed. In the narrow annotation scheme, the tenses of auxiliaries were annotated without reference to the tense expressed through the auxiliary-complement combination. Instead, the grammatical tense of the auxiliary was marked: *have* in ‘I have finished the book’ was annotated as “present”. Participles and infinitives bear no tense, so “participle” and “infinitive” were also used as their tense annotations. The broad annotation of tense

referenced the tense encoded by the entire complex predicate, and both the auxiliary and its complement receive the same annotation: *have* (perfect) *finished* (perfect).

Finally, tokens were coded for a number of phonological variables, describing the tokens themselves as well as their phonological environments. The variables included were: the segments preceding the inflection, the segment immediately following the inflection, the syllabicity of the required inflection, and the syllabicity of the inflectional ending used by the learner. Preceding and following segments were coded with variables of different levels of detail. On the broadest one, there were three categories: sonorant, obstruent, or sentence (clause) boundary. The next classification incorporated manner of articulation and included vowels, fricatives, stops, affricates, nasals, and approximants. Approximants included glides and liquids, which were grouped together because individually their counts were low.

Syllabicity was coded for the inflectional ending required and that supplied by the learner. For example, if the ending required by context was *-st* but *-e* was supplied, Syllabicity Needed was coded as “no”, because the ending would not have created a new syllable, and Syllabicity Used was coded as “yes”, because the supplied ending introduced a new syllable.

**Predictions.** If the difficulties with producing inflectional endings, as documented in learners of English, arise due to non-felicitous syllable structures, learners of German should exhibit similar patterns of inflection omission. In particular, their production of inflected forms should be more accurate when the form required is one bearing a syllable-forming ending, rather than one creating a consonant cluster: *-e*, *-en* > *-t*, *-st*. Within the group of inflections realized as consonants, the one with a simpler

phonotactic structure, *-t*, should be produced at higher rates of accuracy than the more complex one, *-st*. Moreover, the nature of errors matters as well. Inaccuracies are expected to go in the direction from more complexity (of the required inflected form) to less phonotactic complexity (in the supplied inflected form). Therefore, not only should non-syllabic inflections be produced less accurately, the errors should either involve dropping the ending altogether or substituting it with a more sonorous, syllabic one, and not vice versa.

**Data analysis.** The data were analyzed using the *lme4* package for estimating generalized linear models, implemented in R software (R Core Team, 2013). The outcome variable of interest was the correctness of inflectional endings used by learners. This approach follows variable rule analysis, in that it investigates the factors that are conducive to the correct production of inflection.

Models of different complexity were estimated and included predicate type, required inflection, syllabicity of required inflection, and phonological classes of previous and following segments. Most of the analyses involved different coding schemes of these variables, ranging from the most general to detailed. Unless indicated otherwise, separate models were run on: a. all data—involving tokens in the indicative, imperative, or subjunctive moods; b. data in the indicative only; c. tokens in the present tense of the indicative only. This was done to remove the effects of learners' mastery of the different verbal moods and their semantics.

## 6.2 Results

**Predicate type.** According to the most general classification, which separated thematic from functional verbs, there was no observable difference in accuracy. This held

true regardless of which data were included: with or without the subjunctive, with or without imperative forms, restricted to the present tense or including the preterite (Table 22).

The model with a more detailed, syntactic classification, however, revealed a significant overall effect of predicate type in the dataset with all moods, as indicated by a Wald test:  $\chi^2(3) = 10.6, p = .01$ . When restricted to the indicative ( $\chi^2(2) = 3.2, p = .2$ ) or the present tense only ( $\chi^2(1) = 2.7, p = .1$ ) this effect did not hold. The effect of predicate was driven by the contrast between imperatives and simple predicates, with imperatives being less likely to be produced accurately than simple predicates ( $\chi^2(1) = 9.4, p = .002$ ) or auxiliaries (as indicated by the negative regression coefficient in the model).

The hybrid formulation of predicate type did not change the results: overall effects of predicate were significant in the model run on data in all moods: ( $\chi^2(4) = 12, p = .017$ )—but not on the subsets of indicative ( $\chi^2(3) = 3.6, p = .31$ ) or present-tense tokens ( $\chi^2(2) = 3, p = .22$ ). The baseline in this model was the category of thematic uses of functional verbs, and auxiliary uses of functional verbs and imperatives were both significantly less likely to be supplied accurately than the baseline. Simple thematic predicates trended in the direction of less accuracy than thematic functional verbs. However, pairwise Wald tests did not reveal any significant difference among the non-baseline categories: auxiliaries were not significantly different from simple thematic predicates or complements, and complements did not differ measurably from simple thematic predicates.

Table 22 *Effects of predicate type (four coding schemes) on inflection accuracy*

	Coding Scheme for Predicate Type			
	Lexical	Syntactic	Hybrid	Detailed
	All modes and tenses			
Intercept (= baseline)	1.92***	1.53***	3.18***	3.87***
	Closed-class	Auxiliary	Aux thematic use	Aux thematic use
Coefficient ( <i>p</i> ): all other categories relative to the intercept	-		<b>Aux:</b> -1.51 (< .05)	<b>Aux:</b> -2.44 (.03)
		Complement: 0.42 (.29)	Complement: -1.14 (.15)	Copula: -2.047 (.07)
		Imperative: -0.89 (.06)	<b>Imperative:</b> -2.54 (.002)	<b>Modal:</b> -2.262 (.03)
				Infinitive: -1.74 (.11)
				Participle: -2.08 (.07)
				(.003)
	Thematic: -0.27 (.33)	Simple: 0.53 (.10)	Simple: -1.41 (.06)	<b>Simple:</b> -2.10 (.04)
Null deviance (df)	386.04 (461)	376.77 (454)	372.9 (453)	372.59 (452)
Residual deviance (df)	385.07 (460)	389.07 (461)	359.5 (449)	355.24 (445)
AIC		375.01	369.5	371.24
	Indicative only (incl. present, past tense)			
Intercept (= baseline)	1.91***	1.44***	2.97***	3.66***
Coefficient ( <i>p</i> ): all other categories relative to baseline			Aux: -1.35 (.08)	<b>Aux:</b> -2.24 (.04)
				Copula: -1.84 (.10)
				<b>Modal:</b> -2.19 (.04)
		Complement: 0.59 (.16)	Complement: -0.94 (.24)	Infinitive: -1.53 (.16)
				Participle: -1.87 (.10)
	Thematic: -0.05 (.87)	Simple: 0.57 (.10)	Simple: -1.20 (.11)	Simple: -1.89 (.07)
Null deviance (df)	312.89 (399)	311.75 (395)	311.75 (395)	311.46 (394)
Residual deviance (df)	312.86 (398)	308.73 (393)	307.15 (392)	303.13 (388)
AIC	316.86	314.73	315.15	317.13
	Present tense only			
Intercept (= baseline)	1.83***	1.43***	2.92***	3.61***
Coefficient ( <i>p</i> ): all other categories relative to baseline			Aux: -1.33 (.08)	<b>Aux:</b> -2.18 (< .05)
				Copula: -1.85 (.10)
	Thematic: -0.05 (.88)	Simple: 0.56 (.10)	Simple: -1.15 (.13)	<b>Modal:</b> -2.16 (< .05)
				Simple: -1.84 (.08)
Null deviance (df)	242.37 (297)	241.76 (295)	241.76 (295)	241.76 (295)
Residual deviance (df)	242.35 (296)	239.18 (294)	237.74 (293)	234.39 (291)
AIC	246.35	243.18	243.74	244.39

*Note.* In the Detailed classification, “copula” refers to uses of the verb *sein* (‘be’) with nominal or adjectival complements (e.g., “I am tired”, “His eyes are blue”), whereas the “auxiliary” category includes analytical forms with verbal complements (e.g., “I have traveled”).



Finally, the set of models run with the most detailed classification of the predicate revealed differences among the factor levels (of predicate type) but no overall significant effect of the variable. In the data set that included all moods and tenses, there was a trend approaching significance: ( $\chi^2 (7) = 13, p = .07$ ). In the data set of indicative tokens only, there was no measurable effect of predicate type: ( $\chi^2 (6) = 5.4, p = .49$ ). Nor was it present in the data set of present-tense tokens: ( $\chi^2 (4) = 4.4, p = .35$ ).

However, individual levels of predicate type did sometimes differ from each other. In the data set with tokens in all moods, thematic uses of functional verbs were significantly more likely to be used correctly than non-thematic, auxiliary uses, as well as modal verbs, imperatives, and simple predicates with one thematic verb (Table 22). In the data set with indicative tokens only this effect extended to auxiliaries (used as such) and modals being produced less accurately than thematically used auxiliaries, whereas simple thematic predicates only trended in that direction. These differences still held when the examination was restricted to present-tense tokens.

Thus, one of the most robust findings in this set of analyses was the advantage of predicates formed by functional verbs used with their primary lexical meanings over their auxiliary uses and imperative, while *tending*, less reliably, to also be more accurate than simple one-verb predicates. This may reflect the beneficial role of semantic load on learners' production of grammatical elements and its corollary—redundancy. Even though more accurate production of auxiliaries is typically interpreted as reflecting the overlearning of closed-class elements, it was noteworthy that performance on this group of verbs varied depending on how they were used in context. The advantages of the auxiliaries' high frequency were only apparent when they also carried some semantic

function in the sentence, while disappearing when in the presence of a semantically superior complement.

**Syllabicity of ending.** Inflectional endings were not more likely to be produced correctly in environments requiring a syllabic ending than in the environments requiring a non-syllabic ending (Table 23). Thus, the endings that form a new syllable (*-e* and *-en*) were as likely to be supplied accurately as those that are expressed as consonants or consonant clusters (*-st*, *-t*). The reverse also held: when syllabic and non-syllabic endings were supplied by learners, they were equally likely to be the correct ones (Table 23).

Table 23 *Effects of syllabicity on accuracy of production*

	All moods and tenses	Indicative mood only	Present tense only
<b>Syllabicity of Required Ending: Baseline—Non-syllabic ending required</b>			
Intercept (= baseline)	1.95***	1.85 ***	1.85***
Coefficient: all other categories relative to baseline	Syllabic Needed: -0.17 (.53)	-0.03 (.92)	-0.15 (.64)
Null deviance (df)	351.68 (443)	320.32 (399)	245.96 (297)
Residual deviance (df)	351.29 (442)	320.31 (398)	245.75 (296)
AIC	355.29	324.31	249.75
<b>Syllabicity of Supplied Ending: Baseline—Non-syllabic ending supplied</b>			
Intercept (= baseline)	1.69***	1.73***	1.80***
Coefficient: : all other categories relative to baseline	Syllabic supplied: 0.13 (.61)	0.19 (.5)	-0.06 (.86)
Null deviance (df)	385.73 (460)	320.61 (400)	245.96 (297)
Residual deviance (df)	385.46 (459)	320.16 (399)	245.93 (296)
AIC	389.46	324.16	249.93

**Previous phonological segment.** When examining the role of the previous segment, we excluded instances of irregular verbs. Irregular verbs form their different person-number combinations not by deriving them from the “dictionary” entry of the verb but through suppletion. Without a clear “inflection”, it cannot be clearly determined

what “precedes” it. We also conducted separate analyses on the data, with forms of the subjunctive included or excluded. The subjunctive presents a curious case for testing the effects of syllabicity and sonority: in German it is formed with the suffix *-te*, followed by the same endings as in the indicative, except for the first and third persons singular, which are not followed by an ending. Thus, phonological considerations would predict higher accuracy on the subjunctive than indicative forms.

The broadest classification of segments preceding an inflection included the factor levels “obstruent” and “sonorant”, and “subjunctive” as a separate class. There was no effect of preceding segment on inflection accuracy: with subjunctive included, the coefficients for obstruents and sonorants were not significant, and neither was a Wald test for the overall effect of previous segment ( $\chi^2(2) = 4, p = .13$ ). No effect of preceding segment was observed when only indicative and imperative tokens were included, nor when the analysis was restricted to the present tense (Table 24).

Table 24 *Effects of previous segment class on accuracy of inflectional ending: Obstruents versus sonorants*

	All moods and tenses (except irregular)	Indicative mood only (except irregular)	Present tense only (except irregular)
<b>Obstruents vs Sonorants vs Subjunctive</b>			
Intercept (= baseline)	3.04**	1.75***	1.62***
	Subjunctive	Obstruent_Ending	Obstruent_Ending
Coefficient: all other categories relative to baseline	Obstruent: -1.55 (.13) Sonorant: -1.12 (.28)	Sonorant: 0.19 (.54)	Sonorant: 0.33 (.37)
Null deviance (df)	348.02 (408)	279.52 (348)	205.77 (248)
Residual deviance (df)	343.09 (406)	279.15 (347)	204.95 (247)
AIC	349.09	283.15	208.95
<b>Subjunctive removed: Obstruents vs Sonorants</b>			
Intercept (= baseline)	1.49 ***		
	Obstruent_Ending		

Coefficient: all other categories relative to baseline	Sonorant: 0.43 (.13)
Null deviance (df)	337.24 (386)
Residual deviance (df)	334.95 (385)
AIC	338.95

---

A more detailed classification of preceding segments separated vowels and consonants, while further subdividing the consonants according to manner of articulation. There was only one token where the inflectional ending was preceded by an affricate, which was excluded. The baseline category for these comparisons was approximants. None of the models showed an overall effect of the manner of articulation of the preceding segment (Table 25). However, there were significant differences between factor levels in the analyses of present-tense tokens: the regression coefficient for fricatives differed significantly from that of vowels ( $\chi^2(1) = 4.5, p = .03$ ), and approached a significant difference from stops ( $\chi^2(1) = 3.7, p = .056$ ), whereas vowels and stops did not differ.

Table 25 *Effect of previous segment on inflection accuracy: Manner of articulation*

	All moods and tenses (regular)	Indicative mood only (regular)	Present tense only (regular)
Baseline: approximant			
Intercept	1.61*	2.30*	2.30*
Coefficient	Fricative: -0.18 (.82)	-0.80 (.45)	-1.08 (.32)
	Nasal: 0.13 (.88)	-0.63 (.56)	-0.85 (.44)
	Stop: -0.05 (.95)	-0.15 (.89)	-0.02 (.99)
	Vowel: 0.49 (.55)	-0.19 (.86)	-0.03 (.98)
Null deviance (df)	336.89 (385)	279.23 (348)	205.45 (247)

Residual deviance (df)	333.67 (381)	275.58 (347)	198.19 (243)
AIC	343.67	285.58	208.19
Wald test (overall effect)	$\chi^2(4) = 3, p = .55$	$\chi^2(4) = 3.6, p = .46$	$\chi^2(4) = 7, p = .14$

**Following segment.** Following segments were coded at two levels of detail: the most general classification involved obstruents, sonorants, and sentence or clause boundary; the detailed classification included a division into vowels and consonants, which were further subdivided based on manner of articulation. Approximants were excluded due to their low number (three tokens).

In the most inclusive data set with tokens of all moods, no effect was observed for the class of following phonological segment: this was reflected in the regression coefficients for factor levels “obstruent” and “sonorant”, neither of which was statistically significant. This was mirrored by the results of the Wald test, which did not reveal an overall effect for this variable. The detailed classification based on manner of articulation did not yield a statistically significant effect. No measurable effect was found in the datasets of indicative tokens or present-tense tokens either (Table 26).

Table 26 *Effect of following segment on inflection accuracy*

	All moods and tenses	Indicative mood only	Present tense only
<b>Obstruents vs Sonorants: baseline—sentence (clause) boundary</b>			
Intercept (= baseline)	2.18***	2.44***	2.20 (.04)
Coefficient: all other categories relative to baseline	Obstruent: -0.39 (.35)	-0.55 (.25)	-0.37 (.73)
	Sonorant: -0.65 (.12)	-0.82 (.09)	-0.42 (.69)
Null deviance (df)	384.76 (457)	311.75 (395)	237.24 (292)
Residual deviance (df)	382.10 (455)	308.48 (393)	237.06 (290)
AIC	388.1	314.48	243.06
Wald test (overall effect)	$\chi^2(2) = 2.5, p = .28$	$\chi^2(2) = 3, p = .23$	$\chi^2(2) = 0.16, p = .92$
<b>Manner of articulation: baseline—affricate</b>			
Intercept (= baseline)	1.25 (p > .1)	1.79 (p < .10)	1.79 (p < .10)

Coefficient: all other categories relative to baseline	Boundary: 0.93 (.29)	0.65 (.57)	0.40 (.79)
	Fricative: 0.50 (.57)	-0.02 (.98)	-0.21 (.85)
	Nasal: -0.04 (.97)	0.15 (.90)	0.51 (.70)
	Stop: 0.60 (.48)	0.15 (.89)	0.14 (.90)
	Vowel: 0.37 (.65)	-0.22 (.84)	-0.11 (.92)
Null deviance (df)	384.76 (457)	311.75 (395)	237.24 (292)
Residual deviance (df)	380.87 (452)	308 (390)	235.88 (287)
AIC	392.87	320	247.88
Wald test (overall effect)	$\chi^2 (5) = 3.9, p = .56$	$\chi^2 (5) = 3.5, p = .63$	$\chi^2 (5) = 1.3, p = .94$

Therefore, in the present data inflectional endings were not more or less likely to be produced correctly depending on the phonological segment that followed them.

**Composite models.** The predictors were also tested in multiple-predictor models, to account for the possibility that some of the effects are only apparent in the presence of the others. Because the predictors are factors, the number of their levels included in these models was kept to a minimum to create only broad classifications, to prevent data sparseness. Irregular verbs that employ suppletion in their paradigms were excluded. The models were estimated based on tokens in the indicative and imperative moods and excluded forms of the subjunctive (see section “Previous Phonological Segment” above for rationale).

Table 27 *Joint effects of phonological environment on inflection accuracy*

Parameter	Predictor Level	Estimate ( <i>p</i> )*	Wald test
<b>Additive effects of preceding and following segment</b>			
Intercept (= baseline)	<b>Obstruent_Ending_#</b>	2.36***	
Previous Segment	<b>Sonorant_Ending_#</b>	0.19 (.55)	
Following Segment (baseline: sentence boundary)	Obstruent_Ending_ <b>Obstruent</b>	-0.52 (.30)	$\chi^2 (2) = 3.5, p = .18$
	Obstruent_Ending_ <b>Sonorant</b>	-0.87 (.07)	
<b>Interacting effect: Preceding x following segment</b>			
Intercept (= baseline)	<b>Obstruent_Ending_#</b>	2.99***	
Previous Segment (baseline: obstruent)			

Following Segment (baseline: sentence boundary)	<b>Sonorant_Ending_#</b>	-1.01 (.26)	
	<b>Obstruent_Ending_Obstruent</b>	-1.18 (.14)	$\chi^2 (2) = 4.6, p = .1$
	<b>Obstruent_Ending_Sonorant</b>	<b>-1.63 (.03)</b>	
Interaction	Sonorant_Ending_Obstruent	1.26 (.22)	
	Sonorant_Ending_Sonorant	1.52 (.14)	
Model comparison			$\chi^2 (2) = 2.37, p = .30$

The interaction between the nature of the following and previous segments did not reach statistical significance, but the inclusion of this interaction term made the effect of following segment statistically significant (Table 27; Figure 10). Despite this, in neither model was the overall effect of following segment class significant. The model with the interaction did not fit the data better than the additive one (Table 27).

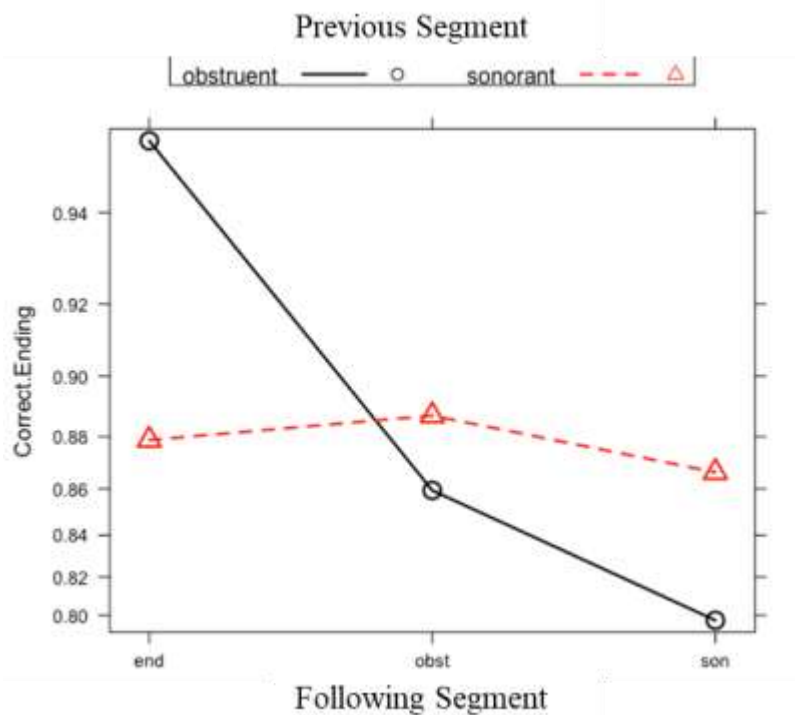


Figure 10. Interaction between class of following phonological segment (x axis) and previous phonological segment (y axis) in affecting inflection accuracy.

Another set of models combined phonological environment with the syllabicity of required ending. The effects of these three predictors (syllabicity, previous segment, following segment) were tested with and without an interaction between the previous and following segment (Table 28).

Table 28 *Combined effects of syllabicity of ending and phonological environment on inflection accuracy*

Parameter	Predictor Level	Estimate	Wald test
(p)*			
Syllabicity of Required Ending + Environment (Syllabicity + Preceding Segment + Following Segment)			
<b>Additive model</b>			
Intercept	Obstruent_Ending [Non-syllabic]_#	2.25***	
Syllabicity of Inflection (baseline: non-syllabic)	Syllabic	0.10 (.79)	



Previous Segment (baseline: obstruent)	Sonorant	0.24 (.51)	
Following Segment (baseline: sentence boundary)	Obstruent	-0.50 (.33)	$\chi^2(2) = 3.4, p = .18$
	Sonorant	-0.86 (.08)	
<b>With two-way interaction (Syllabicity + Previous x Following Segment)</b>			
Intercept	Obstruent_Ending [Non-syllabic]_#	2.81***	
Syllabicity of Inflection (baseline: non-syllabic)	Syllabic	0.19 (.62)	
Previous Segment (baseline: obstruent)	Sonorant	-0.97 (.28)	
Following Segment (baseline: sentence boundary)	Obstruent	-1.14 (.16)	$\chi^2(2) = 4.7, p = .09$
	<b>Sonorant</b>	<b>-1.63 (.04)</b>	
Interaction	Sonorant_E_Obstruent	1.29 (.22)	
	Sonorant_E_Sonorant	1.61 (.12)	
Model comparison			$\chi^2(2) = 2.59, p = .27$

---

Looking at the exact nature of learners' departure from expected inflections the morphemes expressed by segments that arguably create phonotactic difficulties, *-t* and *-st*, were used correctly in 23/27 and 16/16 cases, respectively. By contrast, *-e* and *-en*, both salient by virtue of forming a separate syllable and by including vowels, as well as easy to articulate, were supplied less accurately: 26/37 and 35/44 cases, respectively. When one examines what forms were used in their stead, one sees that *-e* and *-en* were sometimes used interchangeably, and substitutions of one for the other are the most frequent error class for both morphemes: for *-e*, four out of nine errors were uses of *-en*; for *-en*, six out of nine errors were uses of *-e*. Both *-e* and *-en* were also omitted (2/34 and 2/42, or 2/9 errors on *-e* and 2/9 on *-en*)—in contrast to *-st*, which was always produced correctly (13/13) and *-t*, which was omitted in three instances out of 27 against an overall highly accurate backdrop with only five total errors. Admittedly, a salience-

based account of learning can accommodate these findings if one specifies that learning may be reflected in the suppliance versus omission of inflection and not only in the suppliance of a *correct* inflectional ending.

### 6.3 Conclusions

In contrast to English, all of whose overt tense-marking endings would be considered infelicitous, German possesses a range of endings in its inflectional paradigm that vary in their phonological properties. The analysis of both syntactic and phonological environments allowed me to rule out phonological processes as an exhaustive explanation of learners' difficulties with morphology, at least as far as written production is concerned.

Overall, learners did not appear to incorrectly produce the inflectional endings considered phonotactically disfavored at disproportionate rates. There were no effects of the syllabicity of required ending on production accuracy, and the following or preceding segments did not appear to influence accuracy. Of interest was the finding that functional verbs, which belong to a closed class, are produced at a higher accuracy—an observation that mirrors prior research. However, this general pattern was refined in the present study, where it only applied if the closed-class elements were used thematically. This likely illustrates the benefits of semantic emphasis, which is standard for general-cognitive proposals (e.g., DeKeyser, 2001, 2005; VanPatten, 2005), compounded, potentially, by the higher probability of rote learning of these verbs (due to their being closed-class). On an account featuring semanticity only, there should have been no difference between thematically used auxiliaries and open-class thematic verbs. On an account that attributes the learning benefits to the closed-class verbs' higher overall frequency, no difference

should be expected between auxiliary and thematic uses of these closed-class verbs. Furthermore, imperatives were less likely to be produced correctly, despite being formed by means considered to be felicitous—bare verb stems or, for some verbs ending in consonant clusters, the ending *-e*.

The variable rule approach has been applied in the past to data from learners of English, but its applicability to other target languages, such as in this case—German, is less straightforward. The data from learners of German analyzed in the present chapter imply that the variation in learners' production of inflectional morphology may be more nuanced than what variable rule analysis presupposes. In this sample, inaccurate productions were hardly a simple matter of suppliance or omission. They included not only bare and infinitival forms, but also finite forms substituted for one another, inaccurate application of root alternations or inflectional patterns associated with a different verb class. To capture even a fraction of these options in learner behavior, I had to create multiple outcome variables that described varying degrees of accuracy, while focusing selectively on the endings, to the exclusion of omitted main verbs and complements, as well as overapplication of root processes, or selection errors.

These results may mean that phonological processes do not fully account for the full scope of learners' inaccuracies in the use of inflectional morphology. At the same time, it is entirely plausible that phonological competence may impose a ceiling on what can be articulated in production, or, to a more limited extent, expressed in writing. The written channel of the task enabled an analysis of samples obtained in a high-stakes context, in which learners can be expected to recruit monitoring processes and demonstrate maximum possible accuracy. However, a disadvantage of this lies in the fact

that the phonology of learners' interlanguages was not studied directly but through the lens of writing. In addition to monitoring, writing also provides more opportunities for revision and multiple checks of one's production output, in contrast to oral spontaneous speech. Therefore, oral speech data could be expected to show the lower bounds of learners' potential accuracy. Despite these differences in the demands of oral and written production, those differences do not fully explain why syntactic environments would and phonological environments would not affect production accuracy.

## Chapter 7: Discussion and Conclusions

This chapter will recapitulate key research aims and findings, concentrating above all on the takeaways they offer for theories of L2 learning and their place within other insights into L2 morphological development accumulated up to this point. I will then discuss the limitations of this research and its contributions to the field.

### 7.1 Key research aims and findings

In this dissertation, I set out to accomplish two major goals. The first one was to provide a set of facts of morphological learning that would be simultaneously cross-linguistic and rooted in learner productions. The second aim was to use these cross-linguistic data to test the predictions and assumptions extended from theories of L2 learning, including the learning mechanisms that are and are not compatible with them and the effect that the complexity of the linguistic system to be learned might have on the rate of learning.

The empirical findings (reported in Chapters 5 and 6) are recapitulated below in condensed form. Their interpretation in light of current theoretical models of learning is the subject of the following section.

**Learners of Czech were as successful as, and learners of Italian—more successful than, learners of German at using inflected, finite forms when those were required by context.**

Learners of all TLs were equally likely to substitute inflected forms for one another, but learners of German were more prone to use uninflected and non-finite forms in place of finite ones, compared to learners of Italian, and were as likely to use them as were learners of Czech. Between Czech and Italian, learners of Czech were more likely to use

infinitival forms (when finite were required) and equally likely to use bare uninflected forms. This pattern is surprising in that learners of Czech do not perform significantly worse, compared to learners of German, despite the demonstrably higher number of morphological contrasts expressed through distinct morphemes in Czech. Learners of Italian had an edge over learners of Czech, likely due to the slightly higher number of distinct inflections in Czech and the less predictable system of verb classes. Overall, the lower than expected prevalence of these errors in learners of Czech and Italian would support an account of richer paradigms pushing learners to acquire the principle of needing *some* form of agreement marking.

**Learners of Italian and Czech were less likely to overuse inflection than learners of German.** Learners of German were also less able to constrain inflection appropriately, overusing it in contexts where non-finite forms were required at a higher rate than learners of Italian and Czech. Learners of Italian and Czech were equally likely to make these errors.

**Learners of German made fewer verb class errors than learners of Czech and were at least as good as learners of Italian.** In a contrast with errors of infinitive and bare form use, errors pertaining to verb classes did show evidence of increasing hand-in-hand with the complexity of the TL's verb class system. For learners of German, a language with a simpler, two-way class system provided an advantage, compared to the learners of Czech, which has the most complex verb-class system of all TLs. Learners of Italian were, for the most part, as good as learners of German on this attribute (with the exception of one CEFR level, B1, where they were outperformed by the learners of German).

**Learners of German made fewer phonological errors than learners of Czech and performed similarly to learners of Italian.** Phonological errors involved incorrectly supplied segments of the root or inflectional ending that resulted in a non-existent form. By contrast, an incorrectly supplied ending that resulted in an existing form would be classified as a substitution error, whereas an incorrectly supplied root segment that resulted in an existing form would be classified as a root process error. These errors possibly signify a failure to retrieve the correct inflected form or inflectional ending by learners of Czech and, less so, Italian. This, in turn, may reflect the processing costs associated with storing a more diverse set of inflected forms in the mental lexicon.

**In German, there was no accuracy advantage for syllabic, sonorous inflectional endings (compared to non-syllabic, obstruent endings).** While phonological factors are undoubtedly at play during the learning of morphology and its production, Chapter 6 showed on the basis of German data that phonological explanations alone are insufficient: the nature of phonological segments preceding and following inflectional endings in German did not appear to influence the accuracy of inflection use. Nor was there a measurable difference in accuracy between endings that are expressed as full syllables (-e, -en) and those expressed as obstruents and clusters of obstruents (-t, -st).

**In German, non-syllabic inflections were *not* replaced with syllabic ones.** Learners' errors did not follow the patterns predicted by accounts positing phonological simplification as the cause of morphological errors. In particular, marked endings were not omitted or substituted with less marked ones. Most of the substitution errors involved

substitutions between *-e* and *-en* in both directions, whereas *-t* and *-st* were supplied correctly<sup>10</sup>.

**In German, learners were more accurate on functional (closed-class) verbs used in their primary lexical meaning than on auxiliaries, copulas, and modals.**

Going beyond a simple distinction between closed- and open-class elements, accurate production was facilitated by the syntactic function of closed-class elements. When they were used as thematic verbs (i.e., in their primary lexical meanings, cf. *I have a dog* vs. *I have bought a dog*), accuracy was higher.

The findings related to developmental patterns and their differences among the TLs were the least clear. The proficiency variable did not meaningfully participate in interactions with TL and error type, with a couple of exceptions among the many contrasts explored<sup>11</sup>. On the one hand, this may imply that the cross-linguistic differences described (Chapter 5) are a steady influence throughout development. On the other, this absence of differences may be a type II error stemming from the cross-sectional nature of the data (see Limitations below). It may also be the case that the construct of proficiency adopted in the CEFR framework was too broad to correlate reliably with just one aspect of grammatical performance.

## 7.2 Theoretical implications and takeaways

Overall, different types of errors formed two broad clusters with respect to differences among TLs: errors of substitution, infinitive use, bare form use, and inflection

---

<sup>10</sup> It is possible that the overall distributional properties of the TL desensitize learners to marked phonological material, and consonantal clusters in inflectional endings lose their difficulty as learners become more familiar with clusters across the board. However, this is unlikely in this sample of learners who were at the extreme low end of the proficiency spectrum.

<sup>11</sup> As detailed in Chapter 5, the analyses incorporated corrections for multiple comparisons.



overuse patterned in cluster; and errors of verb class and phonological processes formed the other. This implies that they might rely on shared learning mechanisms. The first group engages rule-like compositional properties that could be more amenable to facilitation at an abstract level. The second group shares the similarity of being grounded in morpholexical learning, which is item-specific and, therefore, closer to the lexical end of the continuum and less prone to benefitting from abstract generalization.

For the first, morphosyntactic, group of processes, German did not show a learnability advantage over Czech, and both were disadvantaged compared to Italian. Therefore, higher paradigm complexity is at least not detrimental to learning and can even be facilitative to acquiring the need for agreement marking. For the second, morpholexical, group of processes, learners of German were at an advantage over learners of Czech and at no disadvantage compared to learners of Italian. It may also be the case that this study did not have sufficient power to capture a difference between German and Italian, which only differ by one verb class.

Thus, on morphosyntactic dimensions of inflectional morphology use learners of TLs with more complex paradigms were at an advantage, and their errors occurred at a rate that was out of proportion to the TL complexity. In other words, their error rates were lower than what would be expected simply from extrapolating error rates from L2 English data. On the morpholexical dimensions, learners' errors increased accordingly if their TL had a more complex system in this respect (morphophonological alternations, verb classes).

The patterning of errors in this way may suggest that the types of knowledge underpinning them may require different amounts and kinds of evidence from the input,

while also being differentially responsive to this evidence or even instruction. Even though syntactic accounts do not explicitly posit such a split, this pattern is broadly consistent with some of the ideas inherent to them. The lack of a noticeable handicap for learners of Italian and Czech (relative to German) can be taken as a learnability benefit resulting from those TLs' higher number of distinct (non-homophonous) morphemes, which provide unambiguous evidence every time agreement marking is encountered in the input. By contrast, learners of German, while also receiving input with overt agreement marking (on all person-number combinations), are presented with forms that are homophonous with the infinitive for one-third of the paradigm. While this homophony is still situated in a paradigm where all person-number combinations are overtly marked—in contrast to English, for example—homophony with the infinitive in particular may be especially damaging to grammar building.

Syntactic accounts do posit the necessity for morpholexical learning, however, and the disadvantage on errors of verb class and phonological errors for learners of Czech is broadly consistent with this notion. Since such learning is often deemed to be outside of the scope of syntax, theories of L2 syntactic learning do not go any further in specifying just how it occurs.

In explaining the German data (Chapter 6) on the role of syntactic environments in the production of inflection, syntactic accounts are less successful. The distinction between free and bound morphology that they employ translates in this study into the categories of “auxiliary” and “simple predicate”, which did not sufficiently explain learners' accuracy. Instead, a finer-grained classification proved predictive of learners' performance, in which closed-class freestanding morphemes (“auxiliaries”) were treated

differently depending on their syntactic function (as thematic, or “main”, verbs versus true auxiliaries).

With respect to general-cognitive views of learning difficulty, this dissertation’s results run counter to these views where morphosyntactic learning is concerned—that is, the use of bare and infinitival forms and overuse of inflection. The results concerning morpholexical learning, however, are broadly consistent with them: a higher number of verb classes in a TL coincided with more learning difficulty and higher error rates.

In addition, the data on the influence of syntactic environments on the production of inflection in L2 German (Chapter 6) have the strongest affinity with the GC account among all the findings in this paper. Verbs belonging to the so-called “closed-class”—that is, auxiliaries and copulas—were more likely to be inflected correctly when they were used in their primary lexical meanings and not as auxiliaries or copulas. They were also used more accurately than lexical verbs that do not belong to the closed class (so-called thematic verbs). Viewed through a syntactic lens, this is highly surprising: after all, on a few syntactic accounts free morphemes (i.e., auxiliaries) are learned before bound morphemes (i.e., inflections on thematic verbs). Since it is computing syntactic relationships that drives difficulty, according to these accounts, any verb used thematically should be more, not less, difficult to produce. Conversely, if one were to argue that these closed-class verbs still benefit from their “free” nature *over time*, then there should be no difference depending on the manner of their use in any given sentence. Alternatively, this discrepancy may also reflect layered influences of processing and longer-term learnability during the production of any given token of a closed-class verb. For instance, long-term learnability may, in fact, benefit from the salience associated with

their being free-standing. Yet, at the same time, the correct production in instances of *thematic* use may be additionally facilitated by learners' attention directed towards them when their semantic meaning is essential to the utterance.

On the GC account, by contrast, auxiliaries are considered more salient by virtue of their free-standing nature, similarly to their treatment in syntactic accounts (even though it is not clear whether the inflectional endings on auxiliaries benefit from this fact as well). Additionally, the GC model allows for auxiliaries to benefit from their overall high frequency, which is combined over both types of use—as true auxiliaries and in their primary lexical meanings. Finally, the semantic dimension of salience would be facilitative as well, due to the lower redundancy of these elements once they are used in primary lexical meanings, compared to uses where the carrier of lexical meaning is the complement (e.g., infinitive or participle).

Finally, the phonological data from L2 German run counter to some of the predictions of the general cognitive account, on which morphological elements that are syllabic are more phonologically salient and, thus, learnable. In this study, endings realized syllabically were not produced any more accurately by learners of German than were non-syllabic ones, at least in the written channel. This also serves as a counterpoint to data from L2 English, which have been explained through the lens of salience (in particular, with respect to the production of third-person singular *-s*). Learners of German show that, in principle, there is nothing impossible about producing a non-syllabic ending, which is just as redundant in German as it is in English.

Phonological factors have also been considered by theoretical approaches other than the general-cognitive account of grammatical learning, including the literature on

interlanguage phonology (and its interactions with morphology). These have included consonantal cluster simplification and examinations of segments preceding and following inflectional endings. Again, the German data examined in Chapter 6 did not find evidence consistent with these explanations. In particular, learners did not substitute inflections in the predicted pattern (more sonorous or syllabic endings for less sonorous, non-syllabic ones).

On balance, the lack of a “slowdown” in learning observed among learners of Italian and, to some degree—Czech, compared to learners of German strongly suggests that it is premature to discard syntactic explanations of morphological learning. The current pattern of findings can be most economically explained if one posits some degree of abstract knowledge at the level of syntactic features or at the level of presence/absence of inflection. While the general-cognitive approach can account for the findings related to morpholexical errors, its current formulation as difficulty scores assigned to individual morphemes does not accommodate any abstract facilitation, nor does it explain the learning of zero-marked forms. That is not to say that a general-cognitive account of these cross-linguistic data cannot be formulated in principle; only that to put forward such an account one would need to address the non-linear patterns in learning and to explain the acquisition of contrasts between marked and zero-marked forms, and the acquisition of constraints that prevent the overuse of inflected forms.

### 7.3 Limitations to consider in future research

The first group of limitations concerns the methodological tradeoffs that were made inevitable by time and resource constraints. The second group of limitations pertains to conceptual issues, such as the choices that were made with respect to

operationalizing key constructs and the limits that places on the interpretation of the results.

First among the methodological tradeoffs was the use of existing learner corpus data in the service of the cross-linguistic focus of the study. This use of existing corpora entailed a reliance on cross-sectional data as a stand-in for longitudinal data. Learner proficiency was used as a proxy for points along the development trajectory. A better way, of course, would be to use length of residence or amount of instruction in the target language. The approach taken in the present study poses at least two problems. First, using proficiency as a proxy for time on task, length of learning, or amount of linguistic input creates a circularity: grammatical accuracy, of which the mastery of inflection is a part, is one of the elements in the construct definition of proficiency. Therefore, learners who have received the same “quantities” of input or instruction and show differences in accuracy would be rated at different proficiency levels, which would then erase any differences in error rates between the proficiency levels. Second, the role of proficiency in the analysis as an ordinal variable, rather than a ratio or even an interval, limits the interpretations that can be drawn from the data. Any regression coefficients associated with proficiency levels would only indicate that accuracy is “different”, without showing a stepwise increase in accuracy going along with a stepwise increase in proficiency.

Finally, there is the issue of proficiency being based on raters’ judgments and their interpretation of what it means to be an “A2” speaker of their language. This opens the door to the possibility that raters in different target languages treat inaccurate production of inflection differently, depending on a learner’s use of other linguistic

resources and the learner's success at communicating propositional and even pragmatic meaning—achieved through whatever means.

The use of previously collected corpus data also meant having limited data about learner language backgrounds. More information related to learner L1s and the contexts in which they had learned the TLs, length of residence or instruction, and other languages spoken would all be highly desirable. With respect to the diversity of L1s, this limitation was somewhat mitigated by the regularities found in the distributions of L1s in the three TL groups (Chapter 4). On the one hand, the three groups of learners differed with respect to L1s and, more importantly, the linguistic distance between the L1 and the target language. For example, among learners of Czech there was a sizeable group of native speakers of Slavic languages, and among learners of Italian—of Romance languages, whereas among learners of German there was no similarly situated group. Despite these imbalances in the composition of the learner groups, the facilitation predicted by simple transfer was not traceable in the results. For learners of Italian and Czech (each with a group of speakers of closely related languages), one would predict higher familiarity with the roots, inflectional endings, and, more generally, the phonological properties of the target languages, resulting in fewer phonological errors. Conversely, learners of German—who are speakers of L1s not closely related to it—would be expected to make more phonological errors, due to the lack of such familiarity. However, it was the learners of Italian and Czech whose productions contained a higher number of phonological errors.

On the other hand, despite the imbalances in the specific learner L1s or L1 groups, the majority of those L1s belonged to what would be typically considered

inflectionally “moderately rich” to “rich” languages. In previous research (c.f., Murakami & Alexopoulou, 2015) the extent to which L1 properties were considered as potential influences on L2 morphological production was rather basic—limited to the presence or absence of the feature of interest, such as “Tense”, in the L1. Judged against this baseline, this study’s distribution of L1s, although not ideal, can be considered adequate, since the vast majority of L1s cleared the bar of having inflected forms to mark agreement on verbs.

I viewed these methodological decisions as part of a tradeoff between scale and descriptive detail, resolving it in the end in favor of achieving scale and capturing a high number of productions and individual learners. While scale took priority over rich description of learning contexts and histories in this study, future research on this topic might set different priorities.

The use of learner production data—as opposed to controlled elicitation techniques—makes it possible for learners to avoid using any features of the TL they have not fully acquired. This may be doubly true of productions captured in the context of a language proficiency examination. Even though the examinations in question were not high-stakes, one can presume that they still resulted in a fair amount of monitoring and attention to the formal aspects of TL use by learners. In this respect, it would be informative to test the findings of the present dissertation against learner performance on tasks of differing monitoring demands.

Finally, the cross-sectional nature of the data also entails a certain (and unknowable) amount of self-selection by learners into the TL groups. While I prioritized ecological validity when developing the data collection strategy, the possibility of



experimentally controlling assignment to the target languages had to be given up. Future research may resolve this tension differently and test the hypotheses that this dissertation's findings have spawned experimentally, rather than cross-sectionally. For example, artificial and semi-artificial language learning paradigms may prove instrumental to isolating some of the various influences on morphological learning that the present research has identified.

Other limitations were more conceptual in nature. With the exception of a few error types that have appeared in the literature, error categories were developed bottom-up from the data. A principled taxonomy that would draw on mental operations proposed under multiple approaches would merit a separate study of its own. In a perfect world, the error taxonomy would have been independently validated with respect to the mental processes it claims to reflect.

In particular, one should consider how psycholinguistic theories would fit into this space and what error types could be mapped onto some of the mental processes they posit. Even though I make a number of assertions as to what errors originate from faulty processing, I have not demonstrated a connection to processing data. Additional research could show whether, in fact, the errors that I have conceptually linked to processing warrant this interpretation. This would, however, amount to a multiple-experiment research program that would have well exceeded the feasibility of any single dissertation.

Whereas this study concentrated on only three target languages, they were conceived as representing continuous variation in paradigm richness but treated as levels of a factor variable for the purposes of the analyses. Future research might usefully approach this variation from the angle of quantitative typology: instead of treating TLs as

exemplars of “rich” and “poor” languages, *degrees* of richness/poverty would be related as continuous predictors to observed learning difficulty. Such treatment of typological differences as continuous has been successfully applied in studies on child language acquisition (Chapter 1) and allowed for more robust conclusions to be drawn.

This graded view of complexity should ideally apply both at the paradigmatic level, i.e. the number of distinct inflectional endings a language has, and to the number of verb classes in a language. It is possible that it is not only verb groupings designated as classes (by descriptive grammars) that meaningfully relate to learning, but also any other sub-regularities that are no longer productive in a language—such as, for example, *quasi* rules of the German ablaut (cf., English *catch – caught, swipe – swept, ride – rode – ridden*). Any number of patterned regularities among lexical items could be subject to statistical learning, and these need not be limited to those recognized in linguistic descriptions. While the current study took a simplified view of these matters, such nuances might be pursued in future research. This would amount to comparing the predictive power of different operationalizations of a TL’s complexity. Applied to the realm of verb classes, for instance, the traditional partition in German that posits a two-way grouping (“strong” and “weak”) would be pitted against one that recognizes further distinctions within the “strong” verbs based on the no longer productive ablaut system.

Extending this reasoning a step further, one may also conceive of complexity as a property of all of a language’s inflectional paradigms, and not merely the complexity of the domain related to a particular part of speech—such as, in the case of this study, verbal inflection. Thus, learning verbal morphology could be facilitated not only by the richness of the verbal paradigm itself but could also be affected by the variety in the adjectival and

nominal paradigms (e.g., gender, number, case agreement marking on nouns and adjectives). Sweeping examinations of this kind have not been pursued, to my knowledge. Limiting the study of morphological learning to just one part-of-speech domain has been more a matter of following established research procedures than an independently justified methodological choice. There is only one obvious reason to restrict an examination of morphological learning in this way: namely, if the purpose of a study is to focus strictly on abstract syntactic features, the domain has to be limited to one part of speech, or else the forms studied will have no features in common (e.g., I see *the neighbor [case]* vs. He *jogs [tense, person, number]* in the park every day). There is no *conceptual* reason to presume that morphological knowledge or processing are modular in this way. This issue, however, was not pursued in this dissertation and will have to be left up to future research to sort out.

Another obvious difficulty with cross-linguistic comparisons lies in the differences in the ways that languages realize morphological contrasts. The phonological factors influencing the production of inflectional morphology are the subject of a whole research literature. This study attempted to do it justice by including a separate investigation into the role of phonological learner processes in the production of L2 inflectional morphology in German (Chapter 6). Even though it revealed that the patterns of learner errors in German cannot be reduced to the phonological processes posited for the interlanguage, similar analyses were not carried out for the other target languages. To mitigate this limitation, an effort was made to describe the phonological properties of the TLs as they relate to permissible syllable structures and consonant clusters, along with the properties of learners' L1 represented in the sample (Chapter 3).

The results of this research cannot speak to some of the debates within the research literature. Due to the focus being squarely on inflectional morphology, this study does not speak to the issue of its relative developmental sequencing vis-à-vis the word order operations that are associated with morphology (Chapter 2). Similarly, the different dimensions of salience proposed under the general cognitive account have largely fallen outside the scope of this study, which concentrated, instead, on extending the GC account cross-linguistically.

Furthermore, it is impossible to know with any certainty just what the status of any inflected forms used by learners might be in their grammars and mental lexicons. While some psycholinguistic studies and data were noted briefly, these approaches were only marginally incorporated into the present study. Thus, any references made to the retrieval of inflected forms could apply in equal measure to the retrieval of inflectional elements (roots and endings), and this study is emphatically agnostic on this point.

Another limitation of this study is the lack of directly comparable English data, even though it was the literature on morphological L2 learning in English that motivated it and has remained a red thread and point of comparison. Considering how deeply rooted in English data much of theory building is when it comes to L2 morphological learning, the inclusion of English data would lend more force to this dissertation's arguments centering around the role of paradigm complexity in learning.

#### 7.4 Contributions

This dissertation combined a cross-linguistic approach with a focus on learner errors to yield data that allowed me to test two theoretical approaches simultaneously (syntactic, general-cognitive). In doing so, it supplied data from less commonly

researched and taught languages (Italian, Czech), which provided points of comparison to the more widely known patterns in L2 English. German, while not formally recognized as a “less commonly taught” language, is also not widely represented in the research on L2 and foreign language learning, especially in studies utilizing production data. Despite not being the central focus of the study, the fine-grained error data on these languages can inform future research and instructional materials design. More importantly, however, this study showcases the vital role that cross-linguistic comparisons have to play in the building of SLA theories meant to uncover the forces behind the learning of *any* language, as long as the TLs are carefully selected for their specific properties.

If the findings of this study stand the test of replication, they may inform foreign language instruction and the development of instructional materials. Exposing learners to a diverse range of inflected forms, or presenting only a subset of the diverse forms offered by the target language, or, conversely, inflating the diversity of forms learners encounter in input (beyond what would be natural in the language at large)—are only a few of the choices that this research opens up. Even though the present research did not manipulate input directly (as it was experienced by learners), its results speak to the possible superiority of input that is richer in inflected forms, at least as far as suppliance of inflection is concerned.

However, optimal courses of action may differ for different domains of knowledge: it may be that the input that promotes suppliance of *any* inflection (as opposed to omitting it or using an infinitive) is not the same kind of input that is beneficial for learning the intricacies of morphological alternations. If the two are indeed at odds, then their diverging developmental timelines will also need to be respected in

instruction. For example, early exposure to the *full* set of person-number combinations would be balanced against a *curated* approach to introducing lexical items belonging to different classes, or vice versa.

Further implications for instruction may arise from exploring whether the relationship between complexity and learnability is linear or whether any benefits of complexity taper off after a certain point. It is also possible that this point is specific to each learner and is determined not only by the properties of the target language but also by learner-internal factors and processing capacity.

In its data analysis, this dissertation employed not only Poisson regression modeling but also out-of-sample validation, which involved testing the models against data that had been withheld during model building. It is my hope that out-of-sample validation of statistical models will be adopted in future research on this topic and that this dissertation will offer a blueprint for doing so. The availability of learner corpus data may allow researchers to take advantage of this approach, in addition to the ongoing push in the field to step up replication efforts.

Conceptually, the present study adds to the literature on morphological learning its characterization of the “product to be learned” as the morphological paradigm, or system of oppositions among morphemes—and not a set of self-sufficient morphemes. This aspiration dictated the choice to track learners’ overall accuracy on inflection and not to limit the study to a particular morphological feature. This approach is closer aligned with structuralist views of language.

This study also follows the longstanding tradition of drawing inspiration from theoretical work on child language acquisition. The recent work on grammar competition

originating in that field has proven quite stimulating and applicable to L2 learning. Nevertheless, in the process of developing its extensions to L2 learning, I have discovered important modifications that it needs in order to account for L2 data, as well as its limits. The modifications are necessary due to the sheer variety of learner errors in L2: rather than focusing on suppliance versus omission, I examined the full gamut of departures from target-likeness. Some parts of this spectrum were better served by the general-cognitive model, as detailed above (Theoretical implications and takeaways)—such as the gradual, item-specific learning of morpholexical regularities and any alternations as they affect specific lexical items. However, other parts of non-targetlike production, such as overuse of inflection, are not easily explained by general-cognitive principles. If all inflections have positive difficulty “scores”, then supplying one should be more difficult than supplying a non-finite form. Similarly, it is not clear why learners would supply “bare” forms instead of providing a non-finite form (if it is the default)—considering that their experiential basis with bare forms may be zero, if such forms are not allowed in their target language.

As one of the main contributions of the study, I want to highlight its taxonomy of learner errors that I attempted to connect to different manifestations of learner knowledge. In doing so, I was forced to confront the layered meanings behind what it means to have “learned” inflectional morphology and to acknowledge the many facets of knowledge. While certain error types were deemed to be irrelevant to the purposes of this dissertation, it is nevertheless useful to be cognizant of the full range of learner behaviors in the face of morphological complexity, and other researchers may direct their attention to those parts of it that were outside of the scope of this study.

Despite the limitations (outlined above) related to the process of developing the error taxonomy, I count among this dissertation's contributions the attempt to examine different aspects of learner performance on the same task, rather than gathering evidence from multiple tasks. Whether the typology of errors is validated in the exact form it was proposed in this dissertation or undergoes refinement, the overall approach of testing the predictions originating from different theories on the same data merits further application. After all, in the words of George E. P. Box, all models are wrong, but some are useful.



## References

- Aksu-Koç, A., Ketrez, F. N., Laalo, K., & Pfeifer, B. (2007). Agglutinating languages: Turkish, Finnish, and Yucatec Matya. In S. Laaha & S. Gillis (Eds.), *Typological perspectives on the acquisition of noun and verb morphology: Antwerp Papers in Linguistics 112*, (pp. 47–57). Antwerp: University of Antwerp.
- Amaral, L., & Roeper, T. (2014). Multiple grammars and second language representation. *Second Language Research*, 30(1), 3-36.
- Anderson, J. (1987). The markedness differential hypothesis and syllable structure difficulty. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system*. New York, NY: Harper and Row. 279-291.
- Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.
- Bayley, R. (1996). Competing constraints on variation in the speech of adult Chinese learners of English. in R. Bayley & D. R. Preston (Eds.), *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins. pp. 97-120.
- Bley-Vroman, R. (1997, October). *Features and patterns in foreign language learning*. Plenary talk presented at the Second Language Research Forum. Michigan State University.
- Bonilla, C. L. (2015). From number agreement to the subjunctive: Evidence for Processability Theory in L2 Spanish. *Second Language Research*, 31(1), 53-74.

- Bowden, H. W., Gelfand, M. P., Sanz, C., & Ullman, M. T. (2010). Verbal inflectional morphology in L1 and L2 Spanish: A frequency effects study examining storage versus composition. *Language Learning*, *60*(1), 44-87. doi: 10.1111/j.1467-9922.2009.00551.x
- Brunh de Garavito, J. (2003). The (Dis)association between morphology and Syntax: The case of L2 Spanish. In S. Montrul & F. Ordoñez (Eds.), *Linguistic Theory and Language Development in Hispanic Languages* (pp. 398-417). Cambridge, MA: Cascadilla Press.
- Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language*, *58*, 265-289.
- Chew, P. (2003). *A computational phonology of Russian*. Universal-Publishers.
- Clahsen, H. (1988). Parameterized grammatical theory and language acquisition. In S. Flynn & W. O'Neil (Eds.), *Linguistic theory in second language acquisition* (pp. 47-75). Dordrecht: Kluwer.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3-42. doi:10.1017/S0142716406060024.
- Coughlin, C. E., & Tremblay, A. (2015). Morphological decomposition in native and non-native French speakers. *Bilingualism: Language and Cognition*, *18*(3), 524-542.
- Davidson, L., & Roon, K. (2008). Durational correlates for differentiating consonant sequences in Russian. *Journal of the International Phonetic Association*, *38*(2), 137-165.

- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(04), 499-533.
- DeKeyser, R. M., Alfi-Shabtay, I., Ravid, D., & Shi, M. (2017). The role of salience in the acquisition of Hebrew as a second language: interaction with age of acquisition. In S. Gass, P. Spinner, & J. Behney (Eds.), *Salience and SLA* (pp. 131-146). London: Routledge.
- DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, 31(3), 413-438.
- Di Biase, B., & Kawaguchi, S. (2002). Exploring the typological plausibility of Processability Theory: Language development in Italian second language and Japanese second language. *Second Language Research*, 18(3), 274-302.
- Dressler, W. U., Stephany, U., Aksu-Koç, A., & Gillis, S. (2007). Discussion and conclusion. *Typological Perspectives on the Acquisition of Noun and Verb Morphology. Antwerp Papers in Linguistics*, 112, 67-71.
- Dryer, M.S. (2013). Prefixing vs. suffixing in inflectional morphology. In M.S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/26>, Accessed on 2017-04-26.)
- Dyson, B. (2009). Processability Theory and the role of morphology in English as a second language development: a longitudinal study. *Second Language Research*, 25(3), 355-376.

- Eckman, F. (1986). The reduction of word-final consonant clusters in interlanguage. In J. Leather & A. James (Eds.), *Sound patterns in second language acquisition*, (pp. 143-162). Dordrecht: Foris.
- Ellis, R. (2015). Researching acquisition sequences: Idealization and de-idealization in SLA. *Language Learning*, 65(1), 181-209.
- Eubank, L. (1994). Optionality and the initial state in L2 development. In Hoekstra, T., & Schwartz, B. (Eds.), *Language acquisition studies in generative grammar*. Amsterdam; Philadelphia: John Benjamins. 369-388.
- Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44), 17897-17902.
- Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, 27(3), 339-374.
- Foote, R. (2015). The storage and processing of morphologically complex words in L2 Spanish. *Studies in Second Language Acquisition*, 1-33. <http://dx.doi.org/10.1017/S0272263115000376>
- Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17(3), 213-247.
- Frank, V. (2000). Impact of in-country study on language ability: National Security Education Program undergraduate scholarship and graduate fellowship recipients. Technical Report, The National Foreign Language Center.

- Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, *90*(1), 58-89. [doi: 10.1353/lan.2014.0013](https://doi.org/10.1353/lan.2014.0013)
- Geyken, A. (2007). The DWDS corpus: A reference corpus for the German language of the 20th century. In Fellbaum, Ch. (Ed.), *Collocations and Idioms: Linguistic, lexicographic, and computational aspects*. London, UK. pp. 23–41.
- Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “natural order of 12 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning*, *51*(1), 1-50.
- Gor, K., & Chernigovskaya, T. (2005). Formal instruction and the acquisition of verbal morphology. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 131–164). Berlin: Mouton De Gruyter.
- Gor, K., & Jackson, S. (2013). Morphological decomposition and lexical access in a native and second language: A nesting doll effect. *Language and Cognitive Processes*, *28*(7), 1065-1091.
- Gregg, K. (1996). The logical and developmental problems of second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition*. London: Academic Press.
- Gregová, R. (2015). The CVX theory of syllable: The analysis of word-final rhymes in English and in Slovak. In *Belgrade English language & literature studies*, Vol. III (pp. 111-127).
- Grijzenhout, J., & Joppen, S. (1998). First steps in the acquisition of German phonology: A case study. In *Theory des Lexikons; Arbeiten Sonderforschungsbereichs 282*, Nr. 110.

- Grodzinsky, Y. (1984). The syntactic characterization of agrammatism. *Cognition*, 16(2), 99-120.
- Halle, M. (1959). *The sound pattern of Russian: a linguistic and acoustical investigation*. Gravenhage: Mouton.
- Hamdi, R., Ghazali, S., & Barkat-Defradas, M. (2005). Syllable structure in spoken Arabic: A comparative investigation. In *INTERSPEECH-2005* (pp. 2245-2248).
- Hawkins, R. (2001). *Second language syntax: A generative introduction*. Wiley-Blackwell.
- Haznedar, B., & Schwartz, B. D. (1997). Are there optional infinitives in child L2 acquisition? In *Proceedings of the 2<sup>nd</sup> annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press (pp. 291-302).
- Herschensohn, J. (2001). Missing inflection in second language French: Accidental infinitives and other verbal deficits. *Second Language Research*, 17(3), 273-305.
- Hilbe, J.M. (2016). *COUNT: Functions, Data and Code for Count Data*. R package version 1.3.4. <https://CRAN.R-project.org/package=COUNT>
- Hoover, J. R., Storkel, H. L., & Rice, M. L. (2012). The interface between neighborhood density and optional infinitives: Normal development and specific language impairment. *Journal of Child Language*, 39(4), 835-862.
- Institut für Deutsche Sprache (2017). Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2017-I (Release vom 08.03.2017). Mannheim: Institut für Deutsche Sprache. PID: 10932/00-0373-23CD-C58F-FF01-3.

- Ionin, T., & Wexler, K. (2002). Why is 'is' easier than '-s'? Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research*, 18(2), 95-136.
- Janda, L., & Townsend, C. (2002). *Czech*. Slavic and East European Language Research Center (SEELRC), Duke University.
- Jia, G., & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research*, 50(5), 1280-1299.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, 21(1), 60-99.
- Kehayia, E., Jarema, G., & Kądziaława, D. (1990). Cross-linguistic study of morphological errors in aphasia: Evidence from English, Greek, and Polish. In *Morphology, phonology, and aphasia*. New York, NY: Springer.
- Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, 20(4), 543-587.
- Kirkici, B., & Clahsen, H. (2013). Inflection and derivation in native and non-native language processing: Masked priming experiments on Turkish. *Bilingualism: Language and Cognition*, 16(4), 776-791.
- Kopřivová, M., Lukeš, D., Komrsková, Z., Poukarová, P., Waclawičová, M., Benešová, L., & Křen, M. (2017). ORAL: korpus neformální mluvené češtiny, verze 1 z 2. 6. Ústav Českého národního korpusu FF UK: Prague, Czech Republic. Accessed online at <http://www.korpus.cz>.

- Kornfilt, J. (2013). *Turkish*. NY: Routledge.
- Krause, T. & Zeldes, A. (2016). *ANNIS3: A new architecture for generic corpus query and visualization*. in: *Digital Scholarship in the Humanities*, 31. Retrieved from <http://dsh.oxfordjournals.org/content/31/1/118>
- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondříčka, P., Zasina, A. (2016). SYN2015: Representative corpus of contemporary written Czech. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož: ELRA. pp. 2522–2528.
- Laaha S., Gillis S. (Eds.) (2007). Typological perspectives on the acquisition of noun and verb morphology. *Antwerp Papers in Linguistics 112*. Antwerp, Belgium: University of Antwerp.
- Laaha, S., Gillis, S., Kilani-Schoch, M., Korecky-Kröll, K., Xanthos, A., & Dressler, W. U. (2007). Weakly inflecting languages: French, Dutch, and German. *Typological perspectives on the acquisition of noun and verb morphology. Antwerp Papers in Linguistics, 112*, 21-33.
- Lardiere, D. (1998). Case and tense in the ‘fossilized’ steady state. *Second Language Research, 14(1)*, 1-26.
- Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research, 14(4)*, 359-375.
- Legate, J. A., & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition, 14(3)*, 315-344.



- Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6(3), 213-225.
- Lehtonen, M., Niska, H., Wande, E., Niemi, J., & Laine, M. (2006). Recognition of inflected words in a morphologically limited language: Frequency effects in monolinguals and bilinguals. *Journal of Psycholinguistic Research*, 35(2), 121-146.
- Leonard, L. B., Camarata, S. M., Brown, B., & Camarata, M. N. (2004). Tense and agreement in the speech of children with Specific Language Impairment: Patterns of generalization through intervention. *Journal of Speech, Language, and Hearing Research*, 47(6), 1363-1379.
- Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.3. <https://CRAN.R-project.org/package=emmeans>
- Long, M. H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly*, 24(4), 649-666.
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. *Foreign language research in cross-cultural perspective*, 2(1), 39-52.
- Luk, Z. P. S., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural -s, articles, and possessive 's. *Language Learning*, 59(4), 721-754.
- Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 36(3), 349-367.

- Lyding, V., Stemle, E., Borghetti, C., Brunello, M., Castagnoli, S., Dell'Orletta, F., Dittmann, H., Lenci, A., Pirrelli, V. (2014). The PAISÀ corpus of Italian web texts. In *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, Association for Computational Linguistics. Gothenburg, Sweden.
- MacWhinney, B. (1987). Applying the competition model to bilingualism. *Applied Psycholinguistics*, 8, 315-327.
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127-150. [doi: 10.1016/S0022-5371\(84\)90093-8](https://doi.org/10.1016/S0022-5371(84)90093-8)
- Maddieson, I. (2013). Syllable structure. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/12>, Accessed on 2017-04-26.)
- Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1), 1-71.
- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance*, 15(3), 576-585.
- Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3(2), 109-135.

- Meisel, J. M. (1997). The acquisition of the syntax of negation in French and German: Contrasting first and second language development. *Second Language Research, 13*(3), 227-263.
- Mezzano, G. G. (2003). *The development of Spanish verbal inflection in early stages of L2 acquisition*. Undergraduate honors thesis. University of Illinois, Urbana-Champaign.
- Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(6), 1146-1160.
- Miceli, G., Mazzucchi, A., Menn, L., & Goodglass, H. (1983). Contrasting cases of Italian agrammatic aphasia without comprehension disorder. *Brain and Language, 19*(1), 65-97.
- Mimouni, Z., & Jarema, G. (1997). Agrammatic aphasia in Arabic. *Aphasiology, 11*(2), 125-144.
- Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*(3), 503–553. <https://doi.org/10.1111/j.1467-9922.2008.00449.x>
- Morales, A. (2014). Production and comprehension of verb agreement morphology in Spanish and English child L2 learners: Evidence for the effects of morphological structure (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

- Murakami, A., & Alexopolou, T. (2015). L1 influence on the acquisition order of English grammatical morphemes. A learner corpus study. *Studies in Second Language Acquisition*. doi:10.1017/S0272263115000352
- Niemi, J., Laine, M., Hänninen, R., & Koivuselkä-Sallinen, P. (1990). Agrammatism in Finnish: Two case studies. *Agrammatic aphasia: A cross-language narrative sourcebook*, 2, 1013-1085.
- Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, 22, 158-192.  
[doi:10.1080/10489223.2014.943903](https://doi.org/10.1080/10489223.2014.943903)
- Paradis, J., Rice, M. L., Crago, M., & Marquis, J. (2008). The acquisition of tense in English: Distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics*, 29(4), 689-722.
- Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.
- Phillips, C. (1995). Syntax at age two: Cross-linguistic differences. *MIT Working Papers in Linguistics*, 26, 325-382.
- Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Approaches to Bilingualism*, 5(4), 409-453.
- Pienemann, M. (2003). Language processing capacity. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition*. Oxford, UK: Blackwell Publishing Ltd. [doi: 10.1002/9780470756492.ch20](https://doi.org/10.1002/9780470756492.ch20)

- Pienemann, M. (2015). An outline of Processability Theory and its relationship to other approaches to SLA. *Language Learning*, 65(1), 123-151.  
DOI:10.1111/lang.12095
- Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of Processability Theory* (Vol. 30). John Benjamins Publishing.
- Pinker, S. (1984). *Language learnability and language learning*. Cambridge, MA: Harvard University Press.
- Portin, M., Lehtonen, M., & Laine, M. (2007). Processing of inflected nouns in late bilinguals. *Applied Psycholinguistics*, 28(1), 135-156.
- Poupier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory phonology*, 2(2), 243-273.
- Presson, N., Sagarra, N., MacWhinney, B., & Kowalski, J. (2013). Compositional production in Spanish second language conjugation. *Bilingualism: Language and Cognition*, 16(4), 808-828.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103-133.
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519-549. DOI: 10.1075/ijcl.13.4.06ray

- Rayson, P. and Garside, R. (2000). Comparing corpora using frequency profiling. In proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000). 1-8 October 2000, Hong Kong, 1-6.
- Rispoli, M., & Hadley, P. (2012). Input effects on the acquisition of finiteness. GALANA
- Rizzi, L. (1993). Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition*, 3(4), 371-393.
- Rossini Favretti, R., Tamburini, F., & De Santis, C. (2002). CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, pp. 27–38. Munich: Lincom-Europa.
- Sadowska, I. (2012). *Polish: a comprehensive grammar*. Routledge.
- Sánchez, L., Camacho, J., & Ulloa, J. E. (2010). Shipibo-Spanish: Differences in residual transfer at the syntax-morphology and the syntax-pragmatics interfaces. *Second Language Research*, 26(3), 329-354.
- Sato, C. J. (1984). Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning*, 34(4), 43-58.
- Saussure, F. de (1966). In Bally, C., & Sechehaye, A. (Eds.) *Course in general linguistics*. New York: McGraw-Hill.
- Šimáčková, Š., Podlipský, V. J., & Chládková, K. (2012). Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association*, 42(2), 225-232. [doi: 10.1017/S0025100312000102](https://doi.org/10.1017/S0025100312000102)

- Slobin, D. I. (Ed.). (1985). *The crosslinguistic study of language acquisition: Theoretical issues* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339-368.
- Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1-33.
- Stephany, U., Voeikova, M., Christofidou, A., Gagarina, N., Kovačević, M., Palmović, M., & Hržica, G. (2007). Early development of nominal and verbal morphology from a typological perspective-strongly inflecting languages: Russian, Croatian, and Greek. *Antwerp Papers in Linguistics*, 112, 35-47.
- Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3(1), 1-42.
- Timmermans, M., Schriefers, H., Dijkstra, T., & Haverkort, M. (2004). Disagreement on agreement: person agreement between coordinated subjects and verbs in Dutch and German. *Linguistics* 42(5), 905-930.
- Tkachenko, E., & Chernigovskaya, T. (2010). Input frequencies in processing of verbal morphology in L1 and L2: Evidence from Russian. In Gronn, A., & Marijanovic, I. (Eds.), *Russian in Contrast: Lexicon. Oslo Studies in Language* 2(2), 281-318.
- Törkenczy, M. (2004). *The Phonotactics of Hungarian*. Doctoral dissertation. Budapest: Hungarian Academy of Sciences.
- Tsapkini, K., Jarema, G., & Kehayia, E. (2001). Manifestations of morphological impairments in Greek aphasia: A case study. *Journal of Neurolinguistics*, 14(2), 281-296.

- Tsapkini, K., Jarema, G., & Kehayia, E. (2002). A morphological processing deficit in verbs but not in nouns: A case study in a highly inflected language. *Journal of Neurolinguistics*, *15*(3), 265-288.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231-270.  
[doi:10.1016/j.cognition.2003.10.008](https://doi.org/10.1016/j.cognition.2003.10.008)
- Vainikka, A., & Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research*, *12*(1), 7-39.
- VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.) *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Wei, T., & Simko, V. (2017). R package "corrplot": Visualization of a correlation matrix (Version 0.84). Available from <https://github.com/taiyun/corrplot>
- Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D. Lightfoot & N. Hornstein (Eds.) *Verb movement* (pp. 305-350). Cambridge, UK: Cambridge University Press.
- White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research*, *7*(2), 133-161.
- White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge, UK: Cambridge University Press.
- Wisniewski, K., Schöne, K., Nicolas, L., Vettori, C., Boyd, A., Meurers, D., Abel, A., & Hana, J. (2013). MERLIN: An online trilingual learner corpus empirically



grounding the European Reference levels in authentic learner data. In: ICT for Language Learning, Conference Proceedings. Libreriauniversitaria.it Edizioni. Retrieved from [http://conference.pixel-online.net/ICT4LL2013/common/download/Paper\\_pdf/322-CEF03-FP-Wisniewski-ICT2013.pdf](http://conference.pixel-online.net/ICT4LL2013/common/download/Paper_pdf/322-CEF03-FP-Wisniewski-ICT2013.pdf)

Yang, C. D. (2002). *Knowledge and learning in natural language*. Oxford, UK: Oxford University Press.

Yuan, B. (2001). The status of thematic verbs in the second language acquisition of Chinese: against inevitability of thematic-verb raising in second language acquisition. *Second Language Research*, 17(3), 248-272.

Zobl, H., & Liceras, J. (1994). Functional categories and acquisition orders. *Language Learning*, 44(1), 159-180.