LANGUAGE ACQUISITION IN THE PAST

Jordan Kodner

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

Charles Yang, Professor of Linguistics

Supervisor of Dissertation

Mitchell P. Marcus, RCA Professor of Artificial Intelligence

Graduate Group Chairperson

Eugene Buckley, Associate Professor of Linguistics

Dissertation Committee

Donald Ringe, Professor of Linguistics

George Walkden, Professor of Linguistics, Universität Konstanz

LANGUAGE ACQUISITION IN THE PAST

Dedicated to my father

ACKNOWLEDGEMENT

I'm not very good at waxing poetically, so I hope my sincerity comes across despite the brevity of this section. There are several people who I wish to thank for their efforts and support over the last five years in the Departments of Linguistics and Computer and Information Science at the University of Pennsylvania. First among them are my advisors, Charles Yang and Mitch Marcus, who have done more to help shape who I am as a researcher than anyone else. This dissertation would have been much worse without them. This work has also benefited greatly from discussions with my committee members Don Ringe and George Walkden, Tony Kroch, William Labov, Julie Anne Legate, Rolf Noyer, Meredith Tamminga, and my other instructors and mentors during my time at Penn.

I have had several co-authors over the last few years, all of whom deserve mention (in order of first co-publication): Spencer Caplan, Mitch Marcus, Hongzhi Xu, Charles Yang, Chris Cerezo Falco, Dan Roth, Shyam Upadhyay, Caitlin Richter, Kathryn Schuler, Jasmine Lee, Justin Mott, Ann Bies, Stephanie Strassel, and Nitish Gupta, and hopefully several more to come. Caplan, Marcus, Xu, and Yang stand out as particularly close collaborators.

Community has also played a major part in my life here. I have to single out my classmates in the cohort of 2015, the open-minded NLP group, and our licorice-fueled LORELEI team. In particular, I thank my comrade in StarLab Spencer Caplan, as well as all those who made my second stint at Penn far more manageable: Faruk Akkuş, Ryan Budnick, Andrea Ceolin, Ava Creemers, Dan Deutsch, Nitish Gupta, Marcella Hastings, Matthew Joseph, Stephen Mayhew, Ollie Sayeed, Shyam Upadhyay, Lacey Wade, Hongzhi Xu, and many more not listed here, as well as my persistent friends Anand Sundaram and Daniel Wei, who have helped me through Penn twice now.

Most importantly, I must thank my family, especially my mother, who taught me the value of honest work and trusted me to pursue my own interests, and my father and grandparents, who I am sure would have been proud to read this today. – Thank you.

ABSTRACT

LANGUAGE ACQUISITION IN THE PAST

Jordan Kodner

Charles Yang and Mitchell P. Marcus

There is a long tradition in linguistics implicating child language acquisition as a major driver of language change, the classic intuition being that innovations or "errors" which emerge during the acquisition process may occasionally propagate through through speech communities and accumulate as change over time. In order to better understand this relationship, I establish new methods for reasoning about language acquisition in the past. I demonstrate that certain aspects of child linguistic experience may be reasonably estimated from historical corpora and employ a quantitative model of productivity learning to investigate the role acquisition played as the driver of four well-documented instances of phonological, syntactic, and morphological change: transparent /aɪ/-raising in modern North American English, the innovation and lexical spread of the *to*-dative in Middle English, the analogy of the lengthened *$\bar{e}$-grade in Proto-Germanic strong verbs, and the forms of the past participles and *t*-deverbals in Classical and Late Latin. These case studies provide new insights into the implications of sparsity and variation on the first language acquisition process, the role that acquisition plays as the actuator of community-level change, and the complementary nature of acquisition and diachronic evidence for synchronic representation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1 : Introduction

Historical linguistics is fundamentally dynamic – languages undergo change in speech communities whose linguistic compositions are constantly in flux over time. Child language acquisition is similarly dynamic – individuals' internal languages develop as they mature from pre-verbal infants to adult speakers. Communities whose language changes over time are composed of learners who mature over time. This parallel has not gone unnoticed: it has long been held that the two are related, that development on the level of the individual over childhoods plays a causal role in development of languages over centuries (Paul, 1880; Sweet, 1899; Halle, 1962; Kiparsky, 1965; Andersen, 1973; Anderson, 1973; Baron, 1977; Lightfoot, 1979; Labov, 1989; Niyogi and Berwick, 1996; Kroch, 2001; Lightfoot, 1999; Yang, 2002; Kroch, 2005; van Gelderen, 2011; Yang, 2016; Cournane, 2017; Sneller et al., 2019, *inter alia*).

Language change is an inexorable process. Though it is often both reasonable and useful to approach language as a static snapshot (Chomsky, 1965), it is actually dynamic, constantly changing over time and varying person to person (Weinreich et al., 1968). We encode language as *grammars*, cognitive objects or systems which are capable of generating exactly the possible utterances of our languages (Chomsky, 1958), but we are not handed grammars by our communities, only samples of language generated by the grammars of our family, friends, classmates, and strangers. Language acquisition is no easy task for a child: the logical structure of language is far from trivial, and nobody ever receives enough input that they could acquire any meaningful competence through brute force and memorization alone. Nevertheless, children do acquire their native languages, and when they mature, they pass on samples of the languages generated by their grammar to younger people, continuing the cycle. This alternation from internal representation to expression and back is potentially error-prone.

The relationship between acquisition and change is formalized in Andersen's 1973 *Z-Model*

which views change as a cycle of error-prone abductive and inductive learning. Represented in Figure 1, some selection of utterances is output output by Grammar 1, and these samples become the input for subsequent speakers who learn a Grammar 2 of their own. This alternation, repeated again and again across time and populations, results in an accumulation of innovations measurable as language change through a process that explicitly privileges the acquisition process as the diagonal of the 'Z.' In subsequent terms, Andersen's model can also be interpreted as an alternation between *I-language* (the grammars) and *E-language* (the samples) (Chomsky, 1986), and provides a role for *competence* and *performance* (Chomsky, 1965), or representation, learning, and social factors in accounting for change.

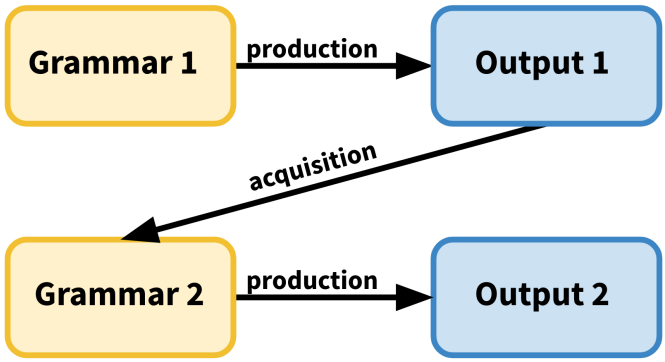


Figure 1: The Z-Model of language change (adapted from Andersen, 1973).

But this conception of acquisition change is not without issue. The Z-model encodes transmission as a simple generational chain, which aside from being formally insufficient (Niyogi and Berwick, 2009), is also empirically insufficient. Speakers are embedded in social networks (Milroy and Milroy, 1985), and transmission of language to children occurs in these (Labov, 2007). But regardless of their social position, children are truly excellent language learners. How can they be drivers of language change if they acquire language so faithfully? This is conflict is *the Paradox of Language Change* (Niyogi and Berwick, 1997). Lehmann (2013 §1.9.2) call on proponents of acquisition-driven change to prove their point by solving the Paradox, "*Advocates of change resulting from language acquisition must demonstrate*

when and how such patterns are adopted." That is what I intend to do.

I take an approach that explains language change as the direct consequence of, and in terms of, the acquisition process. As a goal, an acquisition-driven account should demonstrate how the process by which children acquire some piece of grammar in their native languages yields both the initial innovation and the eventual spread of some change over time. The most direct insights that this provides, then, are on the dynamics of change rather than the formal realization of what has changed. However, it also serves as an alternate line of evidence to motivate formal representations and to clarify points of the acquisition process. A suitable problem for this approach is one which has enough historical data behind it to track its diachronic trajectory and a large enough body of literature from the field of child language acquisition to externally motivate an account of the core learning processes that drives it. Under this view, serious engagement with the acquisition process means a deeper understanding of the drivers of language change in general, and the application of language learning algorithms to specific problems contributes to more rigorous motivation for theoretical analyses.

1.1. Overview of the Thesis

This work consists of a discussion of several aspects of the relationship between language acquisition and change laid out in five studies (Figure 2). Following this introduction and a background discussion in Chapter 2 on language change and language acquisition, I introduce the use of corpora to estimate child linguistic experience in Chapter 3 and explore the feasibility of modeling pre-modern learners with pre-modern data sources. I demonstrate that corpora regardless of genre remain consistent in certain aspects of lexical distribution such that they can be used to model the acquisition of lexical patterns. Chapters 4-7 present four application studies which propose models of acquisition-driven change to account for phonological, morphological, and syntactic phenomena in Modern English, Middle and Early Modern English, Proto-Germanic, and Classical Latin. Together, these build up to a framework for acquisition-driven change which improves upon the classic Z-model (Section 6.6)

and provide a pathway for better understanding synchronic theoretical phenomena (Section 7.7). Chapter 8 ends with a summary discussion.

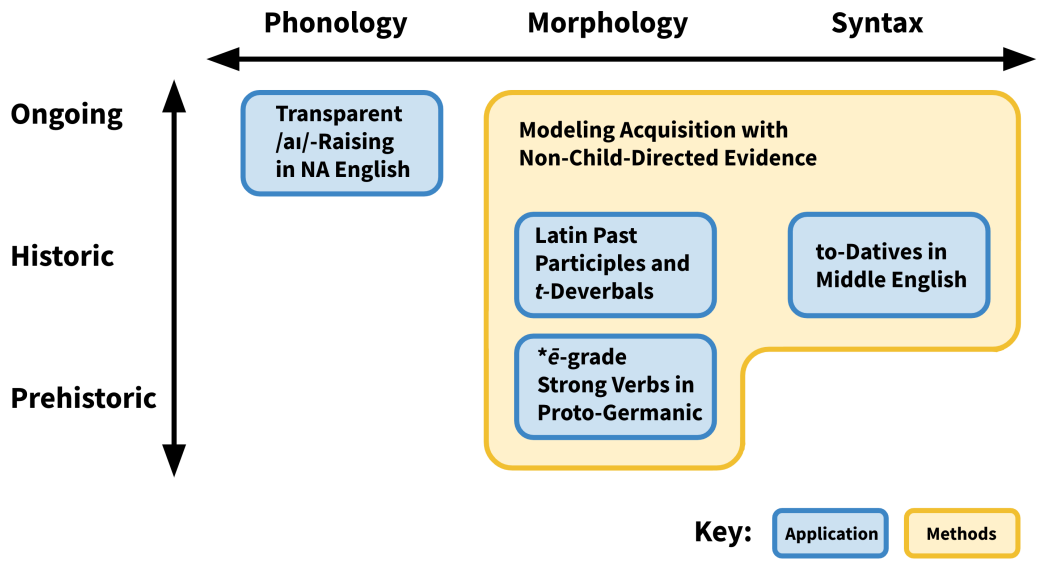


Figure 2: The five dissertation studies organized by time depth and level of the grammar explored.

I begin in Chapter 3 with an analysis of historical text corpora, generally the only direct evidence we have for language as it was in the past. I demonstrate the close relationship between these corpora and those of child-directed speech (CDS) which are often used to estimate child linguistic experience. CDS and non-CDS corpora prove to be quantitatively similar to and often statistically indistinguishable from one another along a number of linguistic dimensions when the pre-processing steps typical of the acquisition literature are applied, therefore they may reasonably be substituted for modern CDS corpora when reasoning about the acquisition process, and we can model how acquisition occurred in the past in the same ways that we do today.

The first case study laid out in Chapter 4 investigates "partial" or "transparent" /aɪ/-raising, a sparsely attested phonological pattern similar to classic /aɪ/-raising or *Canadian raising* of front diphthongs which unusually only raises before surface voiceless segments (as in 'write,'

but not 'writer') (Joos, 1942; Fruehwald, 2016; Berkson et al., 2017). Building on a collaboration with Caitlin Richter (Kodner and Richter, 2020), I introduce a learning model for phonological rule generalization which reconciles the existence of transparent raising speakers with the evidence that transparent raising need not be seen as incipient canonical raising. Noting that transparent raising has only been attested near raising dialect boundaries, we model its development as a kind of contact effect between two minimally distinct varieties of North American English. We propose that a fraction of learners exposed to both canonical raising and non-raising input will innovate transparent raising. Furthermore, population mixtures of canonical, transparent, and non-raising input rapidly transition away from transparent raising, accounting for the lack of transparent raising populations. More broadly, this study emphasizes the observation that variation is a normal part of the acquisition process even in monolingual settings and serves as a demonstration of corpus-based computational modeling of the acquisition process. It casts learners as the innovators of novel grammars which are not attested in their input.

The second case study in Chapter 5 accounts for the actuation and development of the *to*-dative (*Alice gave the book to Bob*) over the course of Middle and Early Modern English. The construction was innovated either in Late Old or Early Modern English then rapidly achieved a wide semantic range (Allen, 1995; Elter, 2018). I challenge existing morphological erosion and borrowing accounts for this change and introduce a two-part acquisition-driven analysis which better accounts for its diachronic development and modern typological distribution within Germanic. First, I present naive learner analyses of superficially recipient-like goal constructions, a process which is consistent with observed modern child errors, as the *to*-dative's ultimate source. This would amount to an allative to recipient shift of the type frequently attested across the world's languages (Cuyckens and Verspoor, 1998). Next, I apply the Tolerance Principle, a cognitively-motivated quantitative model of productivity learning (Yang, 2016), over a 'typical' Middle English learner's lexicon extracted from the Penn Parsed Corpus of Middle English 2 (Kroch and Taylor, 2000). I show that given this lexicon, a cohort of learners could rapidly generalize the new construction from those am-

biguous constructions to its attested Middle English distribution while remaining consistent with the attested chronology of the change. Taken together, these two processes account for the actuation, both the innovation and initial social expansion, of a pattern of English syntax.

The third application in Chapter 6 develops a model of analogical change cast in terms of productivity learning to explain a morphological change that occurred in the prehistory of Germanic. The reconstructed ancestor of the family, Proto-Germanic, had a system of *strong verbs* (as in Modern Germanic, e.g., English *sing* ∼ *sang* ∼ *sung*) which indicated inflection partly by stem changes. Most of the strong verb stem alternations can be accounted for by regular sound change from Proto-Indo-European except for the *lengthened *ē-grade* of the Class IV and Class V past stems. Modern accounts for the lengthened *ē-grade attribute it to analogy from a debated source to Class V verbs then from Class V to Class IV. The productivity-based model for analogical change frames both the innovation and initial social expansion of the change in terms of input sparsity. This model finds tentative support for analogy for one leading hypothesis: that analogy occurred from **etanǫ* 'eat' and rhyming Class V verbs to rest of the class, over an alternative analogy from Class VI to Class V.

The final case study in Chapter 7 explores the synchronic and diachronic implications of productivity learning and analogical change on the Classical Latin past participle and so-called *t-deverbals* which share its stem. As is well known, the Classical past participle is for the most part not reliably predictable from the forms of the present or perfect stem (Lieber, 1980; Aronoff, 1994; Laurent, 1999). Applying the Tolerance Principle to a typical Classical Latin lexicon extracted from the Perseus corpus (Smith et al., 2000) reveals the scope of the problem: most past participles outside of the first and fourth conjugations would need to be memorized by a learner as they cannot be derived by productive generalizations. Additionally, generalization from the fourth conjugation is only feasible if the speaker is permitted to reference the form of the corresponding perfect during the derivation. These observations have implications for existing theoretical treatments of the Latin verb and *t*-deverbals (e.g.,

Aronoff, 1994; Embick, 2000; Steriade, 2016). Further, in connecting productivity learning with analogical change, it provides a first ever principled account for major developments in the Late Latin and Proto-Romance past participle, including the rise of reflexes of *-ūtus*.

Together, the five studies lay the groundwork for a research program in language change from the perspective of child learners and establish some foundational principles. First, the triple roles of learning, representation and diachronic chance in explanation: all three factors contribute to the realization of language as we see it. Balancing their explanatory power allows us to develop simpler theories while also accounting for facts not otherwise captured. Second, the utility of concrete predictive models: committing to and working through the predictions of a concrete model of acquisition can elucidate causes for change which may be otherwise counter-intuitive. Third, the challenge of acquisition: a quantitative view of the data underscores the very real challenge that learners face, both due to sparsity and variability. Language acquisition is tractable, but not trivial. And fourth, a unified quantitative approach: it establishes a quantitative approach to historical data that borrows heavily from work in acquisition and focuses on the similarities between expressions of language, child and adult, modern and historical, rather than the differences. This provides a uniformitarian window into the language acquisition in the past.

CHAPTER 2 : Language Acquisition and Change

Since the earliest days of scientific linguistics, it has been held that children are responsible for language change (Sweet, 1899; Halle, 1962; Kiparsky, 1965; Anderson, 1973, 1977; Baron, 1977; Lightfoot, 1979; Niyogi and Berwick, 1996; Kroch, 2001; Yang, 2002; Cournane, 2017, *inter alia*). Children have been positioned in several roles, as innovators of new grammatical structures (Lightfoot, 1979), and as mere modifiers and re-organizers of existing patterns (*imperfect learning;* Kiparsky, 1965). Furthermore, the centrality of children in these research programs vary immensely, from the subject of experimental studies (Cournane, 2017) to the subject of computational study (Niyogi and Berwick, 2009; Yang, 2002) in conjunction with sociolinguistic insights (Sneller et al., 2019; Kodner and Richter, 2020), from critical primary drivers (Andersen, 1973; Baron, 1977; Lightfoot, 1979) to more abstract entities (Kroch, 2001).

Parallel to these, there exists a body of work that focuses on community interaction as the locus for change. Variationist sociolinguists study the key role that social factors play in both synchronic and diachronic variation (Weinreich et al., 1968; Eckert, 1989; Labov, 1994, etc.). They emphasize the importance of community-level thinking in change (Labov, 1963), the social meaning of linguistic variation (Eckert, 1989), and the role that community networks play in the propagation of variation (Milroy and Milroy, 1985). Usage-based linguists also focus on communication and interactions in the shaping of language (Bybee, 1985; Croft, 2000; Bybee, 2010).

The acquisition and social lines have often been cast as opposed to one another (see Dahl, 2004; Diessel, 2012; Stanford, 2015, etc.), and it is clear that proponents often conceive of their work in that way. And it is entirely possible, it is argued, that children play an active role in change as propagators without them being innovators (Croft, 2000; Dahl, 2004). That said, they are not necessarily always in conflict. Even young children show signs of sociolinguistic competency (Labov, 1989; Roberts, 1994) and can take part in community

interactions while they are still acquiring their native languages. For Labov (2007), *transmission* of language to young learners from members of the community is as important as *diffusion* among members of the community in accounting for variation and change. Even Weinreich et al. (1968), the founding document of variationist sociolinguistics, argues in favor of children as innovators while simultaneously stressing the importance of the community (§2.41).

A full review of over a century of work in language acquisition and change is out of scope for the current work, but I will touch on several key concepts. Section 2.1 clarifies what sorts of things about language change. I draw an informal distinction between "discrete" changes, which reflect categorical features of the grammar, and "continuous" changes, which capture probabilistic changes in usage or production. Section 2.2 discusses the *Paradox of Language Change*, the observation that change happens despite native learners' famous propensity for accurate acquisition. Two features of the acquisition process, data sparsity and variation, may drive learners towards acquiring patterns which are not reflected in their communities. Many of the criticisms levied against acquisition-driven accounts of change center around the Paradox, so by resolving it, we can lay many of them to rest. Finally, Section 2.3 introduces and explains the Tolerance and Sufficiency principles, which will be implemented as a concrete acquisition model in each of this work's case studies.

2.1. What Changes about Language

One important thing to keep in mind about language change is that it is really a cluster of several phenomena, not all of which are closely related. Properties of a language can be altered at every level of the grammar, may or may not vary over individuals' lifetimes, and may or may not rapidly fix in a community. The social aspects of language also change, affecting social meaning within communities, but also how language is used across communities. Changes may be endogenous, emerging from within a community as a result of language acquisition or other processes, or they may be exogenous, through contact and borrowing. There can be no explanatory silver bullet for all change.

It is often far from obvious whether a given change or set of changes is endogenous or exogenous – taking the history of English syntax as an example, even basic phenomena have been attributed both to endogenous and exogenous sources, from phrasal verbs, preposition stranding, spit infinitives, and details of preposition sluicing, among many others (Emonds and Faarlund (2014) vs. Bech and Walkden (2016)), or the syntax and semantics of ditransitives (Trips and Stein (2008) and Elter (2018) vs. Allen (1995) and Bacovcin (2017)) – so it has to be argued rather than assumed. Acquisition-driven innovation is purely endogenous when it is "monolingual" (Section 4.5). And when acquisition intersects with input from language contact, a change may be both endogenous and exogenous.

In this work, I will explore primarily endogenous acquisition pathways for phonological, morphological, and syntactic changes, and rather than distinguishing between levels of the grammar, I find it more useful to make an orthogonal distinction between "discrete" and "continuous" change. To a first order approximation, I mean for "discrete" changes to capture categorical changes in the grammar such as the innovation of a new phonological pattern that was not present previously (Chapter 4), a novel syntactic construction (Chapter 5), or new morphological forms (Chapters 6-7). These are properties of the grammar that are held virtually fixed over the lifetime of an individual, and are standard objects of study in child developmental research (Berko, 1958; Newport et al., 1977; Werker and Tees, 1984; Pinker et al., 1987; Naigles, 2002; Hudson Kam and Newport, 2005, etc).

"Continuous" changes, on the other hand, refers to measurable changes in articulatory and acoustic phonetics, usage rates of different constructions, or distributional aspects of semantics, all of which features of languages that can be quantified with continuous variables. Continuous changes are the most iconic objects of study in sociolinguistics. While young children acquire these properties as well as some understanding of community-level variation (Labov, 1989), they may vary considerably over individuals' lifetimes for extra-linguistic reasons. Continuous changes may be effected by the acquisition process (Yang, 2002), but are also usually what is discussed in relation to the *adolescent peak* (Eckert, 1989, *inter alia*)

among sociolinguists.

Nycz (2013) provides an insightful illustration of this distinction. The study followed Canadian immigrants living in New York City and tracked how well they accommodated to New York's low-back vowel pattern. Canadian speakers have the COT-CAUGHT merger, while New Yorkers make a distinction. Nycz found that individual Canadians accommodated in the position of their vowel to varying degrees, producing items for which New Yorkers have the COT vowel closer to New York COT and items for which New Yorkers have the CAUGHT vowel closer to New York CAUGHT, nevertheless closer inspection revealed that they did not learn to reliably recognize the distinction. In other words, the adult speakers were capable of alternating their (continuous) phonetic realizations, but failed to "undo" their underlying merger, which would have involved a (discrete) change to their phonemic inventories. This study highlights one of the ways in which discrete and continuous change are interrelated.

Discrete properties of the grammar are reflected in continuous variables. This plays out on a population level as well because once a change enters the population, it becomes variation on the community level (Weinreich et al., 1968). Sankoff (1988) goes so far as to define variationism as concerning the distribution of discrete choices, and research programs whose primary interest lies in discrete properties of the grammar may focus on continuous changes as objects of study. This is the hallmark of Kroch's school of diachronic corpus studies of syntactic change which investigate usage frequencies over time in order to uncover insights into discrete properties of the grammar (e.g., Kroch, 1989; Pintzuk, 1991; Taylor, 1994; McFadden, 2002; Bacovcin, 2017).

Since discrete changes are propagated as continuous change, the interesting questions regarding discrete changes lie closer to actuation than to propagation. *Actuation*, under the Labov et al. (1972) definition, refers to both the initial innovation of a change and its uptake in the immediate speech community. I seek to develop solutions to both of these components in this dissertation's case studies. It is, of course, impossible to enumerate the exact conditions, both internal and external, that an individual was subject to at the moment that

they innovated a given change. This is the famous *Actuation Problem* (Weinreich et al., 1968, §3.4). All changes, whether ongoing today or completed in the past, were necessarily innovated at some point before researchers took notice. Nevertheless, we can work out what sorts of conditions would lead to innovation and initial propagation and edge asymptotically closer to the actuation point, which achieves our goal of better understanding language (Walkden, 2017).

2.2. The Paradox of Language Change

The actuation of a change requires both its innovation and its entry into a speech community. Children are famous for the impressive accuracy in acquisition, so one would expect them to grow out of any novel patterns which they innovate. To be sure, children usually do grow out of their innovations, but even if they did not, what kind of community would pick up those innovations from a small child? After all, "*babies do not form influential social groups*" (Aitchison, 2001).

This is the *Paradox of Language Change* (Niyogi and Berwick, 1997). Like all good paradoxes, it challenges our intuitions about the problem: One conception of acquisition-driven change is to "blame the learner," that is, that is to assume that learners sometimes introduce errors into the acquisition process despite receiving enough evidence to make the right choices, whether this is due to a buggy learning algorithm or extra-linguistic cognitive or social pressures. I instead argue for a "blame the environment" approach, that is, to attribute learner innovations to properties of the input without discarding the assumption that learning is accurate. If children are learning from severely sparse evidence, from samples of language drawn from multiple sources, and from a web of variables, they may be forced into reasonable but novel grammars no matter how optimized their language learning faculties are. The very concept of "errors" makes little sense in a situation where there are multiple targets or evidence for the target is absent. Innovations, rather than errors, may be better thought of as faithful learning in the face of sparse or unusually sampled input (Kroch, 2001).

The common idealization of child language acquisition in theoretical literature, that the learner receives clean (albeit impoverished) input that is consistent with a single grammar, and that learner proceeds rapidly towards competence, are really both simplifying assumptions. They date back to the early days of generative linguistics as discussed explicitly in the *Sound Pattern of English* (SPE; Chomsky and Halle, 1968) as the authors lay out their approach to phonological theory,

"*[We] make the simplifying and counter-to-fact assumption that all of the primary linguistic data must be accounted for by the grammar and that all must be accepted as 'correct'; we do not here consider the question of deviation from grammaticalness, in its many diverse aspects... there is another, much more crucial, idealization implicit to this account. We have been describing acquisition of language as if it were an instantaneous process. Obviously, this is not true. A more realistic model of language acquisition would consider the order in which primary linguistic data are used by the child and the effects of preliminary 'hypotheses' developed in the earlier stages of learning on the interpretation of new, often more complex data.*" (SPE §8.1 p. 331)

SPE argues that relaxing these assumptions to account for the complexities of learning from realistic input was beyond the capabilities of the field in 1968. That was probably true, however while the fundamentals it lays out rightly remain relevant today, the field has progressed significantly both in terms of empirical understanding and technical know-how since then. From studies of communities, we know that children receive input from multiple people who themselves may exhibit internal variation, and from mathematical investigation, we know that change is formally *inevitable* in the face of even minimal variation (Niyogi and Berwick, 1997). Further, the effect of the Poverty of the Stimulus is not to be discounted, a point which has only been strengthened by large-scale corpora of child directed speech and the technology to process them. Even the richest innate specifications to the language faculty render acquisition *tractable, not trivial*. Not everything about language is equally easy or hard to acquire, and some properties of language may not actually be reliably learnable

when the input is just too sparse.

Taken together, there are two crucial observations about child language acquisition that drive change: first, that acquisition in the face of variation is the norm rather than the exception, that even minor variation in learner input can force change, and that transmission is a continuous rather than purely generational process, and second, that acquisition presents a real challenge, no matter how well-endowed the learner is because of how limited the input can be.

2.2.1. Acquisition in the Face of Variation

A common idealization in studies of child language acquisition is that there is a single, available, learning target, but the reality is far from this simple. There may be more than one target in the case of variation, either trivially due to the fact that no two parents maintain the same internal grammars or produce identical utterances, or dramatically in the case of full-blown language contact. Children must rectify mixed input as a normal part of the acquisition process, and this in itself renders change inevitable at some level. Acquisition errors presuppose the existence of a single available target to compare against, so the absence of such evidence draws that conception of innovations into question.

Children appear to leverage a few strategies in handling input variation. In response to syntactic variation, they may act as though they have acquired multiple grammars which they draw from probabilistically (Yang, 2002). This kind of *grammar competition* can unfold in a population quite rapidly or over several centuries (Kroch, 1989, 1994; Bacovcin, 2017). Young children, however, have also been shown to regularize mixed input rather than probability match and only transition from categorical learning to adult-like probability matching as they mature. This general pattern has been observed multiple times in laboratory settings: young children tend to regularize mixed input, and only later do they transition towards adult-like probability matching (Singleton and Newport, 2004; Hudson Kam and Newport, 2005; Schuler et al., 2016; Newport, 2019).

Early regularization does not imply a lack of social awareness. Even preschool age children are sensitive to ongoing linguistic variation, and their responses to it are not just simple probability matching either (Labov, 1989, 1994). For example, in the acquisition of stable variation for English "*g*-dropping" (*talking* vs. *talkin'*), it can be shown that children are aware of social conditioning on the variable even before they have mastered all of the grammatical conditions (Labov, 1989). While a follow-up by Roberts (1994) found that younger children showed apparent mastery of the grammatical constraints as well, it was not possible to distinguish their pattern from word-by-word memorization.

Not only are children aware of social variation, but they begin to attend to it by the time they reach school age. This crucially undermines the notion of purely generational transmission (also challenged in Manly, 1930; Weinreich et al., 1968; Roberts and Labov, 1995; Labov, 2001; Nardy et al., 2014, *inter alia*). While it is undeniably true that children, especially very young children, receive most of their early linguistic input from their caregivers, children also receive input from and learn from slightly older children (Weinreich et al., 1968, §2.41). Transmission is a continuous process from cohort to younger cohort as well as from generation to generation when viewed from the perspective of the community.

2.2.2. Tractable, not Trivial Learning

Language acquisition presents a real challenge no matter how rich the innate faculty of language – Universal Grammar renders language learning tractable in the face of the Poverty of the Stimulus (Chomsky, 1959, 1980), but it does not trivialize it. Acquisition takes time, some linguistic patterns are more challenging to acquire than others, and sometimes, sparsity may be so severe in early linguistic input that even the full endowment of UG, whatever that may be, is not enough to ensure that all learners converge on the same grammar. This *Abject Poverty* of the stimulus may even be more common than we think.

One observation is that children sometimes find it easier to pick up on forms than on meanings, and that the former are acquired before they are linked to the latter (Naigles, 2002).

The acquisition of argument structures for novel verbs is a case where this principle applies: Arunachalam and Waxman (2015) connects the diverse and often conflicting results of argument structure acquisition experiments to the difficulty of acquiring argument structure. Essentially, young learners are very finicky and require exactly the right conditions if they are to learn the argument structure of a novel verb, so slightly different experimental paradigms yield results consistent with opposing acquisition models. Furthermore, learner behavior does not necessarily indicate learner knowledge (Arunachalam, 2015), so it is possible for learners to get by with a naive and ultimately incorrect understanding of a verb's meaning in a way that is essentially undetectable day-to-day and even in experimental settings. Form over meaning in argument structure learning has been implicated in syntactic change as well: Anderson (1977) argues convincingly that split ergativity in Indic derives from a reinterpretation of constructions with passive participles as active perfectives. A learner who decided to pursue the superficial perfectivity of those constructions without altering their form would be forced to assume an ergative alignment.

Abject Poverty is the condition where input sparsity is so severe that speakers do not necessarily converge on the same grammar, which can occur when multiple plausible grammars are consistent with the input. Cases are challenging to identify because they require specific probes. If the surface expression of the different grammars were clear and easily accessible, it would not be a case of abject poverty, but if the grammars manifested no differences in their extensions, they would be impossible to distinguish. As such there may be more instances of failed convergence than we can currently observe. Nevertheless, some examples are worth discussing.

Taking example from phonetics, there are the multiple articulations of English /r/ (Baker et al., 2011). Speakers vary in how the phone is articulated. A baby cannot see inside an adult speaker's mouth, so if auditory cues are insufficient to uniquely motivate a particular articulation, they may adopt any one that is consistent. Moving to syntax, Han et al. (2007) find evidence for two populations of Korean speakers who differ in verb raising, but because

Korean is strongly head-final, these can only be differentiated by their judgements of certain uncommon negative scoping constructions. These constructions are rare in the input and those that are present could be generated by one or both grammars, so there is insufficient evidence to uniquely support one kind of raising.

In English morphology, there is evidence to support variability in the decomposition of so-called 'semi-weak' verbs. These are verbs whose past forms contain the regular coronal obstruent, but also employ some kind of stem change as in *tell* ∼ *told* or *sleep* ∼ *slept*. Conceptually, these could be decomposed into a root and the past suffix *tol-d*, or their forms might be learned like strong (*sing* ∼ *sang*) and suppletive verbs (*go* ∼ *went*). One way to investigate this is to look at their rate of "t/d-deletion," a phenomenon of English phonetics which varies according to grammatical context (Labov, 1994). In general, t/d-deletion occurs at a higher rate for mono-morphemes than for the past suffix, so semi-weak verbs are expected to delete at a rate consistent with their morphological decomposition. However, Guy and Boyd (1990) find some variation in deletion rates even among adult speakers, which suggests that most, *but crucially not all*, adults decompose these forms.

A closely related phenomenon which might also be classified as abject poverty is the situation where some part of the grammar goes unspecified, as is the case in paradigmatic gaps. Gaps appear in the inflectional paradigms of many languages, including the past participles for some speakers of English (e.g., *stride* ∼ *strode* ∼ *stridden*?, *strode?*, *strided*?), the genitive plurals of many Polish masculine nouns (Dąbrowska, 2001), some first person singular preterites in Russian (Yang, 2016, for a summary), and Latin defective nouns (*īnstar* nom./acc. sg. only, *vicis* no nom. sg.) and verbs (*inquam* most forms missing, *meminī* perfect stem only). Regardless of how the gap is accounted for cognitively, through a "no-default" approach, competition, or otherwise, these represent cases where the grammar fails to generate a final output for some morphological structure.

Corpus investigations into child-directed speech have begun to quantify the magnitude of input sparsity. Some complex syntactic constructions, such as parasitic gaps in English are

extremely rare or simply not present in the input (Pearl and Sprouse, 2013), so our judgements regarding them must fall out from our general syntactic competence. Constructions like parasitic gaps are fairly obtuse, but surprisingly common ones may be absent in the input as well. The famous *Baker's Paradox* (Baker, 1979) comes from the observation that most items that can participate in some syntactic structure, such as those employed by the English ditransitive constructions, are likely not to be attested in the input. A child then has to work out how to distinguish between verbs which are absent in the construction because they are ungrammatical from those which are absent due to sampling. A similar problem manifests through *paradigm saturation* in morphology (Chan, 2008), where it can be shown that the vast majority of items are attested in only a tiny fraction of their possible inflected forms, even in millions of words of input. In child-directed Spanish for example, the median verb only occurs in only a couple forms, leaving the learner to infer all of the rest.

Taking variation and sparsity together, we can begin to better characterize the learning problem. Borrowing some of the language of computational learning theory, acquisition is a particular class of learning problem in which the input sample and learning objective come with few guarantees: it is unknown to the child how many distributions the input is sampled from or how many distributions to try to fit, some of the input may itself be generated by other learners, and the sample is extremely skewed along several dimensions in ways which may or may not be known to the learner. Language acquisition is a massive undertaking. With such complex input and so few guarantees, any perturbations to the sample have the potential to alter learning outcomes. This is our way out of the Paradox of Language Change.

2.2.3. Some Criticisms of Acquisition-Driven Change

Several criticisms have been levied against acquisition-driven accounts of change over the years, many of which at least implicitly invoke the Paradox (e.g., Dahl, 2004; Lehmann, 2013, §1.9.2). In this section, I will briefly summarize some of the leading objections to the position.

First, it has been pointed out that not all changes reflect child-production errors and that not all child-production errors are reflected in change. For example, consonant harmony is common in children's early productions, but it is typologically rare among the languages of the world (Croft, 2000; Bybee, 2010; Diessel, 2012). However, the claim that acquisition drives change does not imply that *only* acquisition drives change, nor does it imply that all child productions should be reflected in change (Baron, 1977). As discussed earlier, language change is really a diverse set of phenomena, and there is no reason to think that all types of change are initiated in the same way. The uncertainty regarding which changes are initiated by children is an interesting research question, not a reason to dismiss concept. In this dissertation, I implicate child learners as innovators in phonological, morphological, and syntactic changes that can be associated with lexical effects in the input.

Second, it has been suggested that the fact that children often orient towards their peers rather than adults is a problem for acquisition-driven change (Aitchison, 2001; Bybee, 2010; Diessel, 2012), but this seems to be a misunderstanding of both sociolinguistic transmission (Labov, 2007) and of the Z-model. Who children learn from has no bearing on whether they are innovators, and there is nothing theoretically amiss about learning from multiple sources. Related to this is the criticism that children lack the social clout necessary to propagate a change (Aitchison, 2001; Bybee, 2010). But, once it is recognized that transmission is not strictly generational (e.g., Manly, 1930; Baron, 1977), the criticism falls flat. It is not adults that they need to impress, but rather, other children. The fact that children orient towards their peers is a strength rather than a weakness because it identifies exactly what kinds of social ties might lead to the early propagation of change. Children are influential to other children, particularly those just slightly younger than them, so a child who innovates a change may have time to transmit it before growing out of it (Section 6.6).

These children grow up together, so this conception of early propagation is not inconsistent with the observation that teens, not young children, often lead change (Labov, 1982; Eckert, 1989). This has been suggested as an objection as well (Diessel, 2012; Stanford, 2015), but

again, teens leading change does not imply that they are the innovators *per se* any more than children producing interesting errors implies that they are the innovators. Furthermore, the adolescent peak is a measure of continuous change, not discrete.

A conception of early social competence may actually explain why certain types of child productions, such as morphological leveling, have parallels in language change while others such as consonant harmony are much rarer. It has to do with the age at which these manifest: like Aitchison (2001) says, babies do not form influential social groups, so child productions that are expressed exclusively in the earliest productions when there is no younger cohort around to pick them up are not going to enter the community, while productions that continue into school age may be. Consonant harmony, which is not reflected in change, manifests in very early productions, while morphological leveling, which is abundant in change, is attested among school age children.

The Critical Period Hypothesis, the claim that children have a fundamentally greater ability to acquire language than older people, has been challenged as well (Aitchison, 2001, *inter alia*), but this requires a particular selective definition of "grammar" that excludes meaningful difference between native and second language acquisition (e.g., Johnson and Newport, 1989, for criticism). In reality, there are several differences between child and adult acquisition, even in the variationist literature, as discussed above in the summary of (Nycz, 2013). The Sankoff and Blondeau (2007) study of variation in the articulation of /r/ in Montreal French provides a second example. In Montreal, the conservative variant [r] has been replaced in the community by [ʁ]. The authors make two important points regarding the critical period. First, that speakers who were older during the study showed substantially less shift towards [ʁ] over time than younger ones, though a significant minority of older speakers did exhibit an increase in their rate of [ʁ] (Fig. 3, Tab. 12, §7.3). Second, those older speakers who did shift in their [ʁ] rate were already using [ʁ] by the time of the earliest recordings, which indicates that that variant was already present in their grammars. A change in usage rate is a continuous change that can vary over lifetimes rather than a

discrete change of the type narrowly learnable during the critical period.

Finally, even the fact that child over-regularizations of the kind implicated in morphological change are relatively rare has been proffered as an argument that they do not contribute to change (Diessel, 2012). But why is that a problem? One must remember that the development of the languages we see today has unfolded across millennia and over billions of language learners. Errors *should* be rare if children can communicate with their parents, and actuation events should be even rarer if change is not to happen overnight. This is a strength rather than a weakness.

2.3. Learning Patterns from Sparse Input

Children are excellent language learners, and despite their varied early linguistic experiences, there is great uniformity in what children eventually acquire (Labov, 1972). In their first few years of life, going off notoriously sparse evidence, they work out nearly all the rules and patterns of their native languages with astonishing accuracy. This uniformity implies both a shared language faculty across all people and commonalities in individuals' unique linguistic experiences. One of the major learning tasks that children face is determining the existence and scope of linguistic rules, patterns, or generalizations.

The question of whether a proposed pattern is compatible with the data is a non-trivial one which is constrained by the representational faculties of the speaker but is still crucially dependent on the data. To eschew this question is to make another simplifying assumption. To quote SPE again,

"*First, we develop a system of formal devices for expressing rules and a set of general conditions on how these rules are organized and how they apply. We postulate that only grammars meeting these conditions [of compatibility with the language faculty] are 'entertained as hypotheses' by the child who must acquire knowledge of a language. Secondly, we determine a procedure for evaluation that selects the highest valued of a set of hypotheses of the appropriate form...* ***We will not concern ourselves here with the nontrivial problem of***

what it means to say that a hypothesis–a proposed grammar–is compatible with the data*, but will restrict ourselves to two other problems, namely the specification of formal devices and of an evaluation procedure.*" (SPE §8.1 p. 331, emphasis mine).

For SPE, an "evaluation procedure" refers to the means by which a theoretician may rate the quality of an analysis rather than how a learner may do so. The two problems are complementary, and the latter has been taken up by researchers in child development. An overarching result from that field is relationship between *type*-frequency of linguistic patterns in the input, that is, how many items in the lexicon express a linguistic pattern, and generalization learning (Aronoff, 1976; MacWhinney, 1978; Bybee, 1985; Baayen, 1993; Elman, 1998; Pierrehumbert, 2003b; Yang, 2016). The relative uniformity of acquisition is partly explained by these models which rely on the counts of items exhibiting linguistic properties rather than the presence or absence of specific lexical items. The Tolerance Principle and its corollary, the Sufficiency Principle (Yang, 2016), are the most developed of these type-based productivity models and will be adopted for the remainder of this work.

The case studies in this dissertation are all variations on the theme of generalization learning in the context of *tractable, not trivial* learning. What about the acquisition of generalizations can cause language to change? In order to investigate this satisfactorily, we need to adopt some model of generalization learning.

2.3.1. The Tolerance and Sufficiency Principles

The *Tolerance Principle* (TP; Yang, 2016) is a type-based model which casts generalizations in terms of productivity in the face of exceptions. A distinction between productive "regular" patterns and unproductive "irregular" exceptions is prominent in many theories of grammar, and has often been conceptualized in terms of Elsewhere Conditions (Anderson, 1973; Kiparsky, 1973; Aronoff, 1976), in which irregulars have to be looked up in some way before defaulting to a regular pattern. By their very nature, irregulars have to be listed, so even regulars have to be checked against the set of the set of irregulars before they can

be inflected. Since it takes time to search down a list of exceptions before applying the elsewhere condition, an effect consistent with psycholinguistic findings in frequency-ranked lexical access (Murray and Forster, 2004), a cost is established for representing patterns as long lists of lexical exceptions as opposed to shorter lists with elsewhere conditions. The more irregulars, the longer the lookup, which motivates establishing productive patterns when possible.

A question then arises: how many exceptions is too many? At what point does it become better to list items and absorb the cost rather than form a productive generalization? This is where the TP comes in. It is fundamentally a mathematical heuristic, a quantitative *evaluation metric* (Chomsky, 1975, 1965). A child can hypothesize a generalization and then evaluate whether or not the number of exceptions to that generalization are too many to tolerate. If so, it can enter the grammar as a productive pattern.

It is a binary function: either a generalization is productive or it is not. All rules or generalizations are productive and all memorized lexical exceptions are non-productive by definition, so if a pattern is not productive, then it is not learned as a generalization, and vice-versa. It is up to each individual learner to determine whether patterns in the input are productive over a given domain or if they have to be memorized instead. For example, it is clear that English has a default past *-ed* rule, but one may wonder if there are other more specific rules available as well. Consider English *sing–sang–sung*, *ring–rang–rung*, *swim–swam–swum*, *drink–drank–drunk*, and so on. Is there a rule that says that verb stems shaped like *-iN(C)* form their past tenses in *-aN(C)*? If so, then there are a number of verbs that violate the rule: *sting–stung–stung*, *wing–winged–winged*, and *bring–brought–brought*, among others. If there is an *-aN(C)* past rule, then these violations have to be memorized as exceptions to that rule. On the other hand, if there is no rule, all these verbs become lexical exceptions to the default past rule. In worst-case scenarios, learners may not be able to form any generalizations at all under this model, which relates to paradigmatic gaps in abject poverty.

More formally, the *tolerance threshold* is calculated to be the number of exceptions below which it becomes more efficient to hypothesize a generalization than to list items.[1] In order to work this out, the TP assumes that lexical access is correlated with frequency rank (so high frequency irregulars are accessed before lower frequency ones) (Goodman et al., 2008), and items in the input follow long-tailed Zipfian frequency distributions (Zipf, 1949) in which few items are well-attested and others are rarely tested in the input. Zipfian and other similar long-tailed distributions are quite common throughout language (and indeed other domains as well) (e.g., Howes, 1968; Jelinek, 1997; Baroni, 2005; Yang, 2013).

Learners proceed by postulating tentative generalizations, and if the number of exceptions to those generalizations exceed the tolerance threshold for those generalizations, then they can resort to memorization. If the exceptions fall below the threshold, they can learn the generalization, handle the exceptions as needed, and consider a broader one to subsume it. The threshold θ_N itself is determined by the number of lexical items N in the scope of the generalization and the number of exceptions e, the number of words learned so far that fit the conditions for the generalization but do not exhibit the pattern (e.g., *wing*, *bring* for a hypothetical *sing*∼*sang* rule if the child has learned them). (1) provides a formal formulation of the Tolerance Principle, and Figure 3 provides a visualization of it.

(1) **Tolerance Principle**:

If R is a productive rule applicable to N candidates, then the following relation holds between N and e, the number of exceptions that could but do not follow R:

$$e \leq \theta_N \text{ where } \theta_N := \frac{N}{\ln N}$$

[1] See Yang (2016, pp. 10, 144) for the full mathematical derivation.

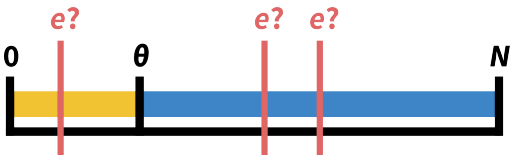
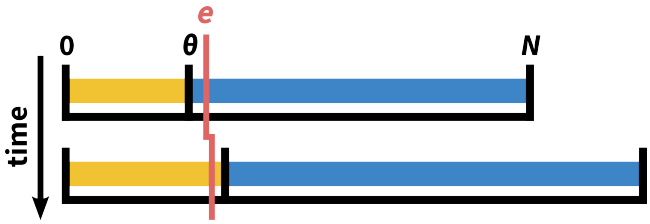


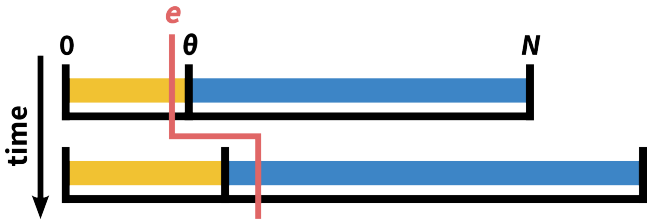
Figure 3: Visualizing exceptions on a number line. e falls in the range $[0, N]$. If it lies below θ (gold), then there are tolerably few exceptions to the hypothesized generalization, so the learner should acquire it and memorize the exceptions. Otherwise, if e lies about θ (blue), then there are too many exceptions to the generalization to tolerate, so the learner should resort to memorization instead.

One important property of the TP is that learners will adjust their decisions about productivity as they learn new vocabulary and their tolerance thresholds shift accordingly (Figure 4a). This means that a generalization that is not productive in a learner's early development can potentially be rendered productive as that learner acquires more vocabulary (Figure 4b). Alternatively, if the number of exceptions to a generalization grows faster than the vocabulary, as may happen when a learner is learning new words from a long tail of exceptions, a generalization that was productive early in development may fall out of productivity (Figure 4c).

(a) Vocabularies grow over development



(b) A pattern may fall into productivity



(c) A pattern may fall out of productivity

Figure 4: a) Visualizing exceptions on a number line during individual development. N grows as the learner's vocabulary grows. θ grows more slowly because of how it is defined as a function of N. A learner's acquisition trajectory depends on how e grows relative to N. b) If θ grows more quickly than e, it is possible for a generalization that was non-productive early in an individual's development to become productive later. Here, e falls in the non-productive (blue) zone earlier, but falls into the productive (gold) zone later. c) If e grows more quickly than θ for a given individual, it is possible for a pattern to fall out out of productivity over the course of development.

This falling in and out is empirically testable because children with over-productive grammars are expected to occasionally make over-production errors. The most famous of these are over-regularized English past tense forms characteristic of Pinker's *U-shaped learning* Pinker et al. (1987), but examples exist across domains, as noted by Gropen et al. (1989), Bow-

erman and Croft (2008) and Yang (2016), among others. The Tolerance Principle provides a principled account for the observation that errors are overwhelmingly over-regularization, or more broadly, over-generalization, rather than over-irregularization. Regular derivations are productive, and irregular derivations are not because of how the TP is defined. For a generalization to apply to novel forms it should be productive.

The *Sufficiency Principle* (SP; Yang, 2016, p. 144) is a corollary to the Tolerance Principle which asks whether the child has heard a sufficiently large number of forms (types) that support a generalization for that generalization to be tenable. This is calculated based on the positive examples that a child has learned without consideration towards negative evidence. While the TP counts the number of attested exceptions e as against a generalization, the SP counts those that have not (yet) been attested as obeying the generalization (m)[2]. Otherwise, the derivation is very similar, as in (2). Visualized, the SP functions just like the TP in Figures 3-4 except that m replaces e.

(2) The SUFFICIENCY PRINCIPLE

Let R be a generalization over N items, of which m items are not (yet) attested to follow R. R can be extended to all N items iff:

$$m < \theta_N \text{ where } \theta_N := \frac{N}{\ln N}$$

2.3.2. Using the Tolerance and Sufficiency Principles

The advent of child-directed speech (CDS) corpora in recent decades containing years' worth of early linguistic input (e.g., CHILDES; MacWhinney, 2000) has facilitated significant progress in the field of native language acquisition. Counts from CDS, and other corpora, as I show in Chapter 3, can be taken together with TP or SP calculations to empirically measure productivity. Four aspects of language learning facilitate this application. To

[2]Yang uses $N - M$, where M is the number of items that *are* attested obeying the generalization

summarize them here, first, the relative uniformity of language acquisition: learners exhibit remarkable synchronic uniformity despite the variability of the input they receive (Labov, 1972). Second, the crucial role of type frequency: convergent results from a wide variety of research programs connect grammar learning to the *number of types* over which linguistic patterns are expressed in the input rather than the attestation of any particular lexical items (Aronoff, 1976; MacWhinney, 1978; Bybee, 1985; Baayen, 1993; Elman, 1998; Pierrehumbert, 2003b; Yang, 2016). Third, token frequency and availability: the relative age at which learners acquire vocabulary items is correlated with their *token frequencies* (Goodman et al., 2008) in the input. And fourth, small early vocabularies: the typical learner knows only a few hundred to a thousand words by around age three (Hart and Risley, 1995, 2003; Szagun et al., 2006). Since children acquire most properties of their native grammars by that age, the bulk of grammar acquisition is undertaken on the basis of relatively few mostly high frequency items rather than large adult-like lexicons.

Lexical variability between CDS corpora reflects the real-world variation in early linguistic experience that leads to precociousness or delays among learners (Maratsos, 2000; Yang, 2002). It also reflects realistic assumptions about learner knowledge. Since higher token frequency items tend to be acquired earlier, young learner's lexicons may be estimated by trimming off the less frequent items from CDS (Nagy and Anderson, 1984; Yang, 2016). Doing so yields approximations of "typical" children's lexicons which are the right size and consist primarily of high frequency items. In contrast, it is notoriously difficult to estimate the knowledge of individual children directly because, for various reasons, child-produced speech does not reflect understanding (Brown and Bellugi, 1964; Schlesinger, 1971; Bowerman, 1973; Dale and Fenson, 1996). It is these properties that make corpora of child directed speech such useful resources for studying grammar learning.

The Tolerance and Sufficiency Principles have been applied to corpus data to address a wide range of problems in language acquisition from phonology, to morphology, to syntax, such as the acquisition of English past tense (Yang, 2016, pp. 81-91), locus of the famous Past Tense

Debates (Rumelhart and McClelland, 1986; Pinker and Prince, 1988), recursive application to German plurals (Yang, 2016, pp. 121-133), the Philadelphian short-*a* system (Sneller et al., 2019), among many others. It has also been implicated in "no-default" accounts of paradigmatic gaps (Gorman and Yang, 2019) and has been shown to accurately predict the generalization behavior of young children in experimental settings (Schuler, 2017).

In the next chapter, I investigate to what extent the Tolerance and Sufficiency principles can be applied to data sourced from adult and historical corpora. This enables the historical case studies employed throughout the rest of the dissertation.

CHAPTER 3 : Historical Corpora as Child Linguistic Experience

The advent of child-directed speech (CDS) corpora in recent decades containing years' worth of early linguistic input (e.g., CHILDES; MacWhinney, 2000) has facilitated significant progress in the field of native language acquisition (Nagy and Anderson, 1984; Pinker et al., 1987; Clahsen, 1997; Lewis and Elman, 2001; Cameron-Faulkner et al., 2003; Mintz, 2003; Li et al., 2007; Goodman et al., 2008; Song et al., 2009; Tomasello, 2009; Perfors et al., 2011; Legate and Yang, 2013; Pearl and Sprouse, 2013; Yang, 2016; Kodner and Richter, 2020, *inter alia*). That said, no CDS corpora exist for the overwhelming majority of the world's languages, and none that do exist date back before the mid-20th century. Without such corpora, the insights that child language acquisition researchers gain from modern methodologies cannot be extended to most of today's world, let alone to past eras. This dissertation relies heavily on such methodologies to model child learners, but it hits an obvious roadblock: the nature of the problems being investigated requires the study of children of the past. Clearly, no CDS corpora exist for Middle English or Latin, and there are no corpora at all of prehistoric Proto-Germanic.

The contribution of this chapter is methodological: I establish that CDS and modern and historical non-CDS corpora are fundamentally similar along some dimensions relevant for child language acquisition despite the differences that intuitively exist between them. This stands to facilitate research into acquisition in the past as well as a more diverse range of modern languages, the overwhelming majority of which lack CDS corpora. There are four enabling properties of language acquisition that play a part here. First, learners show impressive synchronic uniformity even though they each receive unique inputs (Labov, 1972). This means that there is such a thing as a "typical" child who can be modeled. Second, it is the number of *types* over which a linguistic pattern is expressed in a learner's input rather than token frequency or attestation of any particular items that is crucial for grammar learning (Aronoff, 1976; MacWhinney, 1978; Bybee, 1985; Baayen, 1993; Elman, 1998; Pierrehumbert, 2003b; Yang, 2016). The latter must be true given the first point – chil-

dren receive different inputs yet acquire near-identical grammars. Third, token frequency is still relevant, but less directly, as it is correlated with the order of acquisition (Goodman et al., 2008). And fourth, early learner vocabularies are quite small, consisting of only a few hundred to a thousand words by age three (Hart and Risley, 1995, 2003; Szagun et al., 2006). Since children acquire most properties of their native grammars by that age, the bulk of grammar acquisition is undertaken on the basis of relatively few mostly high frequency items rather than large adult-like lexicons.

Lexical variability between CDS corpora reflects the real-world variation in early linguistic experience that leads to precociousness or delays among learners (Maratsos, 2000; Yang, 2002). It also reflects realistic assumptions about learner knowledge. Since higher token frequency items tend to be acquired earlier, young learner's lexicons may be estimated by trimming off the less frequent items from CDS (Nagy and Anderson, 1984; Yang, 2016). Doing so yields approximations of "typical" child lexicons which are the right size and consist primarily of high frequency items, while the full-corpus lexicons more closely reflect learners far past the age of early acquisition (Anglin, 1993; Nation and Waring, 1997). In contrast, it is notoriously difficult to estimate the knowledge of individual children directly because, for various reasons, child-produced speech does not reflect understanding (Brown and Bellugi, 1964; Schlesinger, 1971; Bowerman, 1973; Dale and Fenson, 1996). It is these properties that make corpora of child directed speech such useful resources for studying grammar learning. I leverage this insight in the study of modern North American English partial /aɪ/-raising in Chapter 4.

Once it is established that historical and other non-CDS corpora share these properties as well, researchers can apply models of language acquisition to historical data to work out how, when, and whether the process of native language acquisition effects change. To that end, I conduct four studies, extending Kodner (2019), which elaborate on the similarities between modern and historical non-CDS corpora on one hand and CDS on the other for the purpose of modeling productivity in studies of Middle English ditransitives, Proto-Germanic

strong verbs, and Latin past participles (Chapters 5-7). I begin in Section 3.1 by illustrating the effect that trimming low token frequency items has on CDS and adult corpora in Modern English. This is extended to historical corpora in Section 3.2, where I compare semantic overlap between cross-linguistic modern CDS, historical, and reconstructed lexicons. Section 3.3 moves from lemmas to a study of sparsity in inflectional morphology. Finally, Section 3.4 demonstrates that a type-based threshold learning algorithm to morphological problems yields the same acquisition outcomes in Modern English lexicons taken from CDS and modern non-CDS and to Icelandic lexicons drawn from historical and modern non-CDS.

3.1. Verbal Lexicons Derived from Child-Directed Speech and Adult Corpora

This study establishes the similarity between lexicons derived from adult literary corpora and those derived from corpora of child directed speech. I begin by demonstrating the effect of trimming low frequency vocabulary from the extracted lexicons, and following that, I compare the attested type frequency of various linguistic properties between the adult and CDS-derived lexicons. Type frequencies of these properties are quantitatively similar in these corpora despite superficial differences in specific lexical content.

Adult corpus lexicons are drawn from the Corpus of Contemporary American English (COCA; Davies, 2009), which contains millions of lemmatized and POS-tagged words of text drawn from five genres: spoken, popular magazine, fiction, newspaper, and academic. Each genre contains individual subcorpora for each year, and each subcorpus contains between 2.5 and 5.5 million tokens and between 4,200 and 10,200 verb types when those tagged as auxiliaries or modals are excluded.[1] Child input lexicons are drawn from three lemmatized POS-tagged corpora within CHILDES, each containing roughly 1,000 unique verb lemmas, again with auxiliaries and modals excluded: Brent (n=984; Brent and Siskind, 2001), Brown (n=916; Brown, 1973), and MacWhinney (n=1042; MacWhinney, 1991).[2] These were cho-

[1]Since auxiliaries are excluded, the token frequencies for *be*, *have* and *do* only count the instances of these as main verbs. Since English has few auxiliary and modal types, the choice of whether to include them or not does not meaningfully affect the results.

[2]Lemmas are extracted from the morphological annotations provided in these corpora

sen for their large size relative to other CDS corpora, each containing about a year's worth of child-directed speech. I focus on verbs here for consistency across studies and because they show more interesting inflectional patterns in English than other syntactic categories do. That said, the Zipfian statistical corpus distributions of verb lemmas, inflectional categories, and so on, are the same as those obeyed by other categories (Chan, 2008; Finley, 2018), which is demonstrated in practice by learning behavior in computational morphology learners (e.g., Lignos et al., 2010). The results can therefore be extended to other syntactic categories. Lexicon sizes are summarized in Table 1.

Corpus Type	**Number**	**Lexicon Size (n)**
CDS	3	918, 984, 1042
Academic	28	4,917–7,786
Fiction	28	5,544–8,015
Magazine	28	6,116–9,662
News	28	5,080–7,365
Spoken	28	4,144–5,566

Table 1: Lexicon estimates from CHILDES CDS and COCA.

The most frequent verb lemmas are tabulated for each CHILDES corpus and COCA subcorpus, and four estimates are made from each with the following frequency cutoffs: $n = all$, 1,042 (all types in the largest of the CHILDES corpora), 500, and 100. The three frequency-trimmed conditions can be taken to represent the lexicons of late, middle, and early learners respectively. Given the total vocabulary size estimates of Hart and Risley (2003) and Szagun et al. (2006), a learner who only knows about 100 verbs is certainly less than three years old, while one who knows 500 is perhaps closer to school age.

3.1.1. Raw Lexical Similarity

Measuring lexical overlap between extracted lexicons illustrates the effect of trimming infrequent vocabulary. *Jaccard similarity* $|A \cap B|/|A \cup B|$ is employed to measure the set overlap

between each pair of lexicons (self-similarity excluded). The metric has a range [0,1] where higher is more similar.

Figure 5 shows the range of Jaccard similarities between CDS and COCA corpora on the left and between COCA corpora of different genres on the right. Two observations stand out. First, similarities are much higher for all frequency-trimmed conditions than for $n =$ *all*, which suggests that items which are not shared between corpora are predominately low-frequency. Second, though CDS-COCA similarities are lower than COCA-COCA similarities, their ranges overlap once frequency trimming has been applied, which means that some CDS corpora are more similar to some COCA corpora than some COCA corpora are to one another. Specific lexical items are not necessarily well-shared across corpora regardless of genre, but frequency trimming improves the situation significantly.

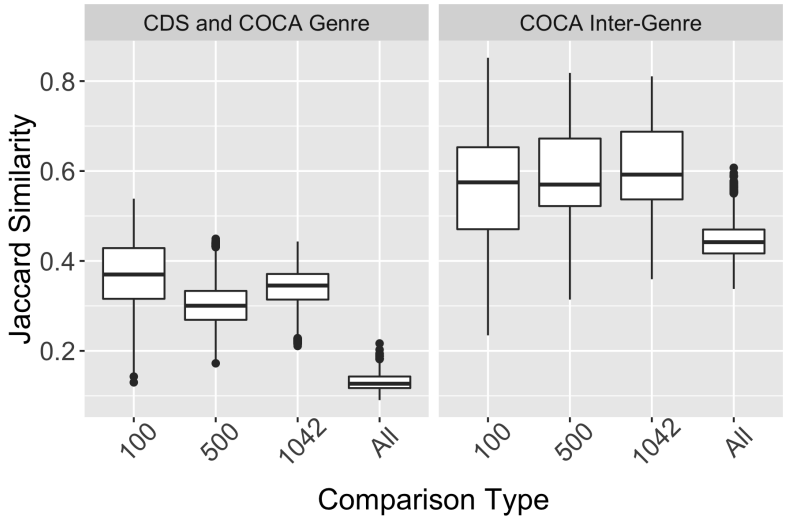


Figure 5: Corpus-derived lexicon Jaccard similarities by comparison type and lexicon size.

3.1.2. Lexical Property Similarity

But what matters for learning is often not the individual linguistic items in the input so much as the properties of those items. As discussed in the introduction, the type frequency of some property, the number of items in the lexicon exhibiting it rather than which specific items

those are, is what is drives productivity learning. This time, I compare the same adult COCA and CHILDES-derived lexicons in terms of the type frequencies of three linguistic properties. These were chosen for coverage: first, *Latinate* verbs are a morphophonological class which is acquired relatively late, around the start of school (Tyler and Nagy, 1989; Jarmulowicz, 2002), second, *irregular verbs* are morphological, learned much earlier, and factor into the classic *Past Tense Debates* about productivity in the acquisition literature (Rumelhart and McClelland, 1986; Pinker and Prince, 1988; Pinker and Ullman, 2002, *inter alia*), and third, *double object alternator verbs* are syntactic and semantic in nature (Rappaport Hovav and Levin, 2008), and their acquisition is one of the classic case studies in argument structure learning (Baker, 1979; Pinker et al., 1987; Gropen et al., 1989; Yang, 2016). The results show that there is less variation between corpora in terms of type frequencies in terms of lexical identity, and that the CDS-derived lexicons are in general quantitatively similar to the adult lexicons in the frequency-trimmed conditions.

Irregular verbs

Irregular verbs in English are those that undergo stem changes or suppletion when forming the past tense and past participle, e.g., *sing* $\sim$ *sang* $\sim$ *sung*, *go* $\sim$ *went* $\sim$ *gone*, or *tell* $\sim$ *told* $\sim$ *told*. A learner acquiring English verbal morphology must work out which of these verbs are inflected according to some generalizable pattern and which are truly one-off "irregulars" that must be listed or memorized (Berko, 1958; Pinker and Prince, 1994). Figure 6 shows the mean number of strong verb lemmas by genre for each frequency cutoff n. It is plain from visual inspection alone that CDS and the COCA genres become much more alike when the rare items are trimmed from COCA. It is also striking that academic writing rather than CDS appears to be the greatest outlier for each trimmed condition.

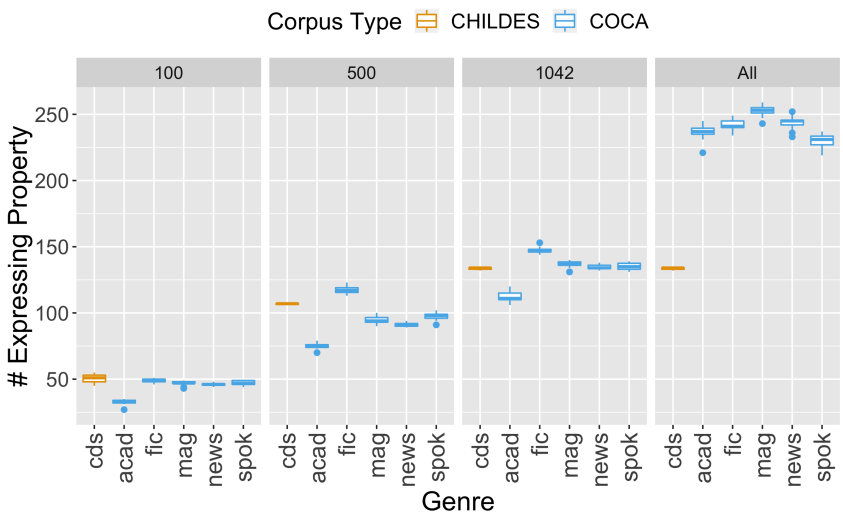


Figure 6: Number of irregular verbs attested per corpus by genre and corpus size.

At $n = all$, the adult lexicons contain far more irregular verbs than the CDS-derived lexicons simply because they are taken from larger corpus samples, but when trimmed to $n = 1042$ and 500, CDS falls within the range of the adult lexicons, while at $n = 100$, CDS overlaps with fiction. A regression predicting the number of strong verbs by CDS/adult status finds no significant difference between CDS and adult lexicons in any of the frequency-trimmed conditions – if one were presented with the box plots in Figure 6 with the genre labels and colors removed, it would not be possible to identify which box corresponded to CDS in the trimmed conditions.

Double object / to-dative alternator verbs

The acquisition of DO/to-dative verbs (e.g., *give*, *send* and *tell*) (Levin, 1993, §2.1) is one of the classic problems in argument structure acquisition. Their attestation in these corpora reveals the same kind of pattern as the irregular verbs: again, trimming the low token frequency items from the COCA-derived lexicons brings them in line with the CDS lexicons.

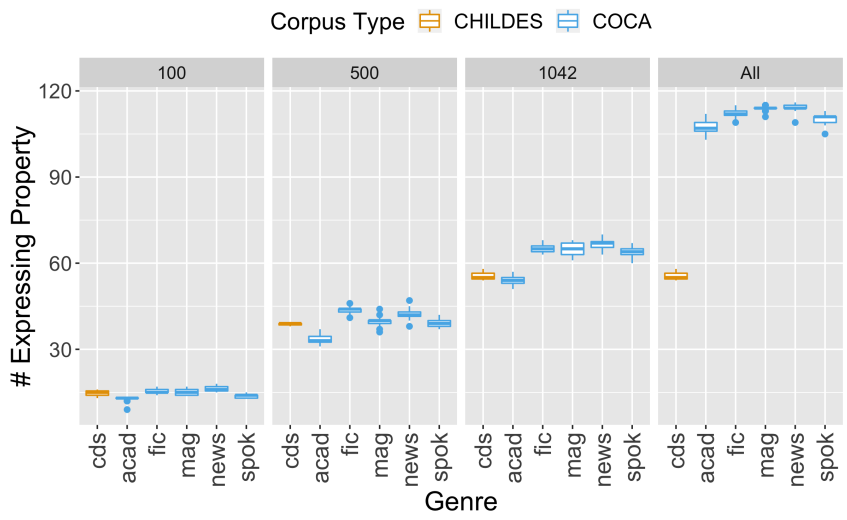


Figure 7: Number of double object alternator verbs attested per corpus by genre and corpus size.

There is no significant difference between CDS and adult lexicons at $n = 500$ or 100, and while CDS is statistically different from adult at $n = 1042$, it is not different from academic, and the difference between CDS and adult means decreased from a factor of about 200% to near 10%.

Latinate verbs

Unlike the previous properties, Latinate verbs are saliently associated with genre (Levin et al., 1981; Levin and Novak, 1991), and many, but not all are high-register (COCA contains *encapsulate*, *irradiate*, *reconstitute*, but also *confuse*, *offer*, and *remember*). Additionally, the morphophonological generalizations associated with English Latinate vocabulary are acquired late, typically not until children enter school. As such, we expect there to exist significant quantitative differences between the rate of Latinate verbs in CDS-derived and adult-derived lexicons as shown in Figure 8.

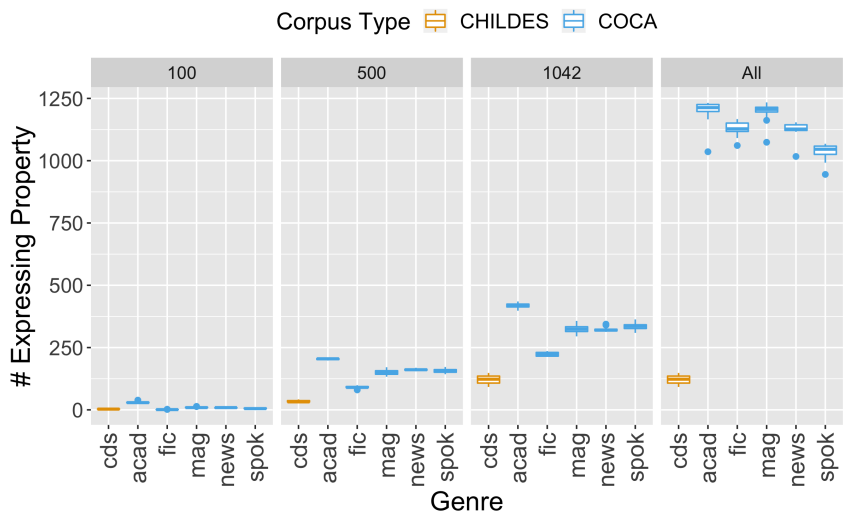


Figure 8: Number of polysyllabic Latinate verbs attested per corpus by genre and corpus size.

This prediction bears out since every test shows a significant difference except for $n = 100$. Nevertheless, frequency trimming brings the type frequencies of CDS and non-CDS much closer together since Latinate vocabulary is disproportionately present among low-frequency items in every COCA genre. Notably, academic lexicons once again differ from all other genres.

3.1.3. Interim Conclusions

These studies show that type frequencies in corpora derived from child-directed speech are statistically similar to frequency-trimmed corpora derived from adult literary genres even though they differ in their specific lexical contents. In every instance, frequency trimming brings CDS-derived and non-CDS-derived type counts much closer together, and in most cases there is no statistically significant difference between the two trimmed lexicon categories. Adult corpora may be reasonably substituted for CDS corpora for the purpose of modeling grammar learning in child language acquisition, since it is these type frequencies that are directly relevant and frequency trimming is just a normal step for approximating

child vocabulary size and composition when analyzing CDS for productivity.

3.2. Verbal Lexicons Derived from Child-Directed Speech and Historical Corpora

Child language acquisition is often implicated as a driving force in language change (Sweet, 1899; Halle, 1962; Kiparsky, 1965; Andersen, 1973; Baron, 1977; Lightfoot, 1979; Niyogi and Berwick, 1996; Kroch, 2001; Yang, 2002; Cournane, 2017, *inter alia*), and some programs which do not privilege child language acquisition still acknowledge a special role for children (Labov, 1989). Children of the past must have acquired language in a way similar to modern children (a straightforward consequence of linguistic uniformitarianism (Labov, 1972; Walkden, 2019)), and the way in which children acquire language makes it amenable to study through modern CDS and non-CDS corpora. So, conceptually, it should be possible to investigate language acquisition in the past by investigating corpora of the past. The obvious obstacle then is practical and empirical: can past corpora be substituted for modern corpora? The answer is not immediately obvious. Corpus linguistics, especially historical corpus linguistics, is more used to focusing on differences, in genre, in style, in lexicon, in grammar, and so on, than on similarities.

This study takes on the basic assumption of irreconcilability between historical and modern corpora by extending the previous analysis back through time to compare the contents of modern CDS-derived and (frequency-trimmed when applicable) historical lexicons. Since linguistic properties like the presence of "irregular" inflection are not conserved across languages, this study compares the meanings contained in each lexicon instead. Items are matched between two lexicons if there is a shared translation between them. For example, English *slide* is matched with Spanish *resbalar* 'slip,' Latin *lābī* 'slip, glide,' and Proto-Germanic **slīdanǫ* 'slide.'[3] Since correspondences between the lexicons are no longer one-to-one, Jaccard similarity does not make sense here. Instead, raw percent overlap is calculated as $|\mathrm{A} \cap \mathrm{B}|/\min(|\mathrm{A}|, |\mathrm{B}|)$. Overlaps are systematically higher than Jaccard similarities because the denominator is smaller.

[3] A full list of correspondences is provided in the supplementary material.

English CDS (Brown) and Spanish CDS from CHILDES (FernAguado, Hess, OreaPine, Remedi, Romero, and SerraSole (Romero et al., 1992; Hess Zimmermann, 2003; MacWhinney, 2000; Aguado-Orea and Pine, 2015)) are compared to two pre-modern lexicons: Latin from all Old and Classical texts in the Perseus online edition (Smith et al., 2000), and Proto-Germanic (PGmc) taken from all securely reconstructable strong verbs in Seebold (1970).[4] The Proto-Germanic strong verbs are chosen because they are not semantically coherent (so do not introduce a semantic confound) and provide a sufficiently large set for comparison. The PGmc lexicon is the limiting factor here, so frequency cutoffs are established for each corpus-derived lexicon to bring them in line with the size of PGmc. To establish a within-language CDS baseline, the overlap procedure was performed between the Brown and Brent English CDS corpora with the same frequency cutoff applied to both, and Brown and Spanish were compared as a cross-language CDS baseline. The lexicons used in these calculations are provided in Appendix A.1.

Comparison	**Sizes**	**Overlap**
EN Brown-EN Brent	260/257	81.71%
English-Spanish	260/263	73.07%
English-PGmc	260/258	66.67%
Spanish-PGmc	263/258	71.32%
English-Latin	260/260	75.77%
Spanish-Latin	263/260	79.62%

Table 2: Modern CDS, historical and prehistoric lexicon raw percent overlaps.

Table 2 reveals a spread of about 15 points between lowest and highest raw percent overlap scores. The within-language English-English baseline is the highest at about 82%, while the cross-language CDS baseline is somewhat lower at 73%. The Latin comparisons are higher than the CDS baseline, while the Proto-Germanic numbers are a few points lower. The high overlap between the reconstructed and modern lexicons is likely due to the fact that

[4] I thank Donald Ringe for his help in sorting through Seebold.

words are securely reconstructable only if they are retained in multiple daughter branches, and that the words that are likely to be retained tend to be frequent everyday terms – the same kind that we expect to find in CDS. For example, the Proto-Germanic words for 'bite,' 'wait,' 'fall,' 'pull,' 'sing,' and 'help' are reconstructable because they were retained in its daughters, and their equivalents are all present in both the modern English CDS corpora since they are common everyday terms.

It seems that cultural differences account for the extra discrepancy between Proto-Germanic and CDS. The PGmc lexicon contains many terms for farming ('sow,' 'plant,' 'thresh'), household chores ('weave,' 'knead,' 'bank a fire') and other aspects of culture ('cast lots,' 'be a retainer') that modern urban children are unlikely to know, but which children growing up in Iron Age agricultural societies must have. These cultural terms account for roughly 3.1 points of overlap, which when added in would put the PGmc comparisons in line with the English-Spanish overlap. Second, the English-Other overlaps are smaller than the corresponding Spanish-Other overlaps, which may be partially explained by the lack of stative verbs in English compared to the other languages and by English's predilection for phrasal verbs which prevent matches, for example phrasal 'look (for)' is used in the corpus instead of 'search.' PGmc, Spanish, and especially Latin have a number of stative verbs, and these cannot be matched in the English lexicon since the corresponding English meanings are either not verbs or rare verbs.

All in all, lexical overlap is conserved between CDS, adult historical corpora, and reconstructed lexicons about as well as between CDS lexicons. They contain largely the same kinds of meanings despite their varied origins, and differences between lexicons can be partially account for by cultural differences rather than corpus differences. The lists collated in the supplementary material show that higher frequency items are more likely to match than low frequency items, even among different CDS corpora. This reiterates the point from Section 3.1 that low token frequency items are more likely to be corpus-specific than high-frequency items. The similarities between Proto-Germanic, Latin, and modern CDS

motivate the methodologies taken in Chapters 6-7.

3.3. Distribution of Forms in Child-Directed Speech, Modern Adult, and Historical Corpora

The previous studies measure the similarities between lexicons estimated from child-directed speech, modern adult, and historical adult corpora in terms of shared verb lemmas and shared classes of verb lemmas. This study extends that comparison to inflected forms to characterize the type of data sparsity present in these corpora. Modern, historical, and CDS corpora are similarly sparse in terms of verbal morphological inflection. Sparsity is quantified with two metrics, *paradigm saturation* (Chan, 2008) and *inverse paradigm saturation*.

This study analyzes three CDS corpora: English Brown, the Spanish CDS from the previous study, and the German Leo corpus from CHILDES (Behrens, 2006). These contain morphological feature annotations as well as lemmatization and POS-tagging. For modern and historical adult corpora, the study uses the English, Finnish, German, Spanish and Turkish as well as Gothic and Latin lemmatized, POS-tagged, and morphological feature-annotated text provided as part of the Universal Dependency Treebank (UD; Nivre, 2018). The Universal Dependencies Latin corpus is a subset of the one used in the previous section and Chapter 7. Table 3 presents superficial statistics for each corpus. The number of verb tokens in each corpus is to the same order of magnitude, with CDS corpora falling within the range of the modern and historical adult corpora. However, CDS token/type ratios are an order of magnitude higher than the rest. In other words, the average verb form in the CDS corpora is used roughly ten times as often as in the adult corpora.

Corpus Type	Language	# Verb Tokens	# Verb Types	Ratio
CDS-CHILDES	English	94,768	916	103.46
CDS-CHILDES	German	96,686	879	110.00
CDS-CHILDES	Spanish	81,351	641	126.91
Modern-UD	English	53,796	3,225	16.67
Modern-UD	Finnish	63,891	3,476	18.38
Modern-UD	German	21,835	2,826	7.73
Modern-UD	Spanish	85,861	5,019	17.11
Modern-UD	Turkish	12,064	968	12.46
Historical-UD	Gothic	12,749	1,172	10.88
Historical-UD	Latin	99,066	2,833	34.97

Table 3: CDS, modern adult, and historical adult superficial corpus statistics.

Paradigm saturation (PS; Chan, 2008; Lignos and Yang, 2018) measures the proportion of a lemma's theoretically possible inflected forms that is actually attested in a given corpus. It has been observed that paradigm saturation tends to be very low in languages with even moderately complex inflectional paradigms and that saturations by lemma follow a long-tailed Zipfian distribution. Table 4 reports maximum, mean, and median saturations for the verbs in each corpus. Note that they are only marginally higher for CDS than for UD despite the much higher token/type ratios. Medians are lower than means across the board due to the Zipfian distribution's long tail. The average CDS German verb is attested in just 2.20 forms while the average German verb in the UD is attested in 1.69 forms, and the average CDS Spanish verb is attested in just 3.29 forms while the average Spanish verb in the UD is attested in 5.57 forms. This is striking given the order of magnitude higher token/type ratios for these corpora.

Corpus Type	Language	|Paradigm|	Max Sat.	Mean Sat.	Median Sat.
CDS	English	5	100%	43.23%	40.00%
CDS	German	29	44.83%	7.59%	6.90%
CDS	Spanish	67	52.24%	8.31%	4.48%
Modern	English	5	100%	42.80%	40.00%
Modern	Finnish	150	27.33%	2.46%	1.33%
Modern	German	29	51.72%	5.83%	3.45%
Modern	Spanish	67	43.28%	4.91%	1.49%
Modern	Turkish	120	99.17%	4.83%	1.67%
Historical	Gothic	52	53.85%	6.31%	3.85%
Historical	Latin	113	81.42%	5.90%	2.65%

Table 4: CDS, modern adult, and historical adult paradigm saturations.

Figures 9-11 show visually that CDS, modern adult, and historical paradigm saturations obey the same Zipfian-like distributions. The most significant difference between the CDS-derived and non-CDS distributions are the longer tails in non-CDS (clearest on the German plots), however, these tails are made up almost exclusively of very low frequency items which would be trimmed for any analysis.

The same patterns that hold for mean saturation in CDS corpora also exist in historical and modern adult corpora, so once again non-CDS corpora may be used to reason about the facts of child linguistic experience. The overwhelmingly low paradigm saturations attested in non-English corpora have significant bearing on the Paradox of Language Change (Chapter 6.6.2): as far as morphological inflection is concerned, children will not necessarily receive enough positive evidence to correct transient errors the accrue during the acquisition process.

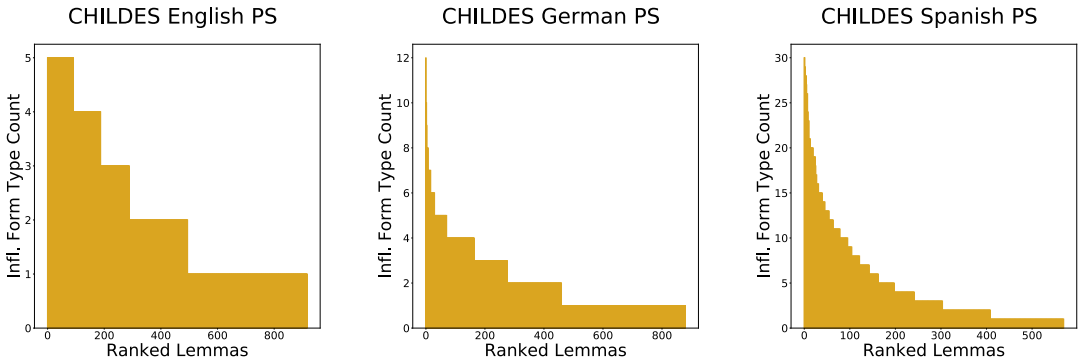


Figure 9: Non-normalized (scaled by maximum PS) paradigm saturation plots for modern CDS English, German, and Spanish.

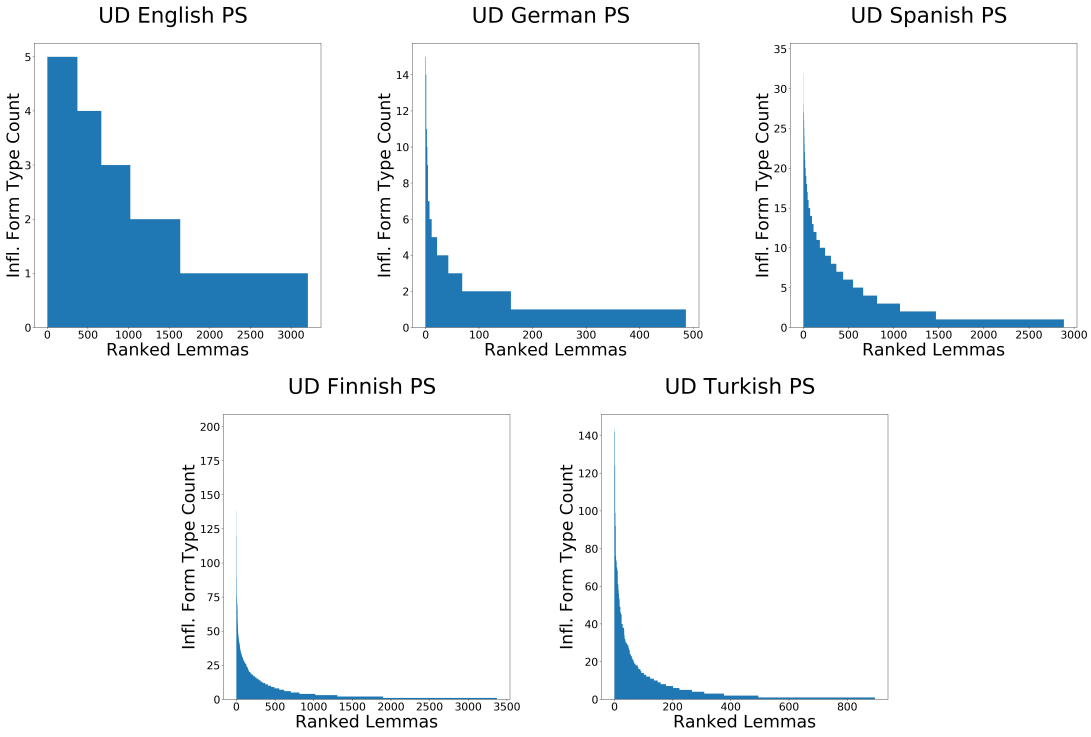


Figure 10: Non-normalized paradigm saturation plots for modern English, German, Spanish, Finnish, and Turkish.

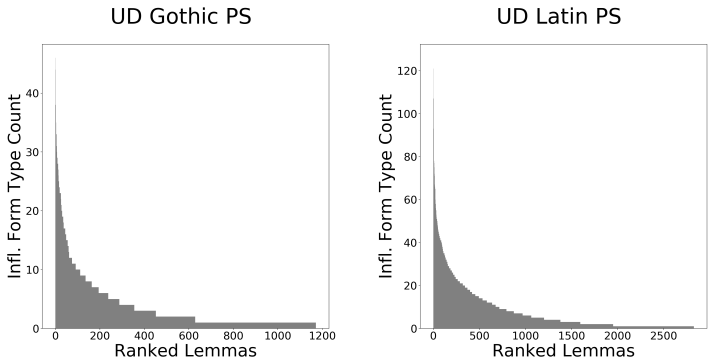


Figure 11: Non-normalized paradigm saturation plots for historical Gothic and Latin.

Inverse paradigm saturation (IPS) is related to paradigm saturation. For every inflectional category, it measures the proportion of lemmas present in a corpus which are attested in that category. IPS tends to follow long-tailed distributions, as with PS, so mean IPSs are low. This means that many inflectional categories are attested with only a few items, while some are attested with relatively many items.

IPS distributions are less tail-heavy than PS, particularly for languages with smaller paradigms. This is visually clear in Figures 12-14. IPS is also not quite as conserved as PS is between CDS and non-CDS, though there is no obvious difference between modern non-CDS and historical corpora. This seems to be a genre effect. Looking at the highest IPS items, these tend to be the 3rd person singular present and past/preterite and some participles in most non-CDS corpora, but the 2nd person singular present is the most common in German CDS and the third most common in Spanish CDS. 1st person singulars have higher IPS as well in CDS. One expects to find similar discrepancies between narratives in COCA fiction compared to COCA news as well. More broadly though, certain patterns are conserved. For example, subjunctives and second person plurals tend to be in the tail region of the IPS curves regardless of CDS status.

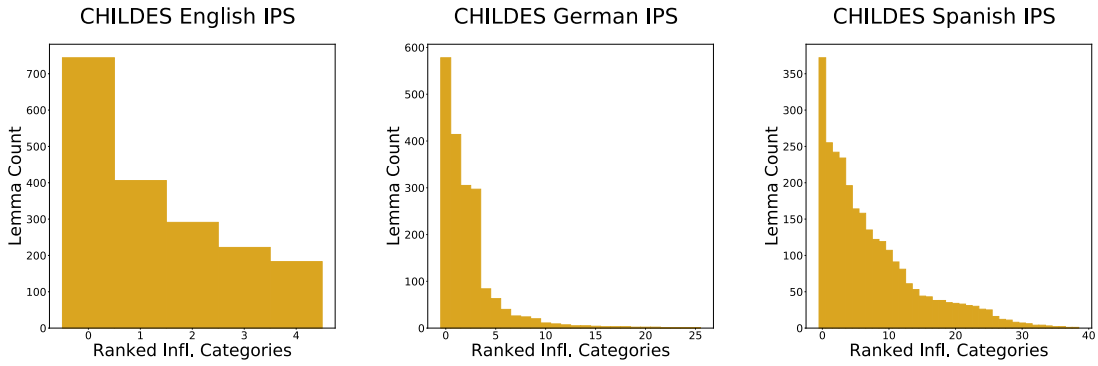


Figure 12: Non-normalized inverse paradigm saturation plots for modern CDS English, German, and Spanish.

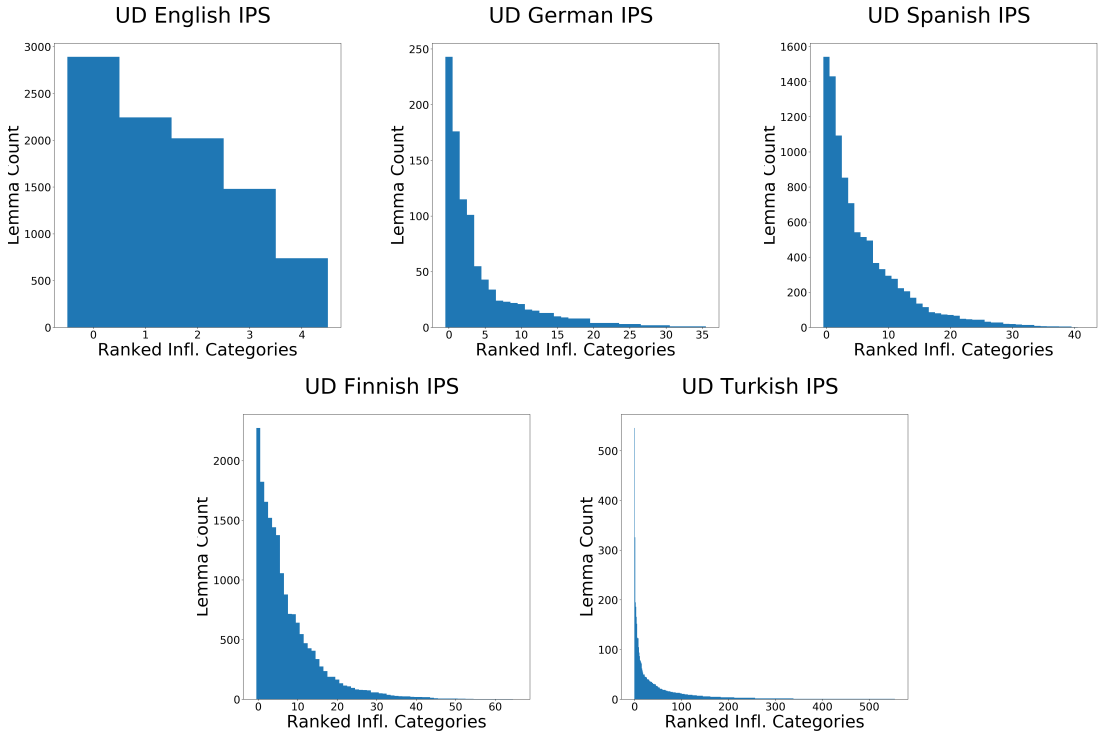


Figure 13: Non-normalized inverse paradigm saturation plots for modern English, German, Spanish, Finnish, and Turkish.

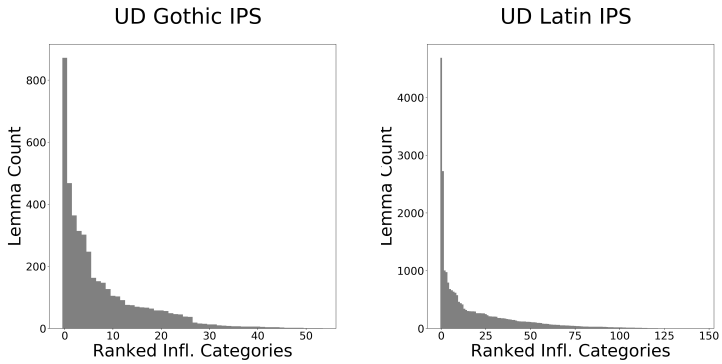


Figure 14: Non-normalized inverse paradigm saturation plots for historical Gothic and Latin.

3.4. Deploying an Acquisition Model

This study compares learning outcomes when a learning algorithm is applied to CDS, modern non-CDS, and historical corpora. First, I compare the acquisition of Modern English productive past *-ed* on lexicons sourced from CDS and adult corpora. Following that, I apply the same algorithm to a past tense generalization in Old and contemporary Icelandic to draw conclusions about child development in the past. In both cases, I apply the Tolerance Principle (2.3), though any type-based acquisition model could be used here.

3.4.1. Modern English Past -ed

To investigate whether CDS-derived and adult-derived lexicons yield similar learning outcomes, I model the acquisition of the English productive past-forming *-ed* pattern. The acquisition of English past tense is a complex and classic problem in morphological learning which has triggered decades of debate (Berko, 1958; Rumelhart and McClelland, 1986; Pinker and Prince, 1988, 1994; Ramscar, 2002; Kirov and Cotterell, 2018, *inter alia*), and the acquisition of a default past *-ed* pattern is one piece of the challenge. In terms of the Tolerance Principle, the pattern being acquired is one that applies *-ed* (with the appropriate morphophonology) to a verb to produce its past tense form. All verb types learned up to a given point in development count towards the N in the formula, while those verb types learned with irregular pasts by that point make up e. Specific lexical items do not matter

in the calculation, nor do the particular values of e and N past establishing whether or not e lies below the tolerance threshold.

Yang (2016) finds that the English lexicon is such that early learners who know 500 or fewer verbs know too many irregulars relative to regularly derived past verbs to learn *-ed* productively. The situation is marginal at 800, and learners can finally reliably acquire the productive past once they know 1,000 verbs.

I reproduce these results. 1,000 CDS-derived lexicons with 1,000 items each are sampled from the 1,515 unique lemmas attested together in English Brent, Brown, or MacWhinney weighted by their token frequencies across those corpora, then the same sampling is performed on the 1,500 most common COCA lemmas to create 1,000 sample adult-derived lexicons. The TP is calculated on each lexicon for the top N=100, 150, and 200 through 1,000 items. For all CDS-derived and adult-derived lexicons, the results at $N = 100$, 500, and 1,000 are identical to what Yang (2016) reports for both sample types: every lexicon fail to generalize past *-ed* at low N but succeed by $N = 1,000$ as shown in Table 5. On its own, this TP calculation would imply that a learner would not acquire a productive past *-ed* until knowing near 1,000 unique verbs' past forms, but see Yang (2016, Chapter 4.1.2)

Corpus Type	**100**	**300**	**500**	**800**	**1,000**
Yang 2016	no	no	no	no	**yes**
CDS Samples	0%	0%	0%	0.02%	**100%**
Adult Samples	0%	0%	0%	68.4%	**100%**

Table 5: Percent of sampled corpora generalizing *-ed* by the Tolerance Principle.

What differences do exist cluster around the $N = 800$ point that Yang reports as marginally non-productive. When plotted in Figure 15, we see that the adult corpus learning curve is shifted somewhat to the left, which reflects the slightly lower average number of irregular verbs in the adult-derived lexicons at that point (117 vs. 127). This is effectively a sample-dependent relative developmental delay of the kind reported in Maratsos (2000) and Yang

(2002). Regardless, the final learning state is identical for every single adult and CDS sample.

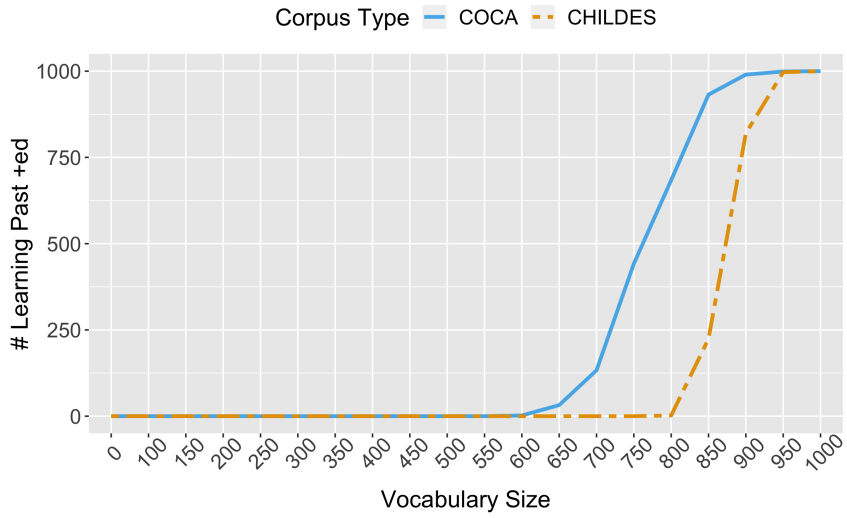


Figure 15: Proportion of learners acquiring productive *+ed* past by vocabulary size.

3.4.2. Icelandic Strong Verbs

Finally, I apply the Tolerance Principle to a learning problem in both Old Icelandic and contemporary Icelandic to compare modern and historical learning trajectories. The remarkable diachronic stability of Icelandic morphology renders it uniquely suitable for this comparison since it allows us to set up a null hypothesis: patterns that emerge among the highest frequency items in a contemporary Icelandic text should be apparent in a contemporary Icelandic text as well. We could run the same test on Old English or Latin, but we would have no hypothesis to test since their modern descendants are so different.

Icelandic, like English, has a significant number of verbs that express past tense by stem mutation (so-called *strong verbs*, e.g., *syngja* ∼ *söng* 'sing'), and a much larger number which express the past through suffixation (multiple classes of *weak verbs*, e.g., *dvelja* ∼ *dvaldi* 'dwell, reside,' *svara* ∼ *svaraði* 'answer, respond'). It is up to child learners to sort out whether any patterns exist over these verbs that indicate which type of inflection to

productively employ. This turns out to be quite challenging – even eight-year-old Icelandic children still make a non-trivial number of errors in which they substitute one class for another (Ragnarsdóttir et al., 1999).

I consider one such generalization that illustrates this pattern of learning: the relationship between monosyllabic verbs (e.g, *dá* 'adore, worship,' *ná* 'get, obtain,' *sjá* 'see, perceive') and strong inflection. Most verbs in this set are weak, but a few are strong as well, so a learner has to determine which ones belong to the productive pattern, if any, and which should be learned as exceptions. To investigate this quantitatively, I extract all verbs which are attested at least once in the past tense from the Old Icelandic (composed no later than 1350) and contemporary Icelandic (written no earlier than 1900) texts in the POS-tagged and lemmatized Icelandic Parsed Historical Corpus (Wallenberg et al., 2011), which results in two sets of 735 and 921 unique verb types respectively. Next, I apply the same sampling procedure as in the above section to generate 1,000 sample lexicons from each era to model the learning trajectories of "typical" learners exposed to these verbs in their input. The resulting developmental trajectories are presented in Figure 16.

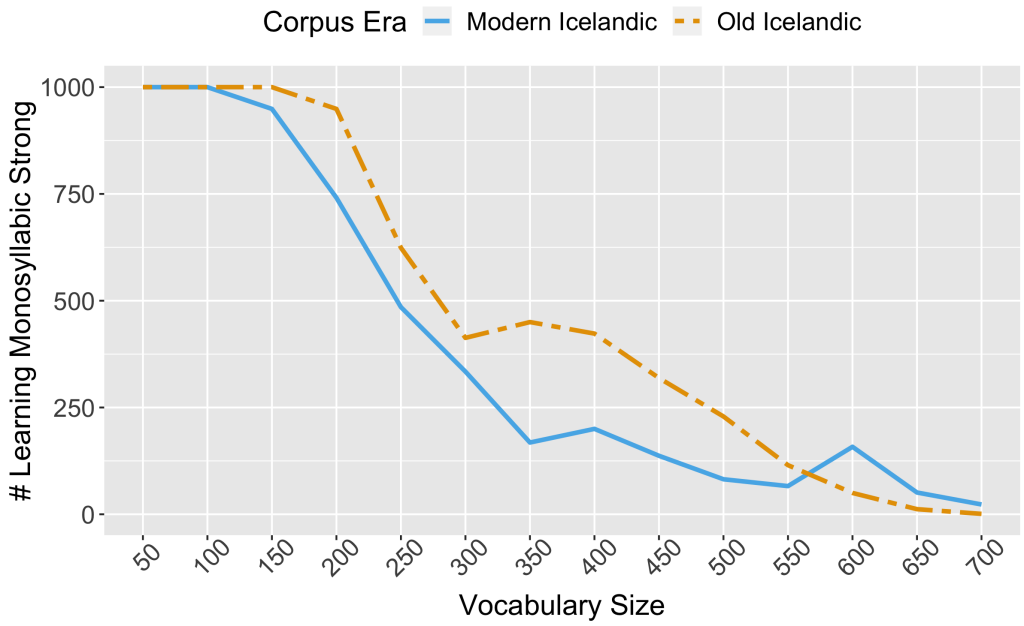


Figure 16: Proportion of learners acquiring productive strong inflection for verbs with monosyllabic stems.

There are two takeaways here. First, the average learning trajectories are closely matched between the Old Icelandic and contemporary Icelandic learners, which confirms our expectations of morphological conservatism in Icelandic and once again demonstrates the insignificance of genre differences when it comes to the type expression of linguistic properties. Second, it shows that all early learners with small vocabularies can productively apply strong verb inflection to monosyllabic verbs, but that they gradually lose this option as their vocabularies grow and monosyllabic strong verbs are revealed to be the true exceptions. This pattern of early spurious productivity is consistent with the widely observed tendency for "irregulars" (here, strong verbs) to cluster among high token frequency items (Bybee, 1985; Baayen, 1993; Yang, 2016, *inter alia*). It drives modern learners to tenable but ultimately incorrect hypotheses about their languages (e.g., Xu and Pinker, 1995; Ragnarsdóttir et al., 1999), and now we can say that it did so for Icelandic learners of the past too.

3.5. Conclusions

The studies presented in this chapter identify substantial similarities between corpora of child-directed speech and lexicons derived from modern, historical, and reconstructed lexicons. Once they are trimmed by token frequency in order to approximate child lexicon sizes (or filtered as a byproduct of the comparative method), they express type frequencies to a degree that is statistically similar to those in CDS corpora. Since it is these type frequencies that are critical for the acquisition of linguistic generalizations, non-CDS corpora can be used to model aspects of child language learning.

These results have relevance for each of this dissertation's case studies. The simple demonstration that the trimming process can be applied to English CDS is useful for the Chapter 4 synchronic study on /aɪ/-raising. The study of the Middle-English *to*-dative in Chapter 5 benefits from the observation that double object alternator and, to an extent, Latinate verbs are conserved between CDS and non-CDS. Since the fairly extensive non-CDS Middle English and Early Modern English corpora available to us are reasonable substitutes for child directed speech when it comes to studying the lexical distribution of the to-dative, we can model the input and behavior of Middle English learners. And, since the Proto-Germanic strong verbs that are reconstructable are those preserved in each daughter branch, which tend to be those high frequency items learned in childhood, the set of securely reconstructable Proto-Germanic strong verbs is a surprisingly good approximation of a prehistorical learner's verbal lexicon. Chapter 6 makes use of this observation in investigating analogical change in the Proto-Germanic strong verb paradigm. Finally, since paradigm saturation is conserved across time and genres, and since the highest frequency known Classical Latin verbs cover the same meaning space seen in modern CDS, we can use the sum total of Classical Latin literature to model the acquisition of Latin morphology. Chapter 7 leverages this fact to better understand the synchrony and diachrony of the Latin past participles and *t*-deverbals.

CHAPTER 4 : Transparent /aɪ/-Raising as a Contact Phenomenon

Canadian Raising is a phonological process which raises the nucleus of both the /aɪ/ and /aʊ/ diphthongs above 60Hz before voiceless segments (Labov et al., 2005, ANAE, p. 205). The /aɪ/ diphthong is raised in much of Canada (Joos, 1942; Chambers, 1973) as well as in many American dialects (Vance, 1987; Dailey-O'Cain, 1997), including the Inland North, resulting in alternations among a large number of near-minimal pairs distinguished by their voicing (3). It is often put forth as a classic examples of phonological opacity because it is canonically conditioned by *underlying* voiceless segments such as the flapped-/t/ in *writer* which yields a minimal pair with the flapped-/d/ in *rider* (4). However, not all /aɪ/-raising speakers exhibit this opaque pattern: so-called *transparent* or *phonetic* /aɪ/-raising speakers only raise before surface voiceless segments (5). The existence of this latter group has renewed debate about the ultimate origins of the raising patterns and the relationship between transparent and canonical raising.

(3) Sample /aɪ/-raising near-minimal pairs

/lʌɪf/ 'life' ~ /laɪv/ 'live'

/spʌɪs/ 'spice' ~ /spaɪz/ 'spies'

/trʌɪp/ 'tripe' ~ /traɪb/ 'tribe'

/brʌɪt/ 'bright' ~ /braɪd/ 'bride'

(4) Canonical raising before underlyingly voiceless segments

/raɪd/ 'ride' ~ /raɪɾɚ/ 'rider'

/rʌɪt/ 'write' ~ /rʌɪɾɚ/ 'writer'

(5) Lack of transparent raising before underlyingly voiceless segments

/raɪd/ 'ride' ~ /raɪɾɚ/ 'rider'

/rʌɪt/ 'write' ~ /raɪɾɚ/ 'writer'

This chapter, which reports on and extends the major results of Kodner and Richter (2020),

contributes to that discussion. we find that the presence of transparent /aɪ/-raising as well as its rare attestation and sparse distribution, can be accounted for with a model of child language acquisition in the face of variable phonological input. It casts transparent raising as a kind of contact phenomenon between two minimally distinct varieties of American English (that differ in as little as the presence or absence of /aɪ/-raising). Since few individuals learn from only a single source, variation is a normal part of the acquisition process, and since change is formally inevitable in the presence of variation (Niyogi and Berwick, 2009), this establishes one of the basic principles of acquisition-driven change.

This chapter begins with a summary of the distribution of transparent /aɪ/-raising and the role it has played in discussion of the origin of /aɪ/-raising more broadly in Section 4.1. Following that, I introduce our model for the acquisition of /aɪ/-raising in variable environments in Section 4.2. The model is applied in in Sections 4.3 and 4.4, and its implications for the transparent /aɪ/-raising problem, lexical raising, and acquisition-driven change more generally are described in Section 4.5.

4.1. The Historical Development of /aɪ/-Raising

The population-level development and spread of /aɪ/-raising has received particular attention because it seems to have occurred rapidly despite its apparent opacity, and communities containing any variants other than the non-raising and the canonical raising patterns of (3)-(4) have been very difficult to find. Joos (1942) described both canonical and transparent-raising individuals in his study of 1940s Ontario, however Chambers (1973) did not report transparent raisers in Canada three decades later. Canonical /aɪ/-Raising has been described in detail several times throughout much of the United States including Martha's Vineyard (Labov, 1963), Charleston and the Tidewater region (Chambers, 1973), Michigan (Dailey-O'Cain, 1997), the Inland North more broadly (Vance, 1987; Labov et al., 2005), and Philadelphia (Fruehwald, 2013), but transparent-raising has remained elusive until recently.

The general absence of the transparent pattern is surprising because it is often framed as

an initial step in the actuation of canonical /aɪ/-raising. Multiple phonetically-motivated pathways have been proposed which leverage *hypocorrection* (Ohala et al., 1981), *off-glide peripheralization* (Moreton and Thomas, 2007), *pre-voiceless shortening* (Joos, 1942), among others.[1] However, in a detailed longitudinal study of the Philadelphia Neighborhood Corpus (Labov and Rosenfelder, 2011), Fruehwald (2013, 2016, 2017) identified increasing prevalence of the raised /aɪ/ in the city's population over time without evidence for any transparent raising pattern, concluding that if any such intermediate stages existed there, they must have been extremely brief. Additionally, /aɪ/-Raising seems to have developed natively in Philadelphia rather than having been imported, since it lacks details such as pre-/r/ raising attested in the Inland North (Dailey-O'Cain, 1997). Taken together, this supports his conclusion that raising in Philadelphia was already a phonological process from its earliest attested stages.

However, using an acoustic production laboratory study, Berkson et al. (2017) have now found transparent raising for the first time since Joos (1942) among canonical and non-raising speakers in Fort Wayne, Indiana (Map 17). Berkson et al. (2017) argue that their findings support the hypocorrection account, intend to continue studying the local population, and posit that the transparent grammar, which has only been observed within the past decade or so, could be completely overtaken by the full raising grammar in as little as a generation.

How, then, might Berkson et al.'s findings be rectified with a phonological account of /aɪ/-raising? We propose that dialect contact is a much more likely source than spontaneously repeated phonetic incrementation and phonologization. In particular, the sudden recent appearance of both canonical and transparent raising in formerly non-raising Fort Wayne, which is not far south of the traditional /aɪ/-raising Inland North as described in the ANAE (Figure 17), is best described as a contact scenario. This contact would only mix canonical (fully allophonic) raising and non-raising speech at first, so the attested transparent raising would not be a result of gradient articulatorily-motivated phonetic incrementation, but

[1] see Fruehwald (2013) for a detailed discussion of these proposals' predictions.

Figure 17: Transparent /aɪ/-raising locations investigated by Joos (1942) and Berkson et al. (2017) superimposed on ANAE Map 14.10. The primarily canonical raising region extending from Western Canada through the Great Lakes into New England is bounded by the bright green line. The North and North-Central dialect regions are bounded by the blue and dark green lines respectively.

rather an instance of actuation resulting from the inconsistent learner input.

Following our proposal, transparent /aɪ/-raising emerges sporadically among native language learners in populations at the geographic frontiers of /aɪ/-raising, implying that children are capable of hypothesizing transparent raising given input that is some combination of raising and non-raising without exposure to transparent raising itself. We investigate this acquisition process with an extension of the Tolerance Principle (Yang, 2016), a model for the acquisition of productive patterns, to describe language learning in variable environments.

We clarify a number of points in our treatment of /aɪ/-raising: first, that the transparent raising pattern is not merely a 'partial' grammar with respect to raising or an incipient step on the way to canonical raising, but a distinct grammar in which raising is allophonic and conditioned by surface voicelessness. Second, raising patterns need not always arise

independently in new areas, since they are easily spread by contact and migration like many other regional dialect features of North American English (e.g., short-*a* tensing in New Orleans; Berger 1980; Labov 2007; or Northern Cities features in St. Louis; Labov 2007; Friedman 2014). The recent emergence of /aɪ/-raising in some North American communities should be viewed as a process of dialect contact, and does not necessarily indicate initial community-internal phonetic actuation of Canadian Raising. Third, the sparse individual attestation and lack of transparent raising communities can be accounted for by the limited range of linguistic environments in which the corresponding grammar is learnable. Finally, the *de novo* innovation of a distinct transparent raising grammar provides a study of children as innovators of change and an empirically testable comment on the classic actuation problem (Weinreich et al., 1968).

4.2. Learning from Mixed Input

A mixed input treatment of transparent raising is a claim about innovation: that a novel grammar, transparent raising, may be learners' best option given their input even if none of that input was generated by transparent raising. A learner whose input sample is drawn from both canonical and non-raising input is subject to variation at the lexical level because any given "raisable" word will be attested in the input as raised, not-raised, or both according to some probability. It is the learner's task to make sense of that distribution.

The acquisition of phonology begins early on in development. Children have a sense of the inventory of surface segments in their native languages by six months to a year (Werker and Tees, 1984; Kuhl et al., 1992), and they are capable of learning allophonic relationships between segments as well towards the end of that time period (Pierrehumbert, 2003a; Pegg and Werker, 1997). Although in a non-raising grammar, pairs like *writer-rider* in which /aɪ/ precedes an alveolar flap are pronounced with the same vowel quality and similar flap realization, they are reliably pronounced with a vowel length difference reflecting the voicing of the following underlying stop. Therefore, flapped /t/-/d/ word pairs have distinct pronunciations in both raising and non-raising grammars, allowing children to recover underlying

stop voicing for words pronounced with flaps. The vowel length information regarding underlying stop voicing must be cognitively available to learners fairly early, as it is already systematic in productions before 24 months; even before children are entirely competent with flap articulation itself, they reproduce the adult pattern of shortening before underlyingly voiceless flaps (Rimac and Smith, 1984; Ko, 2007).

In addition to the basic task of acquiring a native phonology, learners receiving variable input must be content with a scenario where no pattern clearly stands out as the target for acquisition. Notably, young children tend to regularize inconsistent input rather than accurately matching the probabilities of observed variants (Singleton and Newport, 2004). As children mature, their behavior gradually approaches that of adults, who are known to be (fairly) accurate at probability matching and can distinguish between structured, context dependent variation and random variation (Hudson Kam and Newport, 2005; Kam and Newport, 2009; Newport, 2020). The Tolerance Principle makes accurate predictions about children's tendency to generalize mixed input in a laboratory artificial language setting (Schuler, 2017).

4.2.1. Applying the Tolerance Principle to /aɪ/-Raising

We model the acquisition of /aɪ/-raising allophony with the Tolerance Principle (TP; Yang, 2016) building on the treatment developed in Sneller et al. (2019), which applies the Tolerance Principle to variation in the Philadelphia short-*a* system: if two variants of a lexical item are heard, the variant that is heard more often is the one acquired. If a child happens to hear variants generated by one grammar more often than the other, then that child's lexicon will tend to contain more vocabulary which matches that grammar. The TP can then be used to determine which input mixtures should drive children to acquire which grammar by evaluating the learnability of each grammar given each learner's own lexicon.

In their 2019 analysis of the short-*a* system in Philadelphia, Sneller et al. considered learners who were exposed both to the Philadelphia's unique system and the more general nasal

system (ANAE §13.2 for a summary of both systems). The Philadelphian system is a subset of the New York City system which raises /æ/ before front nasals and voiceless fricatives, along with a finite set of adjectives ending in /d/ (e.g., *mad* but not *sad*). The nasal system tenses /æ/ exactly before nasals. Over the past thirty years, younger Philadelphian speakers have been transitioning from the traditional Philadelphia system to the nasal system. The authors hypothesized that this was due to migration of speakers from nasal system regions into the Philadelphia area and that once a critical mass of nasal speakers had migrated, learners were more likely to acquire the nasal system than the traditional one. A learner whose input was drawn from both systems would learn some lexical items which correspond to the Philadelphia system and some which correspond to the nasal system. Items consistent with only one grammar would constitute exceptions to the other, so that the more prevalent one system was in a given learner's input, the more consistent the learner's lexicon would be with that system as opposed to the other. The Tolerance Principle predicted that once enough nasal system speakers relocated to Philadelphia, young learners would be led to hypothesize the nasal system given their mixed input.

/aɪ/-Raising can be treated in much the same way as the Philadelphia short-*a* system, but it provides a number of additional challenges. It amounts to an instance of innovation: transparent raising is thought to emerge *de novo* as some learners' best option given a particular proportion of canonical raising and non-raising input, so a learner would have to propose an unattested third grammar rather than choose either input grammar, unlike the short-*a* scenario. Canonical raising applies to every instance that transparent raising would, so there is no single input that would uniquely pick out transparent raising as opposed to a mixture of non-raising and canonical raising. As a result, transparent raising is only learnable when there is enough evidence in its favor but not enough evidence for canonical raising – there must be "enough" evidence for raising before surface voiceless segments but "not enough" evidence for raising before flapped /t/. This is technically possible in a mixed canonical/non-raising input environment if the learner happens to acquire more surface-voiceless raising items from raising speakers and more flapped /t/ words from non-raising

speakers. There is a sense in which a transparent raising hypothesis is less stringent because it does not depend on the behavior of flapped-/t/ items.

Cast in terms of the Tolerance Principle, a learner may hypothesize two raising grammars, a narrower one which amounts to transparent raising (g_{trans}) or a broader one that amounts to canonical raising (g_{can}), or resort to non-productive raising instead (g_{none}) if neither of the others is tenable. Canonical raising is tenable if there are few enough exceptions among $N_{\text{can}} = N_{\text{trans}} + N_{\text{flap}}$ items with flapped and surface /t/, and transparent raising is tenable for a learner who has learned few enough exceptions among N_{trans} surface /aɪt/ items but too many exceptions among the N_{flap} items for canonical raising. We follow Sneller et al. (2019) in modeling learners who acquire the most frequent variant of each item, since young learners regularize input variation. That is, if a child happens to hear raised 'writer' more often than non-raised 'writer,' that child will initially learn the former rather than the latter, and if the child hears non-raised 'spite' more often than raised 'spite,' the child will acquire the non-raised form, and so on.

This learnability pattern is visualized in Figure 18. There are two number lines of the same type as those introduced in 2.3, the top one for canonical raising which extends from 0 to N_{can} and the bottom for transparent raising which extends from 0 to N_{trans}. The line for canonical raising is longer than the one for transparent raising by N_{flap}. If the number of items that should be raised under canonical raising but were not ($e_{\text{can}} = e_{\text{trans}} + e_{\text{flap}}$) is low, below θ_{can}, then canonical raising is learnable (gold). If canonical raising is learnable, then it subsumes transparent raising since a grammar that applies canonical raising would account for all instances of transparent raising as well.[2] If neither canonical raising nor transparent raising is learnable (blue; $e_{\text{flap}} > \theta_{\text{trans}}$ and $e_{\text{can}} > \theta_{\text{can}}$ on the number), then a child should resort to non-productive raising and memorize those items learned to be raised but not

[2]This is a kind of super-set-subset relationship, but the challenge faced here is not exactly the same as the famous subset problem (Berwick et al., 1985) which asks how a learner that hypothesizes a grammar which is a super-set of the target could ever retreat to the subset. Instead, one of the available targets is the super-set (canonical raising), and we are asking under what conditions a learner would instead hypothesize a subset (transparent raising).

apply raising to other items. And crucially, if the number of underlying /aɪt/ exceptions falls above θ_{can}, but the number of surface /aɪt/ exceptions relevant for transparent raising falls below θ_{trans} (red), then transparent raising is tenable, but canonical raising is not.

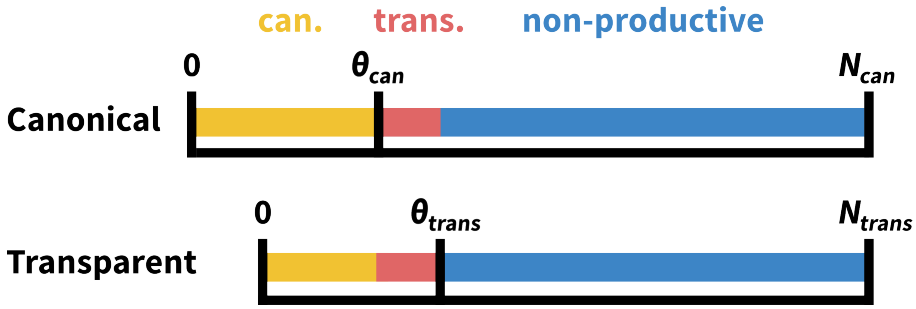


Figure 18: Range of exceptions in which canonical raising is learnable (gold), transparent raising is learnable (red), and neither is learnable (blue). Number lines are oriented such that the red zones line up. The red zone is the range in which exceptions fall above the tolerance threshold for canonical raising but below the threshold for transparent raising. The red zone is as wide as the number of potentially /aɪ/-raised flapped /t/ words in the vocabulary.

The learner can land in the red zone if, by chance, too many flapped /aɪɾ/ words were learned without raising to support canonical raising, but enough faithful /aɪt/ were learned raised to support transparent raising. The number of faithful exceptions (e_{trans}) must be low enough to pass under θ_{trans}, but must be high enough such that e_{trans} plus the number of flapped exceptions (e_{flap}) together exceeds θ_{can}. These bounds are written out in Inequalities 4.1-4.2. e_{trans} is directly dependent on θ_{can}, N_{flap}, and θ_{trans}, and on N_{flap} and N_{trans} indirectly through those, while e_{flap} is directly dependent on e_{trans} as well as θ_{can} and N_{flap}.

$$\theta_{\text{can}} - N_{\text{flap}} \leq e_{\text{trans}} \leq \lfloor \theta_{\text{trans}} \rfloor \tag{4.1}$$

$$\theta_{\text{can}} - e_{\text{trans}} \leq e_{\text{flap}} \leq N_{\text{flap}} \tag{4.2}$$

So it is conceptually possible for a learner to acquire transparent raising given only canonical and non-raising input, but how likely is it? The probability of falling into the transparent

raising zone is dependent on the proportion of canonical and non-raising input, which in turn influences the probability of any given lexical item being acquired as raised or non-raised. That probability can be thought of as a coin flip weighted by the proportion of non-raised input (p_{none}) in the environment.[3] The number of exceptions to a raising generalization is the number of coin flips that come out 'non-raising' rather than 'raising' after N trials, one for each item learned. The number of exceptions to canonical raising e_{can} is modeled for N_{can} coin flips, and the exceptions to transparent raising e_{trans} is modeled for N_{trans} flips, both weighted by p_{none}.[4] As with coin flips, whether a given word happens to be learned one way or the other is independent of how the other words were learned.

A sequence of weighted coin flips is modeled with a binomial distribution. The probability of e_{can} falling between 0 and θ_{can} is the probability that canonical raising is tolerable (the learner falls in the gold zone in Fig. 18), which is calculated with the binomial cumulative density function provided in Equation 4.3.

$$p(g_{\text{can}} \text{ tolerable}) = \sum_{e_{\text{can}}=0}^{\lfloor \theta_{\text{can}} \rfloor} \binom{N_{\text{can}}}{e_{\text{can}}} p_{\text{none}}^{e_{\text{can}}} \, p_{\text{can}}^{N_{\text{can}} - e_{\text{can}}} \tag{4.3}$$

The function for calculating the probability of learning transparent raising is more complex, but can be derived by plugging in the bounds from Inequalities 4.1-4.2. The goal is to calculate the probability that e_{trans} and e_{flap} fall within those bounds (Equation 4.4).

[3] Roughly, the proportion of non-raising speakers in the environment.

[4] The binomial distribution, which is used to model binary outcomes like coin flips, is weighted here by p_{none}, but equivalent results could have been calculated with weighting by p_{can}.

$$p(\text{learn } g_{\text{trans}}) = \sum_{e_{\text{trans}}=\theta_{\text{can}}-N_{\text{flap}}}^{\lfloor\theta_{\text{trans}}\rfloor} \left(\binom{N_{\text{trans}}}{e_{\text{trans}}} p_{\text{none}}^{e_{\text{trans}}} \; p_{\text{can}}^{N_{\text{trans}}-e_{\text{trans}}} \sum_{e_{\text{flap}}=\theta_{\text{can}}-e_{\text{trans}}}^{N_{\text{flap}}} \binom{N_{\text{flap}}}{e_{\text{flap}}} p_{\text{none}}^{e_{\text{flap}}} \; p_{\text{can}}^{N_{\text{flap}}-e_{\text{flap}}} \right) \tag{4.4}$$

At this point, we have defined the conditions under which transparent raising is learnable with the Tolerance Principle and laid out how to model the number of exceptions that a learner might be exposed to. The last ingredient needed before we can calculate the learnability of transparent raising is an empirical measurement of the lexicon. To approximate the relevant portions of a typical learners' lexicon, N_{trans} and N_{flap}, we take measures from the Brown (Brown, 1973) and Brent (Brent and Siskind, 2001) corpora of child-directed speech at varying frequency cutoff thresholds, using the methods described in Chapter 3. The vocabulary sizes of both frequency-trimmed lexicons roughly correspond to those of a child around school age, while the full-corpus lexicons reflect learners far past the age of basic phonological acquisition (Anglin, 1993; Nation and Waring, 1997). Therefore, these estimated lexicons contain a sufficiently complete representation of the input data that learners would acquire their /aɪt/-raising system from, and the convergent predictions we obtain from modeling the various lexicons are reliable predictions of acquisition outcomes (and input to subsequent learners). Estimates of N_{can} and N_{trans} ($N_{\text{flap}} = N_{\text{can}} - N_{\text{trans}}$) are summarized in Table 6.

Corpus	**Freq. Cutoff**	$N_{\textbf{can}}$ **(Types)**	$N_{\textbf{trans}}$ **(Types)**
Brown	≥ 5	53	45
Brown+Brent	≥ 5	82	69
Brown	≥ 1	122	103
Brown+Brent	≥ 1	182	155

Table 6: Number of potentially /aɪ/-raisable items by sample lexicon.

There are far fewer flapped /t/ items than surface [t] items in each sampled lexicon, so the

space in which transparent raising is learnable is present but relatively small. In other words, transparent raising is unlikely to be learned regardless of the mix of grammars that a learner's input is drawn from, which is consistent with the empirical evidence that transparent raising is rare – it should not emerge frequently in the grammars of young people.

4.3. Innovation of Transparent Raising

It is now possible to calculate the distribution of learning outcomes for any input distribution of canonical and non-raising input for any of the lexicon samples enumerated in Table 6. The results for the Brown+Brent ≥ 5 lexicon are shown in Figure 19. In this plot, the x-axis is the percent of input that is non-raising, with p_{none} increasing to the right. Since there is no transparent input at this stage, $p_{\text{none}} + p_{\text{can}} = 1$ and the proportion of input generated by a canonical raising grammar p_{can} increases towards the left. The y-axis is the probability that a learner exposed to the input mixture specified on x should acquire each raising grammar, that is, the probability that e falls into the gold, red, or blue zone for that learner's Tolerance Principle evaluations (Figure 18). With independence assumptions, the y values can also be taken to represent the proportion of learners who acquire each type of raising grammar for the specified input mixture.

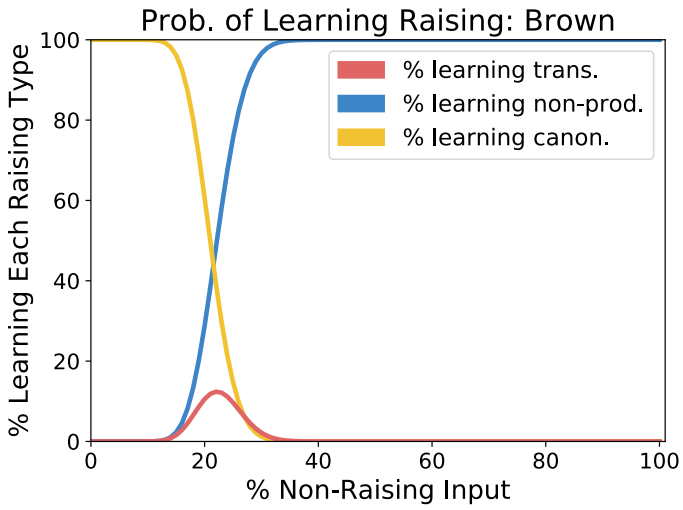


Figure 19: Learner outcomes by input mixture for the Brown+Brent ≥ 5 sample lexicon. The x-axis represents the proportions of non-raising and canonical raising input, and y-axis represents the probability of a learner acquiring each variant raising grammar.

It is clear from the plot that most learners should acquire either canonical raising or a non-raising grammar regardless of the input distribution to which they are exposed. Nevertheless, there is a narrow range, between about 15% and 30% non-raising input, where some learners do indeed acquire novel transparent raising, showing that it really can be innovated *de novo* in dialect mixing situations and from non-raising and canonical raising inputs alone. Calculations on the other lexicon estimates in Figure 20 yields convergent learning outcomes with the peak transparent raising rate consistently occurring between 20% and 30% non-raising input. The fact that learnability remains roughly constant regardless of estimated vocabulary size (the Ns estimated from Brown+Brent ≥ 1 are about three times larger than those from Brown ≥ 5 (Table 6) suggests that learners who innovate transparent raising early on will tend to stick with it as they mature, which is not always the case (Section 2.3 and Chapter 6).

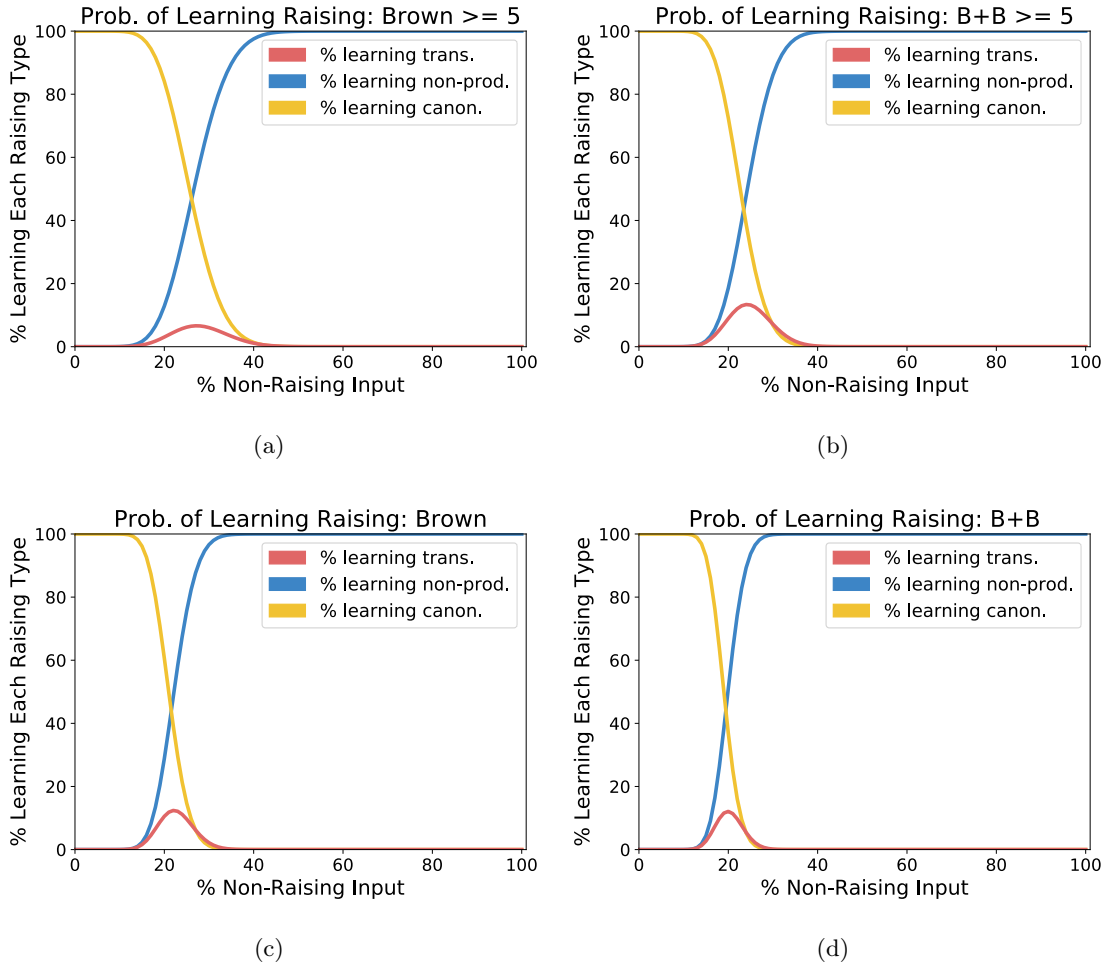


Figure 20: Learner outcomes by input mixture for each lexicon estimate. The x-axis represents the proportions of non-raising and canonical raising input, and y-axis represents the probability of a learner acquiring each variant raising grammar.

These results indicate that transparent raising can be innovated sporadically during contact between raising and non-raising varieties such as at the dialect boundaries where it has been attested in individuals (Figure 17). The next section considers how populations should evolve over time to answer the question of why no transparent raising *populations* have been observed.

4.4. Transparent Raising over Time

It is possible for transparent raising to emerge as a novel grammar at low rates in mixed populations of non-raising and canonical raising speakers. As those new transparent raisers mature, they will provide additional input to their younger peers, so future cohorts will receive transparent raising input as well as non-raising and canonical raising. Modeling how populations of non-, canonical, and transparent raising speakers evolve over time reveals why transparent raising speakers have been so sparsely distributed.

Learning outcomes for all possible three-way distributions of non-, canonical, and transparent raising input are provided in the ternary plot in Figure 21, with colors corresponding to those in Figures 18-20. Points closer to the bottom left represent higher proportions of transparent raising, those towards the top represent more canonical raising, and those towards the bottom right represent more non-raising input. The right edge of the ternary plot represents the circumstance where there is zero transparent input and corresponds to the Brown ≥ 1 plot in Figure 20.

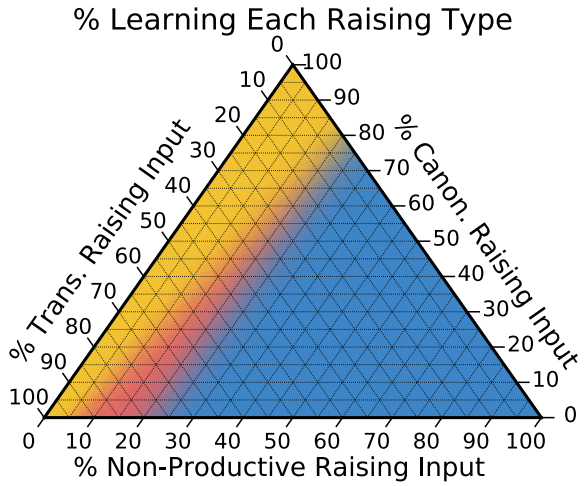


Figure 21: Learner outcomes by three-way input mixture for the Brown ≥ 1 sample lexicon. Colors corresponding to 18 (Blue: non-productive raising, Gold: canonical raising, Red: transparent raising) indicate relative proportion of learners acquiring each type.

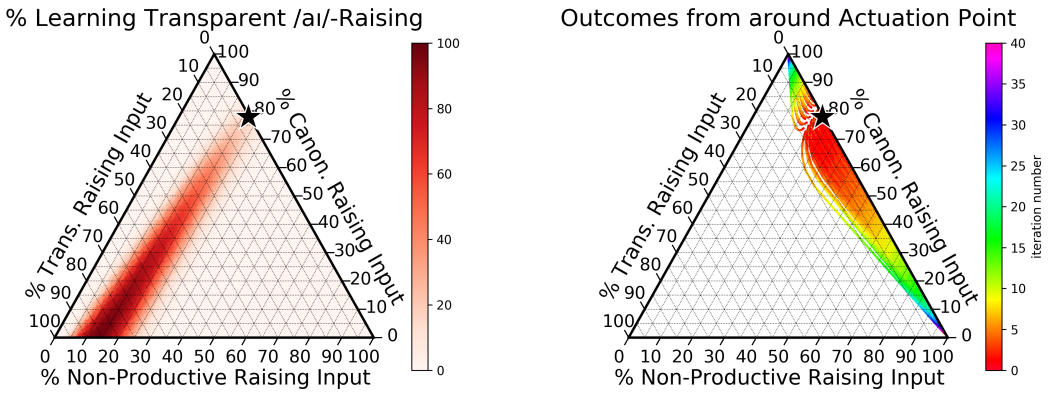
Along the right edge, the transition between canonical and non-raising dominance occurs at less than 80% canonical raising input (more than 20% non-raising input), which corresponds to the transition range in Figure 20. The gradient change in color indicates that no raising type is dominant exactly at this point. Transparent raising is only dominant in the narrow red band between the canonical and non-raising zones in the plot, and interestingly, is not dominant around 100% transparent input. This warrants further discussion in Section 4.5.

The narrow range over which transparent raising is feasibly acquired has implications for its population-level viability, visualized in Figure 22. The left ternary plot focuses on transparent raising's band of viability with darker shades of red indicating a higher probability of acquiring the transparent pattern as opposed to the other options. The band is wider and darker near the bottom left of the plot, and the point at which transparent raising is most likely to be innovated (corresponding to the peaks in the Figure 19-20 plots) is indicated with a black star. One question then is whether the deep red regions of the space are actually reachable from the point of initial innovation.

We set up an iterative simulation to test what would happen if a population were to evolve from the innovation point. Since slightly different initial conditions may yield vary different outcomes in such a dynamical system, several populations were initialized around the most likely innovation point indicated by the black star in Figure 22. These populations were each allowed to evolve iteratively: the usual calculation was run to estimate the learning outcomes of an initial cohort g_0, then that cohort's output was used as the input for the next p_1. This was repeated for forty iterations for each initial state, and the results were plotted. Each iteration's input was a weighted average of the previous iteration's input and output $p_t = 0.9p_{t-1} + 0.1g_{t-1}$ to reflect the fact that only a fraction of a given population is acquiring its native phonology at any given time.[5] The right plot in Figure 22 visualizes these populations' evolution over time, with rainbow coloring indicating iteration number. In every case, the population rapidly falls off the band of viability and progresses toward

[5]However, even a 100% update ($p_t = g_{t-1}$) yields the same final outcomes.

100$ canonical or 100% non-productive raising depending on initial conditions.



(a) (b)

Figure 22: a) Proportion of learners acquiring transparent raising (depth of red) for each three-way input mix. Transparent raising is only viable on a narrow band of input mixtures, indicating that specific input conditions are needed to support it. b) Phase plot showing the rapid defeat of transparent raising. The community always shifts away from the conditions in which transparent raising is actuated (around the black star) towards canonical or non-raising (top or right corners) after several iterations. Rainbow coloring indicates iteration number and progresses from red through to magenta.

No primarily transparent raising population ever has a chance to develop. This can explain why populations of transparent raisers have never been observed, even at dialect boundaries. Transparent raising is self-defeating in practice, and can only exist sporadically.

4.5. Discussion

The computational model laid out here accounts for the sporadic innovation of transparent /aɪ/-raising as phonological acquisition in a mixed canonical and non-raising environment. This has several implications, from narrow to broad, for the distribution of and relationship between the raising patterns, for the position of lexical raising, and for the role of 'monolingual' acquisition in language change.

4.5.1. Patterns of /aɪ/-Raising

This approach recognizes the plausible distinction between how canonical /aɪ/-raising first arose and how it has since spread, giving it a number of advantages over previous treatments. It accounts for its distribution at dialect boundaries such as 1940s Ontario and 2010s Indiana without casting transparent raising as necessarily an incipient form of /aɪ/-raising. Furthermore, it offers an explanation for why no populations of transparent raisers have been found: it is innovated stochastically in certain variable environments, but cannot gain a foothold in a speech community. This has implications for Fruehwald's observation that Philadelphia never passed through an incipient transparent raising stage on its way from non-raising to canonical raising and the hypothesis that raising developed in the phonology from its earliest stages. Since the presence of transparent raisers need not imply phonetic raising, the presence of transparent raisers in the Berkson et al. study does not necessarily conflict with Fruehwald. On the other hand, it suggests that some individual transparent raisers should have existed in Philadelphia in the time span covered by Fruehwald's studies who may have been missed by his population-level statistical analysis. An individual-by-individual investigation, such as the laboratory experiments undertaken in Fort Wayne, might be a necessary supplement if we want to find more transparent raisers going forward.

The model makes a number of testable quantitative predictions regarding the relationship between phonological input and change in progress. It suggests where researchers should look for additional transparent raisers – synchronically at dialect boundaries, or diachronically in corpora such as the PNC at the points in time when the community was mixed – and importantly it gives the ratio of non-raising and canonical raising input at which transparent raising may be innovated in a given community. A comparison to Berkson et al.'s experimental results on young adults provides very rough validation and suggests a path for validation in the future. They find that at the time of study, 37% of their 27 young adult participants exhibited Pattern 0 or Pattern 1 (in our terms, non-productive raising), 33% exhibited Pattern 2 (transparent raising), and 30% exhibited Pattern 3 (canonical raising).

Our model predicts that a speech community that has just innovated transparent raising, roughly twenty years previously in Fort Wayne, would exhibit the three grammars at 48%, 13%, and 39% respectively. This is reasonably in the ballpark given that the computational study of learners and laboratory study of young adults are not directly comparable.

4.5.2. Lexical Raising

In addition to the rule-governed raising pattern discussed in this chapter, the introduction of the Tolerance Principle to this problem has additional implications for lexical or "exceptional" raising, that is, the circumstance where a finite set of items is /aɪ/-raised even though they do not follow one of the raising patterns. Lexical raising could conceivably be implemented with the raised diphthong in underlying forms, which renders /ʌɪ/ phonemic, or through a rule which applies to a list of items. Some commonly cited examples include *high school* as ['hʌɪ.skul] when the speaker shows no other raising across word boundaries[6] (Chambers, 1973) or raising in a small set of items including *spider* and *tiger* (Fruehwald et al., 2009; Fruehwald, 2013). Conceptually, lexical raising can refer to two phenomena: non-productive raising with only listed raised items, and productive raising with additional listed raised items.

The former, non-productive raising, is a special case of non-raising in which the speaker does raise a few items but overall lacks a raising grammar. In the terms of the Tolerance Principle and our variable application, non-productive raising speakers can simply memorize any items that they happen to have acquired as raised. A child exposed to both non-raising and canonical raising input who fails to acquire a raising pattern is likely to have acquired some amount of lexical raising, since any item that the child happened to hear raised more often than non-raised could be memorized that way. This is why non-raising is referred to as non-productive raising throughout most of this chapter.

The latter, productive raising plus additional raised items, is just a special case of canonical raising which seems to be well-attested. For example, raising of *Snyder*, *spider* and *cider*

[6] My former officemate raises this word but does not have productive raising (Caplan, p.c.)

is attested in the PNC starting around 1940, about a generation after canonical raising is first observed in Philadelphia (Fruehwald et al., 2009; Fruehwald, 2013). Though it is possible that these speakers have underlying /t/ in these words, and that the orthography is misleading, that would not answer why they would have underlying /t/ in the first place. It should also be noted that these forms contain /r/ later in the word, and /aɪ/-raising is known to occur before /r/ in much of the Northern United States. It is possible that some Philadelphia learners were exposed to a limited amount of pre-/r/ raising and either acquired it lexically, or reinterpreted some of those items as containing underlying /t/, or these forms may have been innovated in Philadelphia without outside influence. In any case, the presence of these items is well within the purview of our class of model.

At this point, I would like to return to the bottom left corner of Figure 21, which shows that if nearly 100% of input is drawn from a transparent raising grammar, a learner should actually acquire canonical raising. This is because the number of flapped items, those which are not raised by a transparent grammar, is smaller than the tolerance threshold for canonical raising, so *if* a learner were to hypothesize canonical raising in a primarily transparent raising environment, it would be tolerable. This actually provides a mechanism for transitioning from a narrower pattern to a broader one. It does not match the empirical facts for canonical raising, but it would be interested to consider what other phonological changes may be explained in this way, perhaps a transition from canonical raising to Philadelphian canonical plus *spider*-raising to Northern American canonical plus pre-/r/-raising.

4.5.3. Change and "Monolingual" Acquisition

Transparent raising is described here as a kind of contact phenomenon between two minimally distinct dialects which differ only in the presence of /aɪ/-raising. In the most formal sense, two varieties that differ in even a single feature are different languages, and in that sense, no two individuals actually speak the same language. In other words, minimal variation is ubiquitous and unavoidable. All normal language acquisition has to contend with variation of some kind, even in settings typically conceived of as "monolingual." Individu-

als' experiences differ in how dramatic the variation manifests, from "monolingual" minimal variation that likely lies below the level of social awareness, to dialect contact above the level of awareness, to full-blown contact scenarios as traditionally conceived between two or more distantly or unrelated languages.

In that sense, monolingual acquisition and traditional language contact are more two ends of a spectrum than truly distinct phenomena. As such, it is likely that many of the principles outlined for transparent /aɪ/-raising apply through that spectrum. For example, the finding that change is a formally inevitable result of categorical grammar learning and population-level variation (Niyogi and Berwick, 1996) holds no matter how trivial the variation. A future research program may seek to describe this spectrum, building on work such as this dissertation and aspects of variationist sociolinguistics at the low end and creole linguistics at the high end, where some have already proposed a distinction between creoles in the proper sense and less distinct "semi-creoles" (e.g., McWhorter, 2002). Finally, if variation, to whatever degree, is always present, then it can serve as a critical component in a model of (even monolingual) acquisition-driven change (Sections 2.2.1 and 6.6.2). It challenges the claim that acquisition-driven change cannot occur without early bilingualism (Meisel, 2011).[7]

In the next chapter, I will transition from an innovation that fails to gain a community presence to one that was fully actuated in the distant past. The learners who innovated the *to*-dative in Middle English were able to spread it to their local communities, and it eventually spread to fixation as a standard construction in English syntax.

[7]Or at least early multi-dialectism, still as opposed to traditionally conceived monolingualism (Meisel, 2011, fn. 3)

CHAPTER 5 : The Middle English *to*-Dative

This chapter investigates the historical development of the English *to*-dative ('*Alice sent the message to Bob*'). The dative constructions available to Old English were notably different from the modern ones. Old English had no *to*-dative (Mitchell, 1985), but had a *symmetric* double object (both IO-DO '*Alice sent Bob the message*' and DO-IO '*Alice sent the message Bob*' are well-attested), while Modern English has an asymmetric double object and does have a *to*-dative. Evidence from Middle English shows that the extension of the *to*-dative was nonlinear. It first came into use sometime by the beginning of the Middle English period and rapidly expanded in its usage frequency and lexical distribution until it exceeded that of Modern English, eventually being attested with verbs including *ask*, *save*, *forbid*, and *forgive*. Only later did the distribution of the *to*-dative retreat to its modern pattern. At the same time that the *to*-dative was spreading, the symmetric double objects was lost as DO-IO double objects gradually fell out of use.

The development of the *to*-dative is well-situated as a case study. Its surface expressions are well attested in historical corpora including the Penn Parsed Corpus of Middle English (PPCME2; Kroch and Taylor, 2000) and Penn Parsed Corpus of Early Modern English (PPCEME; Kroch et al., 2004), the synchronic representation and diachronic development of the construction's structures has been thoroughly investigated (Larson, 1988; Pesetsky, 1996; Allen, 1995, etc.), and the process of their acquisition has been studied extensively (Pinker, 1989; Gropen et al., 1989; Goldberg, 1995; Rappaport Hovav and Levin, 2008; Yang, 2016, etc.). I relate the innovation of the *to*-dative to transient errors in modern child-produced speech: individual children exhibit a pattern of generalization and retreat that mirrors that undertaken by English over a much longer time span, and produce overuse errors with both the *to*-dative ('I asked this to you.') and double object ('Jay said me no.') despite not receiving such constructions in their inputs and before eventually settling on adult-like grammars. In doing so, I begin to develop a "learners' perspective" framework for the relationship between acquisition and change which is further expanded on in subsequent

chapters.

There are quite a few moving parts that fit together in this chapter: Section 5.1 begins by discussing relevant synchronic aspects of the *to*-dative and double object constructions. I adopt Rappaport Hovav and Levin's 2008 conception of necessary broad-range semantic classes here. The main takeaway is that in Modern English, the *to*-dative requires that verbs belong to one of two classes: CAUSED POSSESSION or CAUSED MOTION, while the double object requires CAUSED POSSESSION.

Next, Section 5.2 summarizes how the constructions have changed over time from the perspective of which kinds of verbs could participate in them. Cast in terms of broad-range classes, Old English CAUSED POSSESSION verbs did not satisfy the condition necessary to introduce objects with *to* like they do now. That is, *to* could not introduce recipients. The true *to*-dative came about by the beginning of the Middle English period and rapidly rose in *token* frequency over the course of that era. It also rose in *type* frequency indicating an expansion of productivity. The rise in type frequency is not simply the result of increased attestation since there are even recorded instances of the *to*-dative used with verbs that no longer support it in Modern English.

I summarize some of the wide body of literature that child language acquisition brings to bear on the dative constructions in Section 5.3. Research in language acquisition has largely focused on how children learn the lexical distributions of the constructions, and I argue that this is the synchronic analogue to the diachronic problems. I then turn to applying synchronic findings to the *to*-dative's diachrony, beginning with its innovation in Section 5.4, I propose that surface-ambiguous constructions which introduced goals with *to* led child learners to hypothesize a *to*-dative in the grammar which introduced recipients. In terms of broad-range semantic classes, these children can be seen as having realigned CAUSED POSSESSION in their internal grammars to satisfy the necessary condition for the construction. In support of the plausibility of this account, I show that modern child production errors are consistent with this kind of novel mapping between syntax and semantics leading to over-

generalization and that the Middle English texts contain the kinds of surface-ambiguous utterances that could yield it in the first place.

The realignment and over-generalization account aims to provide an explanation for how the *to*-dative originated, but it alone does not explain its eventual lexical distribution. For that, I consider what would play out in a small population of learners obeying the Sufficiency Principle in Section 5.5. Applying the SP iteratively to lexical items extracted from the Penn Parsed Corpus of Middle English (PPCME2; Kroch and Taylor, 2000), I show that a generation in a single area could have reach the historically attested Middle English distribution in a short period of time. Next, I include Early Modern English vocabulary and show that the same learning model applied again effects the *to*-dative's lexical retreat. Section 5.6 surveys previous accounts for the innovation of the *to*-dative and weighs their deficiencies. Finally, Section 5.7 contrasts this case study with previous approaches to the problem and discusses the broader implications of addressing problems of language change primarily from the perspective of child language acquisition.

5.1. The Dative Constructions

Modern English has a pair of ditransitive constructions traditionally called the *dative alternation*: the DOUBLE OBJECT ('Alice sent Bob a message'), and the *to*-DATIVE ('Alice sent a message to Bob'). At a high level, the double object's first argument, the indirect object (IO), fills the role of a goal or a recipient, and its second argument, the direct object (DO), is a patient. In the *to*-dative, the goal or recipient instead follows the direct object and is introduced by the preposition *to*. The traditional name 'dative' refers back to the morphological case which marks recipients in languages like Latin and German, and 'alternation' refers to the relationship between the two constructions in some structural accounts.

Broadly speaking, syntactic analyses of the dative alternation follow two strategies: *Two-derivation* or *alternate projection* accounts propose distinct but often closely related derivations for each construction (Jackendoff, 1990; Pesetsky, 1996; Harley, 1997, 2002; Bruening,

2010, *inter alia*), and *single-derivation* or *transformation* analyses that relate one construction to the other via movement (Chomsky, 1975; Baker, 1988; Larson, 1988, *inter alia*). Two- and single-derivation accounts make different predictions about acceptability, and the debate over them has yielded decades' worth of theoretical literature. Rather than differentiating between these analyses, the present work seeks to uncover why each construction is possible at all and for which verbs in the language regardless of how they are represented in the grammar.

More relevant to the discussion here are the semantic constraints on verbs' participation in each construction. An effective treatment specifies *broad-range* classes which contain all the verbs associated with the constructions (Mazurkewich and White, 1984; Pinker et al., 1987; Gropen et al., 1989). The double object constructions support two broad-range classes, CAUSED POSSESSION (CP), and CAUSED MOTION (CM). This class may be subdivided into *verb-sensitive* (Levin, 2008; Rappaport Hovav and Levin, 2008) and *uniform multipe meaning* (UMM) (Pinker, 1989; Goldberg, 1995) approaches. I mostly adopt the verb-sensitive treatment, though the historical analysis explicated below could also be expressed in terms of UMM.

Under a verb-sensitive account, the *to*-dative requires either CP (e.g., *give*) or CM (e.g., *throw*), while the double object specifically requires CP as illustrated in Figure 23. These classes are closely associated with the recipient and goal thematic roles that all the structural analyses require and so work together with syntactic structure to dictate which verbs can participate in each construction. The classification applies both to concrete possession and movement as well as abstract, so as Gropen et al. (1989) notes, this accounts for the grammaticality of verbs like *tell*, *teach*, and *write* with the double object because they can be interpreted as transferring abstract possession of thoughts or ideas.

Under the broad-range analysis, *give* (CP only) and *throw* (CP or CM) support both constructions because their broad-range classes satisfy the necessary conditions for both (Figure 24). However, broad-range classes cannot be sufficient conditions because membership does

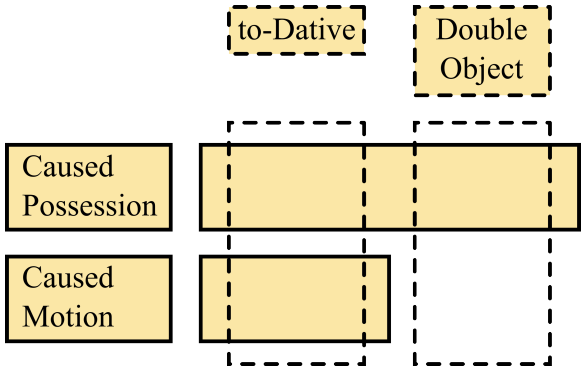


Figure 23: The relationship between broad-range classes and dative constructions in Modern English. The constructions are represented by columns and the semantic classes by rows. An intersection between a class and a construction indicates that that class is a necessary condition for that construction.

not guarantee grammaticality. For instance, many Latinate verbs such *donate* cannot form double objects even when they are CP verbs (Oehrle, 1976; Mazurkewich and White, 1984; Levin, 1993), and as Gropen et al. (1989) notes, neither can verbs like *say* (Figure 25). Research into these possibly arbitrary patterns is a major focus of language acquisition research and so is summarized along with the other acquisition aspects of the constructions in Section 5.3.

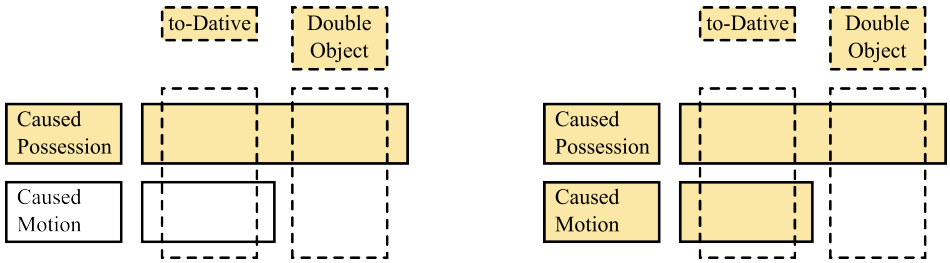


Figure 24: Broad-range semantic classes and dative constructions for *give* and *throw*. Highlighting indicates to which classes a verb belongs and which constructions are possible for that verb.

5.1.1. The Distribution of to-Datives in Germanic

The English *to*-dative and double object have analogues across modern Germanic with a number of frequent differences. First, the analogous constructions outside of English are formed with different prepositions: cognates of German *an* in Continental West Germanic

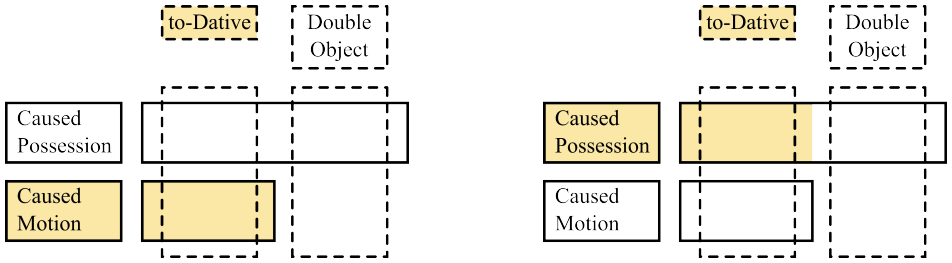


Figure 25: Broad-range semantic classes and dative constructions for *say* and *donate*. Highlighting indicates to which classes a verb belongs and which constructions are possible for that verb.

and cognates of Icelandic *til* or *að* in North Germanic. Second, several languages permit DO-IO double objects (in which the direct object precedes the indirect object as in **Alice gave the book Bob*) at least for pronouns. This is licit for some speakers in the English Midlands (MacKenzie and Bailey, 2016; Biggs, 2015), while in Swedish, it is only acceptable for certain particle verbs (Lundquist, 2014). Third, several modern varieties in North and West Germanic retain overt dative-accusative distinctions on nouns or definite nouns and pronouns, which may provide unambiguous evidence for object roles independent of surface word order in double object constructions. Figure 26 maps the synchronic distributions of the *to*-dative, DO-IO double object, and overt morphological dative-accusative distinction.

High German varieties retain overt dative-accusative marking, have *to*-dative, and support DO-IO (Seiler, 2003; Bacovcin, 2017). Midlands English, Low German, and Swedish lack overt case marking, employ *to*-datives, and DO-IO at least with pronouns or certain verbs (Lundquist, 2014; Biggs, 2015; MacKenzie and Bailey, 2016; Bacovcin, 2017). Faroese, some varieties of Norwegian, and some varieties in Sweden including Elfdalian retain overt case marking, have *to*-datives, and lack DO-IO (Dahl, 2009; Garbacz, 2010; Åfarli and Fjøsne, 2012; Lundquist, 2014). Standard English, Dutch, Frisian, and continental North Germanic varieties not already mentioned have *to*-datives but lack overt case marking and DO-IO (Tiersma, 1985; Dahl, 2009; Lundquist, 2014), and finally, Icelandic has overt case marking, lacks a *to*-dative, and supports DO-IO (Thráinsson, 2007).[1]

[1]The classification described here indicates grammaticality at least with pronominal objects and does not

Most possible combinations of the three variables are attested somewhere in Germanic. And in addition to these general grammatical features, individual languages vary lexically. For example, while THROW-type verbs permit the double object in English ('Alice threw Bob the ball'), they are illicit in Norwegian. A successful account of the constructions' development in English needs to be compatible not only with English diachrony but also with the synchronic cross-linguistic disconnect between the constructions and overt case marking, the languages' independent choice of preposition, and arbitrary lexical variation. In the discussion in Section 5.7, I return to this point and address prior work in light of these facts.

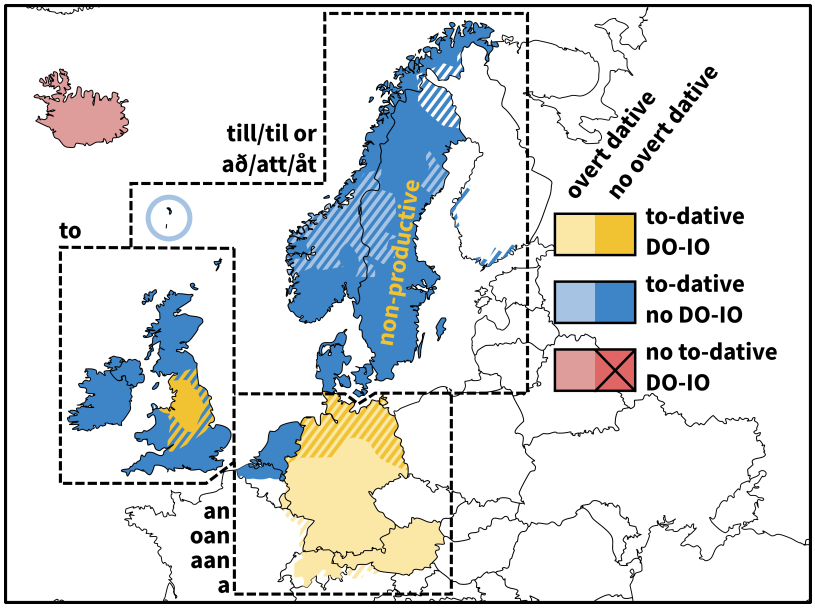


Figure 26: The distribution of DO-IO double objects, *to*-datives, and overt morphological dative-accusative distinctions in modern Germanic. Compiled from data published in Tiersma (1985); Seiler (2003); Thráinsson (2007); Dahl (2009); Garbacz (2010); Åfarli and Fjøsne (2012); Lundquist (2014); Biggs (2015); MacKenzie and Bailey (2016) and especially the appendices of Bacovcin (2017).

5.2. The English Dative Constructions over Time

The dative alternation at the end of the Old English period was different from the current alternation in a few key ways. First, the double object was *symmetric* in that both the

imply high frequency of use.

modern IO-DO ('Alice sent Bob the message') and the now-ungrammatical DO-IO (* 'Alice sent the message Bob') were licit (Mitchell, 1985; Allen, 1995; Polo, 2002, *inter alia*). Second, the *to*-dative (that is, where *to* may introduce recipients) was probably absent from the language (Mitchell, 1985, §1210). And third, the language made an overt morphological distinction between the accusative and dative cases on nouns and pronouns. For most verbs, the direct object was marked with the accusative case and the indirect object with the dative, though there were plenty of exceptions (Allen, 1995, p. 29). Over the course of the Middle English period, the constructions approached their modern form and distribution: the *to*-dative as we know it increased rapidly and dramatically in frequency at the same time as the DO-IO double objects declined in use (McFadden, 2002; Polo, 2002), and less discussed, the *to*-dative also expanded in its lexical distribution such that Middle English overshot the modern English system and permitted the *to*-dative with verbs that forbid it in Modern English (Visser, 1963), including *ask*, *forgive*, and *save*. The remainder of this section reviews the diachronic descriptive facts regarding the dative constructions and existing accounts for their histories.

Double Objects

The symmetric double object orders (6-7) are attested with roughly equal frequency in late Old English and early Middle English. Allen (1995, pp. 48-50) finds that they appear at the relative rates of 54% accusative-dative (mostly DO-IO) and 46% dative-accusative (mostly IO-DO) across 139 instances of the construction, and Koopman (1994) corroborates with 46% DO-IO and 54% IO-DO in a larger corpus of 1,987 instances.[2] Over the course of the Middle and Early Modern English periods, this roughly equal distribution shifted in favor of the IO-DO double object and *to*-dative.

(6) Old English DO-IO (Polo, 2002)

[2]It is important to note here that these are counts of surface forms and cannot be interpreted as the frequencies of any underlying structure because, surface DO-IO can result from multiple derivations (McFadden, 2002).

... þæt he forgeafe godne willan þam seocan hæðenan
... that he.NOM granted good.ACC will.ACC the.DAT sick.DAT heathen.DAT
'that he would grant good will to the sick heathen' (ÆCHom ii.2.12.28)

(7) Old English IO-DO

... gif þu geoffrast Gode ænige lac æt his weofode.
... if you.NOM offer God.DAT any.ACC .ACC at his altar.DAT
'if you offer God any sacrifice at his altar' (ÆCHom 16.19)

to-Datives

The preposition *to* indicated physical or abstract motion towards a goal in Old English, and it is not unusual to find it introducing an abstract goal with verbs of speaking, including *secgan* 'say, speak, declare,' *cweþan* 'say, speak, name, proclaim,' *sprecan* 'speak,' and *cleopian* 'cry, call' (Mitchell, 1985, §1210).

(8) Old Engish (Taylor, 2007)

... and hu miht þu **secgan to** ðinum breðer þus:
... and how might you say to your brother thus
'... and how might you say to your brother thus'

(coaelhom,+AHom_14:146.2080)

Old English *to* is also attested with the goals of verbs such as *bringan* 'bring,' *niman* 'take,' *lætan* 'permit, remain,' and *sendan* 'send.' The language apparently did not permit *to* with *give* and similar recipient verbs under normal circumstances, but Mitchell (1985) points out a few attestations of *give to* and *sell to* when the recipient is a plausible goal, for example '*agifan* to a monastery,' '*(ge)sellan* to a church,' and even 'to shame.' Two of these are plausible goals because they refer to literal locations, the recipient readings coming from metonymy. These two sources of *to*-constructions were apparently quite common in the language – as De Cuypere (2015) uncovers in a corpus study, about 15% of constructions that could have been expressed with a direct object are instead expressed with *to*, which is just a bit lower than the 21% rate expressed in a Modern English corpus.

So did Old English already have a *to*-dative after all? Probably not, because all the verbs which typically allow *to* in Old English fit into Rappaport Hovav and Levin (2008)'s broad-range CAUSED MOTION class, that is their objects are goals and not recipients, so the Old English system may be visualized as in Figure 27.[3] The examples of metonymy then can be thought of as 'recipient-like' goals that can be construed like recipients. In any case, Visser (1963: §687) notes that these examples come from late Old English, so if they are really *to*-datives, the innovation of the *to*-dative is just pushed back a few generations.[4]

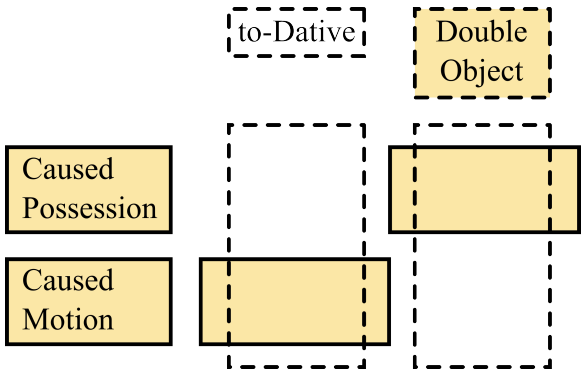


Figure 27: The relationship between Old English broad-range classes and the dative constructions. Contrast with Figure 23 for Middle and Modern English. The construction with *to* does not support caused possession and therefore is not a *to*-dative.

Regardless of their representational status in Old English, *to*-datives expanded rapidly in the early Middle English period, both in its *token frequency* (number of usage instances) relative to double objects (McFadden, 2002), and *type frequency* by spreading to a wider range of lemmas over that period. The rise in type frequency cannot be attributed simply to the increase in token frequency because the construction is periodically attested in Middle

[3]De Cuypere (2015) does not make this terminological distinction and so argues that *to*-datives were already quite common in OE. As far as I understand De Cuypere's position, we are actually in agreement about the facts, just not the interpretation.

[4]A similar pattern occurs in Icelandic, which, like OE, is described as lacking a true *to*-dative but allowing abstract motion with 'say.' Icelandic 'give' can occur in a *to*-dative-like construction only if its indirect object is sufficiently goal-like such as in cases of metonymy (Thráinsson, 2007).

(1) Ég gaf bækurnar til Háskólabókasafnsins
'I gave the books to the University Library.'

This presents a potential problem for Rappaport Hovav and Levin's (2008, p. 140) contention that GIVE-type verbs never satisfy caused motion. If that were true, it would not be possible for Icelandic and Old English, which lack a *to*-dative, i.e., a caused possession *to* construction, to show sensitivity to the 'goalhood' of their indirect objects in this way.

English with verbs which no longer support it in Modern English. Examples exist both with verbs that still support the construction (e.g., *...I shal ȝive* ***to thee*** *a coroun of liif*) and those which no longer do (e.g., *Commaunde ȝe* ***to the puple****, and saye ȝe...*). Visser (1963; §687) presents a large number of such examples which he deems 'unidiomatic' in the modern language, and notes that these aberrations were largely resolved by around 1500. Several more of these 'over-extensions' of the construction are identifiable in the Penn Parsed Corpus of Middle English 2 (Kroch and Taylor, 2000) with *save*, *ask*, and *forbid*, among others. While these instances are not particularly common, they are present both in texts translated from original French and ones which were not. The advance of the *to*-dative to this state and its subsequent reversal and retreat to the modern state both require explanations.

Figure 28 summarizes the changes that Middle English Middle English underwent. The *to*-dative was innovated, amounting to a realignment of broad-range semantic classes, around the same time, the DO-IO double object fell out of use, though it is once again present in certain Modern English dialects (cf. Biggs, 2015). The language lost its overt morphological dative-accusative distinction, and the *to*-dative was temporarily extended to far more verbs than can support it now. It should be noted that there was dialectal variation in the temporal ordering of these occurrences, as further discussed in Section 5.6.

5.3. Learning the Constructions

Research into the acquisition of the dative constructions has largely focused on the apparent arbitrariness of their distribution (Bowerman, 1983; Fodor, 1985; Gropen et al., 1989; Pinker, 1989; Ambridge et al., 2008; Yang, 2016, etc.). As described in Section 5.1, broad-range classes provide necessary conditions for verbs' grammaticality in the constructions, but not sufficient conditions, and the exact patterns by which verbs participate in each construction are language-specific and must be learned. For example, Norwegian has a double object and a *to*-dative like English but differs in that it prohibits the double object with its equivalent of THROW-type verbs while English allows them (Barðdal et al., 2011). This section provides

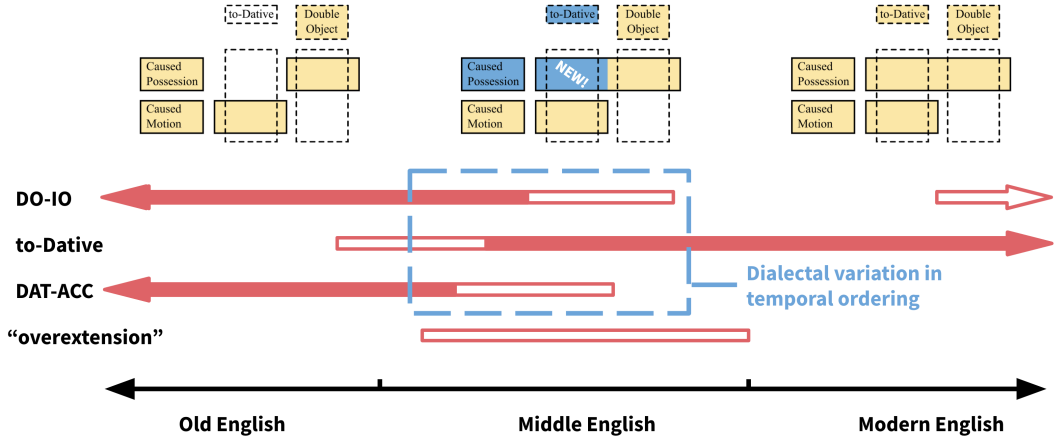


Figure 28: Schematic timeline summarizing the changes that occurred in Middle English. DO-IO fell out of use, the *to*-dative came into use (there was a broad-range semantic realignment), and the overt dative-accusative distinction was lost. An "over-extension" of the *to*-dative occurred for a time as well.

a brief overview of what is known about the acquisition of the dative constructions. I apply the Sufficiency Principle in Section 5.5 to model the *to*-dative's historical development.

5.3.1. Background

A plethora of semantic and morphophonological conditions can be posited to account for the distribution of the double object and *to*-dative, but simply enumerating these and positing them innately is a non-starter because, along with cross-linguistic variation in these conditions, each condition is subject to many exceptions that cannot be predicted. For example, Latinate verbs (e.g., *address*, *demonstrate*, *distribute*, *recommend*, *refer*, and *transfer*) are regarded as exceptions to caused possession because they prohibit the double object (Levin, 1993, p. 46) even when semantically similar Germanic verbs permit it (9). However, these exceptions have exceptions themselves, and several Latinate verbs including *advance*, *refund*, *allocate*, *allot*, *concede*, *extend*, *guarantee*, *offer*, *promise*, *refund*, *render*, and *telephone*, actually allow it (10). A few of them even require the double object and forbid the *to*-dative, including *imagine*, *nominate*, *presume*, *profess*, *refuse*, and *suppose* (Levin, 1993, pp. 45-47). To make matters more complicated for the learner, there even appears to be individual

variation (Fellbaum, 2005).

(9) Germanic/Latinate Doublets (Gropen et al., 1989, p. 206)

a. John told/reported the news to Bill.

b. John told/*reported Bill the news.

c. Kate showed/demonstrated the technique to Alan.

d. Kate showed/*demonstrated Alan the technique.

(10) Latinate Double Objects

a. Bob guaranteed Alice the report by the end of the week.

b. Alice assigned Bob another task.

c. Bob promised Alice a better cake.

d. Alice extended Bob an offer and Bob telephoned his friend.

e. Bob advanced Alice some money and Alice later refunded him.

Acquiring these lexically specific patterns is a daunting task. Strict lexical conservatism is unworkable because even many common verbs, let alone high-register Latinate ones, are unlikely to appear in all their possible constructions in the first few years of a child's life, so children must use their linguistic knowledge to make generalizations instead (Bowerman, 1983; Pinker, 1989, etc.), an ability that has been demonstrated in experimental studies with novel verbs (Gropen et al., 1989; Conwell and Demuth, 2007), yet children must remain moderately conservative so that they do not wildly generalize the constructions past where they belong (Pinker, 1989; Baker, 1992; Ambridge et al., 2008; Bowerman and Croft, 2008). This is the classic problem of Baker's Paradox (Baker, 1979).

Children can leverage their innate knowledge of semantics through broad-range semantic classes and their corresponding thematic roles, but that is only part of the solution. They still need finer-grained sufficient conditions to distinguish between similar verbs and to indicate which of the verbs that broadly could appear in either construction actually do (Oehrle,

1976; Gropen et al., 1989; Pinker, 1989; Grimshaw, 1990; Jackendoff, 1992).

The first step towards addressing this problem is to divide up the dative verbs into fine-grained *narrow-range* classes that, in conjunction with broad-range classes, come close to describing the patterns of grammaticality (Oehrle, 1976; Gropen et al., 1989; Pinker, 1989; Grimshaw, 1990; Jackendoff, 1992). Taking Gropen et al.'s classes as an example, *tell* and *show* are grammatical in the double object construction as VERBS OF TYPE OF COMMUNICATED MESSAGE while *say* is prohibited as a VERB OF COMMUNICATION OF PROPOSITIONS AND PROPOSITIONAL ATTITUDES.[5] Since these classes are very specific, and the distinctions between them are sometimes nuanced, it becomes challenging for the researcher to determine exactly where one class begins and another ends. A few sets of narrow-range classes have been proposed on the basis of careful descriptive work, and while they differ on a few details, they are nevertheless very similar to one another (Gropen et al., 1989; Levin, 1993).

I adopt Levin's (1993) narrow-range classification for the remainder of this study. It represents a reasonable approximation of the sorts of narrow-range semantic classes which children might learn from their inputs, and it has proven useful in describing the acquisition process of the dative alternation (Rappaport Hovav and Levin, 2008; Yang, 2016) and the causative alternation (Bowerman and Croft, 2008) among others. Levin's classes are outlined below:

(11) Double Object & *to*-Dative

- GIVE-type —feed, give, lend, etc.
- TRANSFER OF MESSAGE-type —teach, tell, show, write, etc.
- FUTURE HAVING-type —grant, offer, promise, etc.
- CARRY-type —carry, pull, etc.
- BRING/TAKE-type —bring, take.
- THROWING-type —cast, hit, throw, etc.
- SEND-type —send, ship, etc.

[5]See Gropen et al. (1989; pp. 243-244) for a full list of his narrow-range classes

- DRIVE-type[6] —chase, drive, steer, etc.

(12) *to*-Dative Only

- SAY-type —say, speak, etc.
- MANNER OF SPEAKING-type —call, cry, sing, etc.
- FULFILLING-type —entrust, leave, pledge, etc.
- PUTTING IN A SPECIFIED DIRECTION-type —fasten, lift, raise, etc.
- LATINATE-type —distribute, explain, remit, translate, etc.

(13) Double Object Only

- DO ONLY-type —ask, beget, wish, etc.
- DUB-type —anoint, dub, etc.
- APPOINT-type —allow, appoint, ordain, etc.
- BILL-type —bill, charge, tender, etc.
- DECLARE-type —declare, judge, etc.

Narrow-range classes can be defined distributionally by their syntactic patterning (Pinker, 1989; Braine and Brooks, 1995). This provides a mechanism by which children can hypothesize language-specific mappings between classes and constructions, and experimental evidence suggests that both children and adults are sensitive to these distributions (Gropen et al., 1989; Conwell and Demuth, 2007; Wonnacott, 2011; Ambridge et al., 2012). Children are sensitive to morphophonological distributional evidence as well, as Gropen et al. (1989) shows for Latinate verbs. If narrow-range classes are identified by their apparent semantic functions in the input, then there is also room for individual variation (Fellbaum, 2005).

5.3.2. Learning Generalizations

Semantic classification is useful in solving Baker's Paradox because it allows learners to pool evidence from the members of each class. However, they still need a means for projecting the constructions beyond their input onto those classes. This is where the Sufficiency Principle

[6]Levin places DRIVE verbs here tentatively. My judgment is that they do not allow double objects.

(SP; Yang, 2016, p. 144) comes in as a theory of generalization.

To review Section 2.3, the SP is a corollary to the Tolerance Principle which asks whether a rule or construction that a child has only heard apply to a few forms should be extended productively to forms yet unheard. This is exactly what is needed to describe the acquisition of the dative alternation: Learners hear the *to*-dative with some verbs but not all verbs to which it could conceivably apply, so they have to decide whether they can safely extend the *to*-dative to other verbs (i.e., their absence in the input was just chance) or they cannot (i.e., it would be ungrammatical to do so).

Say that a child has learned some N members of a given narrow-range class and has witnessed m of those have not (yet) been attested with the *to*-dative. It is then up to the child to decide whether support the *to*-dative and just have not be attested yet or if these m verbs represent exceptions to the general pattern of the class. The SP defines the point at which the $N - m$ which have attested the *to*-dative are numerous enough to motivate the generalization.

This process is repeated for all narrow-range classes as visualized in Figure 29 for a hypothetical child and a hypothetical system of four narrow-range classes. Each class has its own N, the number of members that this particular child has learned so far, and its own m, the number of members which the child has not witnessed in a *to*-dative construction so far. These N determine the width of each number line, and m is indicated as a tick on the line. The part of the line below the sufficiency threshold θ is indicated in gold and the part above it in blue. If the red tick falls into the gold region, then the *to*-dative is productive for that class, and if it falls above θ in the blue region, it is not. In this example, Classes 2 and 4 have productive *to*-datives while classes 1 and 3 do not.

It is fairly common for children with small vocabulary sizes to 'guess wrong' because of the current state of their N and m and postulate productive double objects or *to*-datives in constructions where they do not belong. Such children with 'over-productive' grammars occasionally make over-production errors. As noted by Bowerman (1983), Mazurkewich and

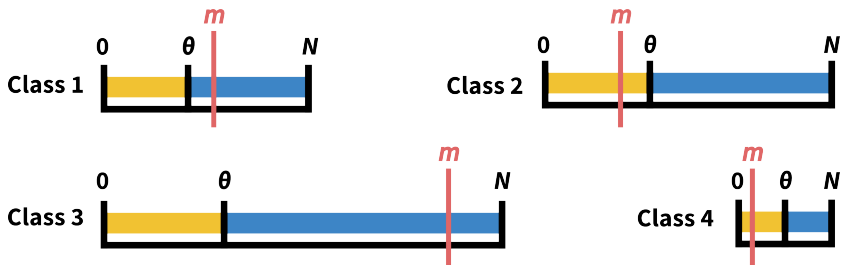


Figure 29: Number lines illustrating the sufficiency principle applied to four classes of differing size. A learner with these classes would conclude that Classes 2 and 4 have a productive pattern because $m < \theta$ but Classes 1 and 4 are non-productive because $m > \theta$.

White (1984), Gropen et al. (1989), Bowerman and Croft (2008), Ambridge et al. (2012) and (Yang, 2016, pg. 192) among others, these types of errors are attested for both the *to*-dative and the double object in Modern English. (14-15) list just a few of the many quoted examples.

(14) Overgeneralized *to*-Dative

a. '*When you gonna feed me? I asked this to you.*' (Child (4;8); Hall et al., 1984)

(15) Overgeneralized Double Object

a. '*Jay said me no.*' (Ross (2;8); MacWhinney, 2000)

b. '*I delivered you a lot of pizzas.*' (Child (3;8) MacWhinney, 2000)

c. '*Mattia demonstrated me that yesterday.*' (Damon (8;0); Bowerman reported in Gropen et al., 1989)

5.4. Actuation of the *to*-Dative

The next two sections lay out an acquisition-driven account for the development of the Modern English *to*-dative. First, constructions with semantically ambiguous objects are introduced as a possible avenue for the actuation of the *to*-dative where plausible analyses on the part of the learner would lead them to innovate a new structure overgeneralizing the meaning of *to*. Not only is type of semantic expansion typologically common (Cuyckens and Verspoor, 1998; Heine et al., 2002), but it has already been proposed in the context of the

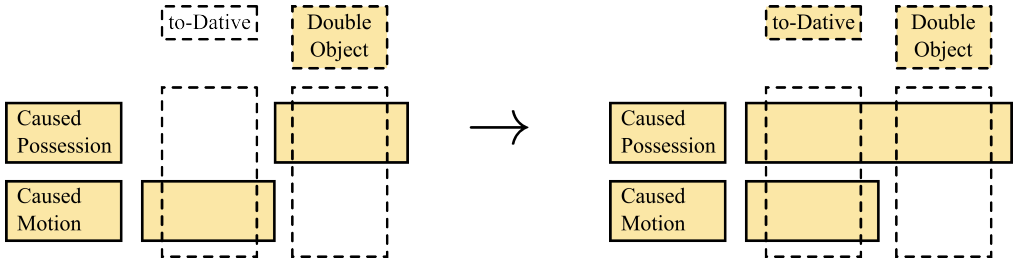
English change (De Cuypere, 2015). I also present evidence from modern child production errors that the relevant innovation is a normal, albeit normally transient, process in the acquisition of Modern English.

Next, a quantitative application of the Sufficiency Principle to Middle English data accounts for the rapid generalization of the construction across the verbal lexicon. This requires using historical corpora to approximate past learners' vocabularies in a similar way to how corpora of child-directed speech are used for the modern language. Following Chapter 3, corpora of sufficient size can be used regardless of genre or age to closely approximate the lexical knowledge of 'typical' language learners. I discuss some key observations about the time course of child language development and vocabulary acquisition and of lexical distributions in corpora which enable this kind of analysis. Then, applying the Sufficiency Principle to this typical Middle English child lexicon shows that the ambiguous *to* constructions presented enough evidence for Middle English learners to extend the novel *to*-dative to a few narrow-range semantic classes, and this in turn provided enough additional evidence for subsequent learners to extend it to an even wider range of verbs. The same application of the SP to Early Modern English data partially accounts for the *to*-dative's retreat from its Middle English 'over-extension' when the rapid influx of new borrowed vocabulary during that period shifted the makeup of the lexicon.

5.4.1. Semantic Expansion

Old English was a language without a *to*-dative, that is, the object of *tō* had to be a goal, and only CAUSED MOTION verbs could introduce their objects with *tō*. In contrast, the modern system allows both CAUSED MOTION and CAUSED POSSESSION verbs to introduce goals and recipients with *to*, so the initial change from Old English to Modern English can be seen as 'realigning' the necessary conditions for the *to*-dative from CAUSED MOTION only to either CAUSED MOTION or CAUSED POSSESSION (Figure 30).

The likely impetus for this realignment comes in the form of ambiguous Old English abstract



(a) Old English (b) Modern English

Figure 30: Realignment of broad-range classes yielding the *to*-dative.

goals and recipient-like goals introduced by *to*. These would be uttered by a conservative speaker with an intended goal meaning, but could be easily interpreted by a naive learner as containing abstract recipients or goal-like recipients instead with no effective miscommunication. Some illustrative Modern English examples of these ambiguous constructions (henceforth "ambig-to") are provided in (16). Especially given that such semantic extension of allative prepositions towards indicating recipiency is quite common cross-linguistically, this treatment can account for the presence of *to*-dative-like constructions with diverse prepositions across Germanic (Figure 26) as well as in Romance and other families as independent innovations.

(16) a. Alice said something to Bob.

b. Alice threw the ball to Bob.

Bob in (16a) must be an abstract goal, but he is very recipient-like. After all, it is hard to find a situation where hearing *Alice said something to Bob* elicits a different response from *Alice told something to Bob* even though the formal semantics of those sentences may differ. Similarly in (16b), it is easy to see how *Bob* could be interpreted with a superficial goal or recipient meaning. SAY-type and SEND-type (e.g., throw) ambig-to verbs are well attested for Middle English in the PPCME2 (17) and Old English (see Section 5.2 as well as De Cuypere (2015: §4)).

(17) Middle English ambig-to

a. and I **seye to** anoþer 'Com'

'and I say to another, 'Come'

(CMWYCSER,366.2491)

b. When Merlyn hade herde al þat his moder hade saide, he **spake to** þe kyng in this maner:

'When Merlin had heard all that his mother had said, he spoke to the king in this manner:'

(CMBRUT3,57.1671)

c. ... tyll þay **broght to** hym þat broþer þat was at home,

'... till they brought that brother that was at home to him'

(CMMIRK,99.2671)

d. Asa **sente** mychil gold and syluer **to** the king of Syrie, to helpe him ...

'Asa sent great gold and silver to the King of Syria to help him ...'

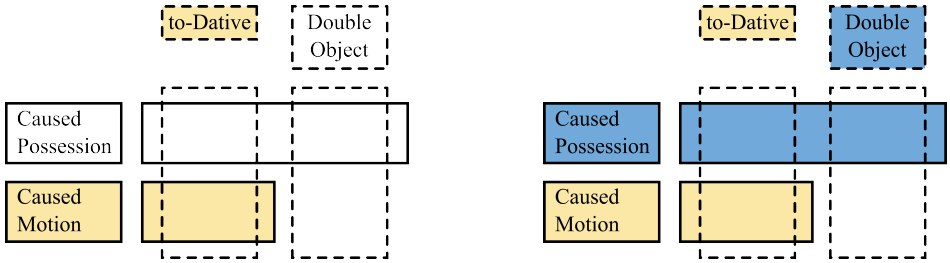
(CMPURVEY,I,22.1050)

De Cuypere (2015) identifies ambig-to constructions as a source for both the initial innovation and the eventual semantic distribution of the change. He notes that the Old English preposition introduced both concrete goals such as locations as well as abstract goals with verbs such as *say*, and in certain situations which he describes as recipient constructions, including the metonymous examples cited by Mitchell (1985) and Visser (1963) that involve a possessional transfer between human recipients. De Cuypere suggests that the lack of more general recipient constructions is due to blocking from the DO-IO double object. However, as discussed in more depth in Section 5.6, this is not a satisfactory explanation in itself because it does not account for the coexistence of these constructions in other Germanic languages.

A simpler explanation is that Old English *to* still did not have a recipiency sense so that these semantically limited examples of plausible recipient objects are actually just recipient-like goals. That is, Old English was like Icelandic and did not truly have a *to*-dative.

Therefore, semantic expansion accounts for only the initial innovation of the construction and an alternative path to is still needed to account for its eventual lexical distribution.

Semantic expansion makes a testable prediction: if a learner has decided that SAY-type verbs, for example, can express CP, that learner might be able to produce SAY-type double objects. This is attested in CHILDES (MacWhinney, 2000), a collection of child-directed speech corpora, as 'Jay said me no' (15) and 'Don't say me that or you'll make me cry' among examples with many other verbs collected by Bowerman (1983), Mazurkewich and White (1984) and Gropen et al. (1989). These modern learners were entertaining a naive grammar with SAY-type double objects at the time that this sentence was uttered, as visualized in Figure 31.



(a) *Say* (Adult) (b) *Say* (CHILDES)

Figure 31: Modern adult and attested naive alignments for SAY-type verbs.

Understanding the impact that Old/Early Middle English GIVE-type + *to* constructions had on the realignment requires a decision on Rappaport Hovav and Levin (2008)'s claim introduced in Section 5.1 that GIVE-type verbs never express CAUSED MOTION. Presumably, a modern child must learn that even *give* itself is a GIVE-type verb since the mapping between the phonological form and meaning is arbitrary. In other words, the child must learn that *give* only belongs to the CP class despite what Rappaport Hovav and Levin (2008) describe as an 'illusory' CM interpretations.

A search of the CHILDES corpus indicates that it would be surprising for a child not to entertain a CM interpretation of *give*. Of the 393 instances of the lemma *give* uttered by the caregivers of CHILDES, 27 are particle verb constructions (e.g., '*give up on*') or idioms (21

total of '*give a kiss*,' '*give a bath*,' and '*give a spanking*') and so are probably not indicative of the regular meaning of *give* itself. Of the remaining 366 instances, the vast majority, 349 (95.4%), involve straightforward apparent physical transfer of possession (e.g., '*You wanna give Melissa her present now?*', '*Just give me the cow.*'), so the vast majority involve plausible concrete goals. It seems likely that children would entertain a CM meaning for *give* at some point in their development given that it introduces plausible goals the vast majority of the time, and the same argument can be made for many other GIVE-type verbs.

It is of course possible that modern English learners grow out of this hypothesis and eventually acquire Rappaport Hovav and Levin's system exactly, but there is still the question of *to*-dative-like constructions with recipient-like goals in Old English and Icelandic. This sensitivity to goalhood is exactly the behavior that we would expect if GIVE-type verbs sometimes expressed CAUSED MOTION in a language where *to* required a CAUSED MOTION verb, so that is the analysis taken for the remainder of the paper.

There is one other observation worth mentioning: a child who analyzed recipient-like goals as goal-like recipients would have no way of correcting themselves, since there is no disambiguating evidence. The practical meanings of the two interpretations are the same despite their slightly different formal semantics, so every time such a child child heard a recipient-like goal with *to*, that child could interpret it as a goal-like recipient with no issue. The mere innovation described in this section is actually 'asymptomatic' in that the extensions of the conservative and innovative grammar are the same, so for all we know, individuals still differ in how they parse ambig-to constructions today. Figure 32 diagrams Old or Middle asymptomatic English innovation graphically. Under the naive learner analysis, a verb expressing SAY-type CAUSED MOTION with *to* is assumed to be a CAUSED POSSESSION verb with *to*. This assumption requires the learner to entertain a grammar with CP *to*-datives and SAY-type verbs that can express CP.

The innovation becomes symptomatic as soon as the *to*-dative expands within the verbal system, is uttered in an unambiguous construction, and becomes input to other speakers

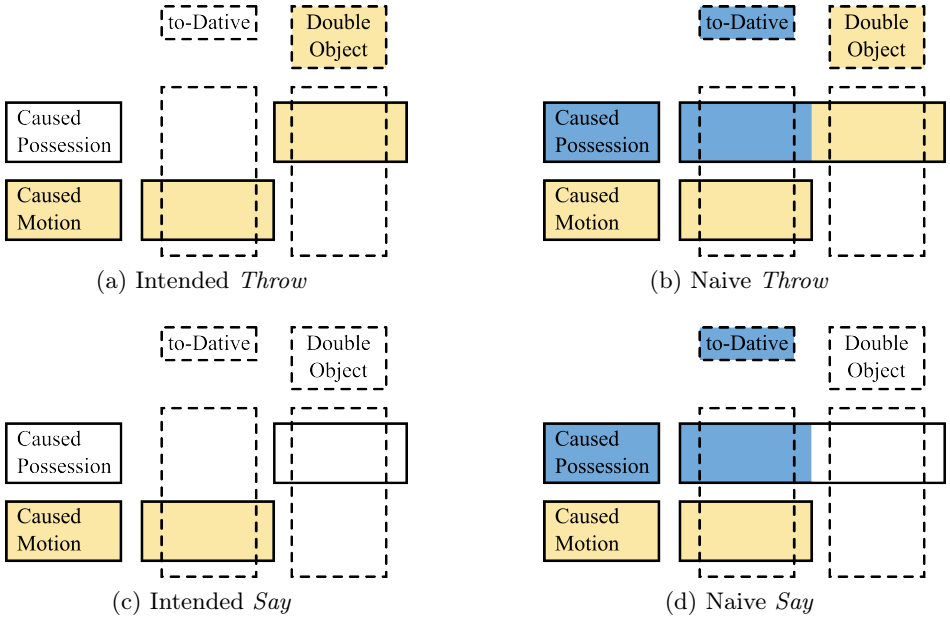


Figure 32: Intended meaning and asymptomatic naive interpretation of pre-modern ambig-to constructions. Naive *to*-dative THROW is always asymptomatic, while naive CP SAY would only remain asymptomatic as long as the learner did not produce a SAY-type double object.

in the population. The following section addresses this lexical extension from individual ambig-to verbs to narrow-range semantic classes and the implications for community-level change.

5.5. Advance and Retreat of the *to*-Dative

Once it was innovated, the lexical distribution of the *to*-dative expanded rapidly. By hypothesis, this extension of the construction to a wide variety of verbs depended on changes to its evidential basis in the language. A *to*-dative is easy to innovate via semantic expansion, so its initial actuation is not the limiting factor in its distribution, and it may have been innovated asymptomatically many times and only gained a foothold once or twice in a given language. Invoking uniformitarianism again, I apply the Sufficiency Principle (SP) as a concrete mechanism for investigating this problem. Following the SP, a learner who had innovated the *to*-dative with ambig-to could extend the construction to unambiguous verbs in a given narrow-range class only if enough ambig-to members above the sufficiency threshold had been witnessed. If the number of ambig-to verbs lay below the threshold,

then speakers who innovated it would have restricted it just to ambig-to verbs and would have remained externally indistinguishable from everyone else despite their unique internal grammars.

The following section investigates whether Middle English learners who had just actuated the *to*-dative were in a position to extend it to additional verbs. This requires approximating the lexicon of young Middle English speakers from the Penn Parsed Corpus of Middle English (PPCME2) in a similar way to how Modern English learners' lexicons are estimated from modern corpora of child-directed speech. Combined with Levin's semantic classification, this quantifies the conditions under which the extension to the attested Middle English distribution was possible. Next, Early Modern English learners' lexicons are approximated the same way with vocabulary from the Penn Parsed Corpus of Early Modern English (PPCEME) to show how learners again employing the SP would have driven the *to*-dative back down toward its modern distribution.

In the modern analogue to this problem, children's linguistic knowledge is estimated by analyzing the child-directed speech (CDS) in corpora made available as part of CHILDES (MacWhinney, 2000) among others. Of course, CDS corpora are totally unavailable for Old and Middle English, but PPCME2 and PPCEME make fine substitutes. As discussed in Chapter 3, historical corpora can be used in lieu of CDS for the purpose of estimating a child's knowledge of double object verbs when infrequent items are discarded. Moreover, the PPCME2 corpus is similar in size to the CDS corpora used in acquisition work, containing about 1.2 million words total of Middle English text (Kroch and Taylor, 2000), which is smaller but to the same order of magnitude as the CDS in CHILDES's combined English corpora. I extract frequent verbs from the PPCME2 and sort them into Levin's semantic classes in the same way one would for a modern CDS lexicon.

5.5.1. From Actuation to a Wide Distribution

I begin by estimating the number of ambig-to verbs available to Middle English-learning children in order to determine what evidence the *to*-dative had in its favor immediately after actuation. To find the relevant verbs in Middle English, I extracted all those occurring in either the double object construction or with a *to*-PP complement in the Penn Parsed Corpus of Middle English 2 (PPCME2; Kroch and Taylor, 2000). The inflected verbs were grouped by lemma (orthography normalized), and lemmas occurring two or fewer times were removed. These cannot give good statistical information about the attestation of two individual constructions, and they are more likely than other verbs to be genre-specific vocabulary that would have occurred rarely if at all in child-directed speech. The cutoff employed here is equivalent to removing all lemmas which occur less than 400,000 times in the corpus. This is stricter than Nagy and Anderson (1984) and Yang's (2016, p. 71) one-per-million rule of thumb for child-directed speech, but gives a comparable lexicon size and effectively removes genre-specific vocabulary.

The resulting items were marked according to which may have an ambig-to meaning. To the extent possible given the small corpus size, these assignments were made based on attested usage in the text, but attested usage is an underestimation of the actual prevalence of ambig-to in the language because of the size and domain of the corpus, so these numbers were supplemented with judgments from the author. Verbs were then grouped into the narrow-range classes presented by Levin, and those not listed in Levin were assigned to her classes by best match. Verbs which have been lost in Modern English (e.g., *scriben* 'to write') were assigned according to their translations but given unique lemma codes (e.g., *scriben* 'write2' to differentiate it from *write*).[7] The conclusions drawn from this work are robust in that they do not depend on the exact semantic classification or exact numbers tabulated at this step, which will be made clear by considering plausible alternative classifications and counts later in this section.

[7]The full list of verbs and classifications is enumerated in the Appendix.

What results is a set of 36 ambig-to verbs out of 75 verbs occurring more than twice in double object or *to*-PP constructions in our approximation of a Middle English child lexicon. Furthermore, six out of the ten most frequent verbs in the Early Middle English (m1) documents belong in the ambig-to set (*give*, *say*, *send*, *bring*, *offer*, and *nimen* 'take'). 48% of the frequent lemmas under consideration and 60% of the 10 most frequent lemmas are verbs which support ambig-to constructions, so it would have been impossible for children not to have heard many ambig-to sentences. 75 represents a typical English learner's experience with these verbs and constructions and is to the same order of magnitude as the number of verbs that Yang (2016; pp. 201, 204, 207, etc.) extracted from modern corpora in his studies of acquisition of the modern dative constructions. Only a few lemmas presented in the Appendix (e.g., *bequeath*) obtained by this method are not plausibly known by children, and the vast majority are much more mundane.

Productivity is calculated for each individual narrow-range semantic class to determine to which verbs an innovative learner might extend the nascent *to*-dative. For each class, N is the number of attested verbs in the class and $N - m$ is the number of lemmas which may plausibly appear in ambig-to contexts. Table 7 provides these numbers as well as the sufficiency threshold $N/\ln N$ for each class and indicates whether the *to*-dative should be extended productively for that class. Ambig-to verbs alone provide enough evidence for a typical Middle English child to extend them to all learned members of 10 of the 16 classes, or from 36 verbs to 47.

Class	**N**	**N/ln N**	**m**	**Extend?**
TRANSFER OF MESSAGE	10	4.34	8	no
GIVE	5	3.11	1	**yes**
FUTURE HAVING	14	5.30	4	**yes**
CARRY	0	-	-	-
BRING/TAKE	4	2.89	0	**yes**
THROWING	1	(1)	0	**yes**
SEND	1	(1)	0	**yes**
DRIVE	1	(1)	0	**yes**
SAY	2	(2)	0	**yes**
MANNER OF SPEAKING	2	(2)	2	no
FULFILLING	3	2.73	1	**yes**
PUTTING IN SPEC. DIR.	7	3.60	3	**yes**
LATINATE	9	4.10	4	**yes**
DO ONLY	6	3.35	6	no
DUB	4	2.89	4	no
APPOINT	3	2.73	3	no
BILL	0	-	-	-
DECLARE	3	2.73	3	no

Table 7: The extension of the *to*-dative from ambig-to to Levin's narrow classes.

This is of course not yet the attested Middle English or Modern English distribution. However, such a speaker would be a symptomatic innovator, and when that speaker produced novel *to*-datives, any peers, younger companions, or descendants could have received these new unambiguous *to*-datives as inputs in addition to ambig-to, providing an avenue for further extension over time. Children are actually sensitive to sociolinguistic variation from a young age and begin to orient themselves towards their peers by age three or four (Roberts and Labov, 1995), so the presence of slightly more mature innovative peers provides an avenue for innovation. Additionally, even if some children did not orient themselves towards their peers' sociolinguistic variants, they would have no reliable means of discerning innovative productions from conservative ones because of Baker's Paradox.

I consider what patterns a subsequent learner receiving these new inputs from older peers should have acquired. First, since the non-ambig-to *to*-datives stand as unambiguous evidence for the construction, they would not have had to reinnovate it for themselves. Second,

they would have had more evidence for the *to*-dative, around 47 verbs on average rather than 36, which could allow for higher-level generalizations. I test two possibilities for such divisions: a grouping of five classes defined below and the grouping of three higher-level classes as defined by Levin.

(18)
- **Class 1** (caused possession, some caused motion) —TRANSFER OF MESSAGE, GIVE, FUTURE HAVING
- **Class 2** (caused motion and caused possession) —CARRY, BRING/TAKE, THROWING, SEND
- **Class 3** (frequent caused motion) —DRIVE, SAY, MANNER OF SPEAKING, FULFILLING, PUTTING IN SPEC. DIR.
- **Class 4** (morphophonologically defined) —LATINATE
- **Class 5** (no caused motion) —DO ONLY, DUB, APPOINT, BILL, DECLARE

Tables 8 and 9 shows the SP's application to these higher-level classes with $M = 47$. The exact same outcomes are achieved for the five-class division as for a three-class division with 1 and 2 grouped together and 3 and 4 grouped together and DRIVE-type verbs are moved from Class 3 into 1-2 as Levin did, showing that this outcome is not an artifact of the choice of groupings.

Class	**N**	**N/ln N**	**m**	**Extend?**
CLASS 1	29	8.61	8	**yes**
CLASS 2	6	3.35	0	**yes**
CLASS 3	15	5.54	2	**yes**
CLASS 4	9	4.10	0	**yes**
CLASS 5	16	5.77	16	no

Table 8: The extension of the *to*-dative from narrow-range classes to the 5-class system.

Class	**N**	**N**/ln **N**	**m**	**Extend?**
CLASS 1-2+DRIVE	35	9.84	8	**yes**
CLASS 3-4-DRIVE	24	7.55	2	**yes**
CLASS 5	16	5.77	16	no

Table 9: Alternative extension of the *to*-dative from narrow-range classes to Levin's 3-class grouping.

This second age cohort of children could extend the *to*-dative from 47 to 59 of 75 verbs, resulting in a system very similar to the modern one. It should be noted that this conclusion is insensitive to the exact grouping of classes, and only somewhat sensitive to the annotation of ambig-to verbs. Many of the classes in both the first and second iteration fall well under their sufficiency thresholds, so even if the list of ambig-to verbs employed here turned out to be an overestimate, the calculations would have the same outcome. Another way to put it is that even children who by statistical chance happened to hear many fewer *to*-datives than the 47 predicted for the 'typical' child would still acquire the wider system. In particular, the three-class and five-class groupings could tolerate up to six and ten fewer ambig-to verbs respectively out of the 36 originally calculated. So for example, if the SAY or THROWING and SEND classes contained zero ambig-to lemmas and the GIVE and FUTURE HAVING classes contained fewer ambig-to lemmas, the calculation would still work out as is.

The younger friends and siblings of these children would receive around 59 *to*-dative verbs as input on average. If these children postulated a broad generalization including all dative verbs, it would succeed, showing that there was sufficient evidence at this point to presume a *to*-dative even for verbs like *ask*, *forgive* and *forbid* as they are attested in Middle English. Modern learner errors (e.g., '*I asked this to you*') indicate that this kind of over-generalization is still possible, at least for a while during development, despite the semantics of these verbs and the lack of evidential basis for such a construction in the modern language.

N	**N**/ln **N**	**m**	**Extend?**
75	17.37	16	**yes** ($16 < 17.37$)

Table 10: Extension of *to*-dative from specific broad classes to all dative verbs.

In predicting its 'over-extension,' this learning model accurately predicts the lexical distribution of the *to*-dative in Middle English. It also makes predictions regarding its geographical expansion: First, the *to*-dative could reach its attested pattern among children wherever it was innovated (there could very well have been multiple sites of innovation) in only a few years, meaning it would attain grammaticality in its broad distribution for at least some local adults in a single generation. Second, since the intermediate stages of the change were transient and the final stage could reached in a single generation, we would expect the final 'over-extended' stage and not the intermediate stages to pass from community to community through contact or migration. Third, since literacy was uncommon in that era, we might expect the *to*-dative in its broad Middle English distribution, to already be spread widely around England by the time it was first attested. These predictions bear out in that the *to*-dative is already well-attested in the earliest Middle English texts, and it already appears with a wide variety of verbs (Elter, 2018). The increasing token frequency of the construction (McFadden, 2002) in texts and the ultimate geographic extent of the construction are questions of usage and sociolinguistics, not grammaticality.

5.5.2. Retreat to the Modern Distribution

The *to*-dative achieved a wide distribution in the Middle English period because the particular sizes of the narrow-range classes of that era allowed children to extend it to all recipient ditransitive verbs. In that sense, its success was contingent on the composition of the lexicon of the time. If the lexicon had been different enough, the distribution of the *to*-dative may have ended up differently too just because the classes were differently sized. This grants the lexicon significant power over the grammar, which is interesting because it provides an indirect avenue for lexical borrowings to trigger syntactic change in the absence of direct

grammar borrowing.

By the late 16th century, the rate of Latin borrowings into English had already eclipsed the rate of French borrowings in the previous centuries (Durkin, 2014), and this massive influx of Latinate vocabulary at the start of the Early Modern English period was more than enough to alter the relevant sufficiency thresholds. The iterative application of the SP on the Middle English data in the previous section suggests that the system-wide *to*-dative was already on shaky ground since its presence in Class 5 was predicated only on support from all the other classes. If changes to the lexicon from borrowing or loss of vocabulary altered the N values enough, learners could have failed to generalize to those class if their respective m values grew too rapidly.

To estimate the effect of lexical change on the status of the construction, Early Modern English dative verbs from the Penn Parsed Corpus of Early Modern English (PPCEME; Kroch et al., 2004) were added to the calculation. PPCEME contains 118 verbs, 57 of which also appear in the PPCME2 Middle English set and 44 ambig-to verbs, 27 of which occurred in the Middle English set. The set difference between the Middle English PPCME2 corpus and Early Modern English PPCEME corpus is taken to approximate the change to the English lexicon over this period: verbs in PPCME2 but not in PPCEME were lost and verbs in PPCEME but not in PPCME2 were gained. Many of the Latinate verbs introduced at this step are present in Yang (2016, p. 208)'s estimated modern lexicon including *administer*, *convey*, *mention*, *return*, and *submit*, which suggests that our Early Modern English lexicon remains a reasonable approximation for these purposes.

Learners at the start of the retreat would have received evidence that the *to*-dative was available for all recipient verbs. To model this, all verbs that were retained from PPCME2 are assumed to support the *to*-dative as well as all new ambig-to verbs. This is a conservative assumption because it implies that at least some members of every class including Class 5 support the *to*-dative at the start of the calculation, stacking the odds against the retreat phenomenon being modeled here.

A child who learns by applying the Sufficiency Principle to this data would find that the *to*-dative cannot be extended to Class 4 or Class 5, meaning that a child who did not hear some member of one of these classes with the *to*-dative would decide that that was due to ungrammaticality rather than data sparsity. Motivating Baker's Paradox, even common verbs are not necessarily attested to modern learners with every possible construction, so this supposition of ungrammaticality could result in erosion of the construction's base over time.

Class	**N**	$\mathbf{N}/\ln\mathbf{N}$	**m**	**Extend?**
Class 1	27	8.19	0	**yes**
Class 2	8	3.85	0	**yes**
Class 3	29	8.61	8	**yes**
Class 4	29	8.61	15	no
Class 5	25	7.77	14	no

Table 11: Extension of the *to*-dative from narrow classes to broad classes in Early Modern English.

The further extension within Classes 4 and 5 is not possible either because there is not enough evidence for the *to*-dative in the Early Modern English lexicon. With a lexicon of this makeup, the wide generalization that was possible in Middle English fails as well, so the other classes cannot force productivity onto classes 4 and 5 as they had in Middle English.

# to-Dative Lemmas	**# Lemmas**	θ	**Extend?**
87	118	24.73	**no** $(31 \gg 24.73)$

Table 12: Extension from broad classes to all dative verbs in Early Modern English.

This neatly brings Class 5 up to its modern status, but it presents a problem for Latinate Class 4 since that class does support the *to*-dative today. A potential diachronic solution to this question lies in the history of Latinate borrowings into English: perhaps the assumption that no new vocabulary automatically supported the *to*-dative was overly conservative since there is evidence that at least some borrowings via French came along with French (i.e., *to*-dative) syntax. This is particularly clear in some documents translated directly from

French, and has been proposed to account for the anomalous Middle English *to*-dative with verbs like *please*, which was borrowed from French (Trips and Stein, 2008). If at least some Latinate verbs were borrowed along with the *to*-dative, this would depress the m count for that class, making it more likely to pass under the productivity threshold.

Alternatively, sensitivity to the morphophonological conditions that define the LATINATE class are known to be acquired slightly later than the other generalizations discussed in this paper (Tyler and Nagy, 1989; Jarmulowicz, 2002). This is reflected in Chapter 3's comparison between CDS and non-CDS vocabulary estimates: though trimming infrequent items brings CDS-derived and non-CDS-derived lexicon sizes much closer together, non-CDS lexicons still somewhat overestimate the rate of Latinate verbs in the CDS lexicon. If nine or more of the new Early Middle English Class 4 verbs were considered high register and not acquired by the time that a child performs this calculation, the *to*-dative would have been rendered productive in Class 4.

5.6. Existing Diachronic Accounts for the *to*-Dative

Several other treatments for the rise of the *to*-dative have been proposed, none of which attempt to account for the Middle English "over-extension." I group these into four types according to their focus and general claims: 'strong' morphological erosion accounts, 'weak' morphological erosion accounts, borrowing accounts, and semantic expansion accounts, and review their predictions. The last type, semantic expansion, is incorporated into the innovation state of the learners' perspective account, but it, along with the other account types, makes some incorrect predictions.

5.6.1. 'Strong' Morphological Erosion

The decline in Middle English DO-IO is closely matched in time with the rise of the *to*-dative, and these developments occurred in roughly the same era that the overt nominal case marking was lost. These temporal correlations have inspired accounts that tie the change in constructions together as a direct consequence of the erosion of overt case marking (Visser,

1963; Allen, 1995; McFadden, 2002; Polo, 2002). The general idea is that the loss of overt morphology makes some linguistic input ambiguous and can render certain syntactic structures unlearnable, motivating young learners to innovate new structures that resolve the ambiguity. One option is to adopt a grammar with more fixed word orders, which may account for the actuation of a change (to repair unlearnable constructions), the advantage behind a change (to reduce ambiguity), and the often-observed balance between morphological complexity and syntactic rigidity. This approach has a long history in historical syntax, for example, Lightfoot (1999) takes this approach in explaining the development of the English split genitive construction, and Rohrbacher (1994) and Bobaljik and Thráinsson (1998) connect the presence of agreement morphology with V-to-T movement in North Germanic, though Heycock and Wallenberg (2013) draws the strong connection between rich agreement and verb-raising into question. Less directly, morphological erosion has been shown to have facilitated rather than directly triggered syntactic changes in the history of Greek (Michelioudakis, 2015). See Biberauer and Walkden (2015) for more examples and a short review.

Strong morphological accounts appeal to the synchronic consequences of the loss of an overt dative-accusative distinction. The immediate necessity of ambiguity resolution forces a change to the syntax resulting in the *to*-dative as an analytic expression of recipienthood. Empirically we would expect the two changes, the loss of overt morphology and the advent of a new syntactic construction, to proceed exactly together. However, this does not bear out diachronically.

An inspection of the path of change in Middle English points to regional variation in the use of the early *to*-dative. There is a long gap between the loss of overt case marking on nouns and the advent of the *to*-dative in at least some regions of England (Allen et al., 2001; Polo, 2002). *Prima facie*, a learner in those regions would not have heard the case marking to clue them into the DO-IO double object and so should have repaired with a new *to*-dative under a morphological account. Polo (2002) addresses this with a more fine-grained

approach, noting that the distinction was lost on nouns before pronouns and arguing that as long as the dative-accusative distinction was at least retained on some pronouns, there was sufficient evidence for children to learn the DO-IO construction for full NP objects as well. In her analysis of the late Middle English *Second Continuation of the Petersborough Chronicles* representing the Southeast Midlands, she finds that the correlation between the loss of the dative-accusative distinction on pronouns and the loss of DO-IO to be closer than the correlation with the loss of the distinction on nouns. However, an earlier additional temporal discontinuity arises if we consider any of the occasionally attested late Old English examples (see Visser, 1963) as true *to*-datives like De Cuypere (2015) would suggest. If that is the case, then the construction arose well before the loss of case marking and so could not have been the result of it.

Additionally, there exist varieties of Modern English (yellow in the Figure 26 map) which support DO-IO: the construction is grammatical for pronominal objects for speakers across large swathes of the English Midlands (MacKenzie and Bailey, 2016), and with full NPs in Liverpool (Biggs, 2015) as in (19). The existence of Liverpool English demonstrates that there exists some underlying structure by which surface DO-IO is effable regardless of an overt dative-accusative distinction. If learners did not have to innovate the *to*-dative when case marking was lost, then a strong morphological erosion approach cannot work.

(19) a. Mary gave the book the teacher

b. Mary gave the package her nan's

5.6.2. 'Weak' Morphological Erosion

It is clear as it stands that morphological erosion on nouns is not necessary to motivate a new *to*-dative construction, and it is not sufficient to force the old DO-IO out of existence. If overt case marking is relevant for the grammaticality of the dative constructions, it has to be more of a functional pressure pushing languages without overt case marking towards a less ambiguous analytic construction and languages with overt case marking towards a

more economical synthetic one. Even approaches that rely on some other factor to account for the construction's innovation still appeal to this functional pressure to account for its expansion (McFadden, 2002; Allen, 2006; De Cuypere, 2015)

Referring back to the synchronic Germanic map (Figure 26), a functional tendency should favor languages like Icelandic (pink) which make an overt dative-accusative distinction and by hypothesis have a DO-IO double object instead of a *to*-dative, and languages like Standard English (dark blue) which lack an overt case distinction and therefore have a *to*-dative instead of DO-IO. These are indeed the most common patterns by geographic spread, but there are many exceptions as well.

One could claim that languages with both a *to*-dative and DO-IO (yellows on the map) are unproblematic because are merely in transition from pink to dark blue. High German varieties retain case marking but allow *to*-datives (thought not common in most dialects) with cognates of Standard German *an* or with *i* in some High Allemanic varieties. Low German varieties lack an overt dative-accusative distinction, yet retain both a *to*-dative, and DO-IO (Bacovcin, 2017), as does Swedish, but only with certain particle verbs (Lundquist, 2014). More problematic is Liverpool English, which has apparently gone from ideal dark blue back to yellow, an *anti*-functional change.

Several regional North Germanic varieties provide a more severe challenge to weak morphological accounts. These languages (light blue) retain their inherited overt dative-accusative distinctions and yet have a *to*-dative instead of DO-IO, which is the most anti-functional state. Faroese (West North Germanic) exemplifies this pattern (20).

(20) Faroese (Lundquist, 2014)

a. Hon gaf Mariu troyggiuna
She gave Maria.DAT sweater.DEF.ACC
'She gave Maria the sweater.' (IO-DO)

b. * Hon gaf troyggiuna Mariu
She gave sweater.DEF.ACC Maria.DAT

'She gave Maria the sweater.' (DO-IO)

c. Hon gaf troyggiuna till Mariu
She gave sweater.DEF.ACC to Maria.DAT
'She gave the sweater to Maria.' (*to*-dative)

The same pattern is attested in Elfdalian, an East North Germanic variety more closely related to Swedish than to Faroese, with a different preposition (Dahl, 2009; Garbacz, 2010). Since these dialects are not closely related and are separated by seas and mountains, they probably reached their anti-functional states independently.

(21) Elfdalian

a. Ig gav kullum dukkur
I gave girls.DAT dolls.ACC
'I gave the girls dolls.' (IO-DO (Garbacz, 2010))

b. * Ig gav dukkur kullum
I gave dolls.ACC girls.DAT
'I gave the girls dolls.' (DO-IO (Garbacz, 2010))

c. Dier åvå selt gardn að
they.NOM have.PRS.3P sell.PAP.N farm.DEF.ACC.SG to
buälaę
company.DAT.DEF.textstsg
'They have sold the farm to the company.' (*to*-dative (Dahl, 2009))

If morphological erosion really does exert a significant functional pressure on these syntactic constructions then the copious violations among English's close relatives require explanation. On the other hand, if the functional pressure is so weak that it readily admits exceptions, then some other explanation is required for the pattern. The mere presence of a functional correlation in variation and change is not enough to posit a functional cause (Weinreich et al., 1968; Labov, 1987), and in any case, the weak morphological erosion hypothesis does not actually explain where the *to*-dative originates in the first place or what caused the 'over-extension' or retreat. Since all of its proponents provide additional actuation accounts anyway, we are better off looking elsewhere to explain the rise of the *to*-dative.

5.6.3. Borrowing

The borrowing account of the English *to*-dative implicates a French calque as the source of the construction (Visser, 1963; Trips and Stein, 2008; Elter, 2018). Middle English is known to have borrowed heavily from the French of the time, and French already used a *to*-dative-like construction for recipients (22), so by hypothesis, Middle English borrowed the *to*-dative with everything else.

(22) Modern French (Demonte, 1995)

Je (lui) donne le livre à Marie.
I CL.3SG give the book to Marie.
'I gave Marie the book.'

Much of the extant Middle English corpus is actually translated directly from French, so it seems likely that the *to*-dative's high rate of attestation during the period is at least stylistically influenced by French. In this way, the borrowing hypothesis is able to account for some of the attested over-extensions, for example, (23) from the Ayenbite of Inwyt[8]

(23) Middle English possible calque

Huet may þe zone betere **acsy to his uader**...
'What may the son better ask his father for...'

However, there are a number of holes in the borrowing hypothesis. First, this would not account for the wide distribution of the *to*-dative in North Germanic varieties which had much less contact with French (light and dark blue on the map) or for French itself. Second, there would need to be a mechanism accounting for why this particular construction eventually made its way into the native grammar when other French constructions did not. And third, the chronology and type distribution of *to*-datives in Middle English texts is not what one would expect if it were borrowed. If the *to*-dative was indeed innovated in

[8]cf., *Que le fils peut-il **demander à son père ?***

Late Old English, then the basic assumption does not hold. And even if it was innovated in Early Middle English, it was already attested in a wide semantic range, with native (non-Latinate) verbs, and in texts which were not translated from French (Elter, 2018). Many of the over-extensions are attested in native texts. It seems certain that borrowing influenced the attested distribution and usage rate of the *to*-dative in Middle English, but something else is needed to account for its innovation and its grammaticality among native verbs outside of translations.

5.6.4. Semantic Expansion

Finally, semantic expansion proposes that *to* used for abstract goals and in metonymous constructions came over time to introduce recipients as well with some verbs (De Cuypere, 2015). The current study strengthens the position of the semantic expansion account as an explanation for the initial innovation of the *to*-dative. However, semantic expansion is more problematic as an account for the construction's distribution since as Elter (2018) uncovers, the *to*-dative already has a wide lexical and semantic distribution by the time of the earliest Middle English documents. This 'saltational' appearance is counter to the gradual expansion which De Cuypere predicts. He treats the lack of attested *to* with caused possession verbs as evidence for a gradual expansion, however this is the condition still attested in Icelandic which fails to meet the semantic definition of a *to*-dative in the first place. The present study accounts for this by analyzing Old English as a language that lacks a *to*-dative but has concrete and abstract allative-to and Middle English as a language that does have a true *to*-dative. The rapid appearance of unambiguous *to*-datives in Old English is accounted by the learners' perspective on the acquisition of generalizations.

5.7. Discussion

In this chapter, I demonstrate how the process of language acquisition repeated in a community over time accounts for the actuation and advance of the Modern English dative constructions. The actuation of the *to*-dative is analyzed in terms of Middle English learn-

ers' naive interpretations of ambiguous allative-*to* constructions that effected a realignment of the language's broad-range classes and constructions. An iterative succession of child language learners operationalized by the Sufficiency Principle then accounts for the attested lexical distribution of the *to*-dative in Middle English. There was enough evidence available to learners of the time for them to extend the distribution of the *to*-dative to more verbs until it reached its full distribution in just a few age cohorts. Then in the face of massive lexical borrowing into Early Modern English, productivity was no longer a tenable hypothesis for some classes of verbs, so the *to*-dative retreated for some semantic classes. The historical course of the *to*-dative in English mirrors that undertaken by individual children as they sort out the constructions for themselves. Many go through a phase of over-generalization which is make possible by their small lexicons, but as they learn more, those generalizations are rendered untenable, and they undergo retreat.

5.7.1. A "Learners' Perspective" Approach to Language Change

It is quite common to invoke the process of native language acquisition when developing explanatory models for change, but treatments vary greatly in the extent to which they focus on the specifics of the acquisition process. The approach laid out here distinguishes itself by placing the child front and center. Everything is conceived of in terms of what is understood concretely about the synchronic behavior of child learners in analogous settings, and the consequences of that behavior are worked out to determine what kind of change it should drive in the long run. In that sense, this is historical linguistics from the perspective of the native language learner.

The learners' perspective on change sheds new light on the utility of historical corpora as well. Corpora are taken as tools for estimating the lexicons which children have at the time that they are acquiring some pattern of grammar rather than evidence in their own right for the grammaticality of a particular pattern. This turns the conventional wisdom of corpus research on its head: only the best-attested properties of the corpora are relevant for estimating child experience, and trimming the less attested items from the corpora removes

most of their individual properties. Lexicons from different genres and slightly different eras can be collated in a way that would be problematic in most other circumstances.

Although the study in Chapter 3 and Kodner (2019) is first to quantitatively validate the use of historical corpora in this way, learners' perspective approaches to change, particularly those leveraging the Tolerance Principle have begun to gain currency in recent years in studies of recent and ongoing change, for example Dresher and Lahiri's 2015 investigation of Romance stress patterns in English, Yang's (2016; ch. 5) findings on the Icelandic 'Dative Sickness' which shows how an ongoing structural change may be driven by superficial lexical statistics, Sneller et al. (2019) which considers migration patterns and the Philadelphia short-*a* system, and Chapter 4, which investigates the effect of mixed dialect input and Canadian raising.

5.7.2. Hypothesis Testing and Building Simpler Theories

One major benefit of the learners' perspective is that it presents a way for us to delimit explanatory roles in accounting for linguistic facts. Contrast this chapter's account of the *to*-dative with (Allen, 1995), (McFadden, 2002), or (Bacovcin, 2017) in which specific theoretical assumptions carry all the explanatory burden. This acquisition-driven account accomplishes the same and more without making additional assumptions about how the representation is implemented: there is underlying syntactic structure of some kind behind the constructions that can be implemented under several Minimalist proposals or LFGs, and there are unique semantic considerations for each construction. In an account that rests on theory, acquisition, and diachrony, explanatory burden can be offloaded from one to the others. In this thesis, the direction of offloading is from theory to the others, but in principle it can move in any direction as long as it is well motivated.

This concept will be taken farther in the next two chapters. First, Chapter 6 on Proto-Germanic verbs takes the learners' perspective in order to explain a long-standing mystery in the Proto-Germanic inflectional paradigm, the lengthened *$\bar{e}$-grade in strong verb past

stems. Several competing analogical accounts for the *$\bar{e}$-grade have been proposed over the last century and some, but they have proven challenging to distinguish. By committing to a quantitative mechanism for acquisition, I am able to find support for one hypothesis, the "'Eat Analogy," over its strongest competitor, something that has thus far eluded traditional historical linguistics. The chapter on Proto-Germanic also further develops the role of speech cohorts and sub-generational transmission in actuation, culminating in an update to the Z-model.

Next, Chapter 7 on Latin morphology also makes use of historical corpora to estimate child linguistic knowledge. In taking the learners' perspective on the Latin verbal paradigm, it not only provides a novel account of key diachronic facts, but has implications for leading synchronic analyses of Latin past participles. One strong point of acquisition research is that it provides a principled account for what is rule governed or listed morphological systems. This is parallel to a standard question in theoretical morphology. When should we attempt to predict productive patterns, and when should we resort to listing or irregulars? Applying a metric for listing to a paradigm delimits what needs to be explained by the theory and what can be offloaded to acquisition and diachrony. This sheds light on the different predictions (or lack thereof) of competing accounts and ultimately leads to simpler theories.

CHAPTER 6 : The Proto-Germanic Lengthened *$\bar{e}$-Grade

In this chapter, I address a long-standing problem in Proto-Germanic morphology, the presence of the lengthened *$\bar{e}$-grade past stem in two classes of strong verbs. Proto-Germanic (PGmc) is the reconstructed common ancestor of the Germanic branch of the Indo-European family, and like its descendants, it has a number of *strong verbs* whose tense is indicated primarily by stem vowel mutations rather than by affixation. Most of these alternations go back to Proto-Indo-European by regular sound change, with the lengthened *$\bar{e}$-grade of the Class IV and V past as the notable exception. Several hypotheses for the *$\bar{e}$-grade's origin have been proposed over the last century and a half, most of which relying on some notion of analogical change. However, it has proven difficult to effectively contrast competing hypotheses without a concrete mechanism for analogical change. In working towards a solution to this problem, I develop the model of analogical leveling and extension grounded in input sparsity and productivity learning which will be further applied in Chapter 7. I contrast the viability of two leading hypotheses with the productivity-grounded model of analogical change: the 'Eat' Analogy, which proposes that the *$\bar{e}$-grade spread from just one verb, **etanǫ́* 'eat' to all Class IV and Class V strong verbs, and the Class VI Analogy, which argues that the length of the Class VI long *$\bar{o}$-grade past analogized to Class V and spread from there.

Within historical linguistics, *analogy* is one of those technical terms that means something a little different to everyone who uses it (Hock, 2003). At a high level at least, it refers to a type of language change where a pattern from part of a grammar is ported over to another part of the grammar. In its broadest sense, analogy is one of the three modes of change recognized the Neogrammarians along with regular sound change and borrowing (Campbell, 2008). Unlike overwhelmingly regular sound change, analogical change is erratic and irregular, though while it has proven difficult to explain, it is often easy to identify. Most work on the cause of analogical change, therefore, has not focused on the cause *per se* so much as cataloguing tendencies observed in analogical change (Kuryłowicz, 1945;

Hock, 1991). While understanding tendencies is useful, a tendency cannot explain why any given analogical change happened or did not happen. A cause is required for that. It seems clear that productivity is closely connected to analogy – analogical *leveling* of productive patterns at the expense of unproductive ones is far more common than the analogical *extension* of unproductive or narrowly productive patterns – but thus far, it has been unclear how to implement analogy with productivity (Hock, 2003, p. 446). To the extent that productivity is associated with synchronic regularity, analogical change may be associated with regularization as productive patterns spread. Thus the character of analogical change may be best summed up as it is put in Sturtevant's Paradox: though analogical change is irregular, it creates irregularity, compared with regular sound change, which creates irregularity (Sturtevant, 1947, p. 109).

Moreover, since the definition of analogy is so broad, there is no reason to think that a single mechanism can account for everything under that title. The Neogrammarians used it as a catch-all for patterns that could not be explained otherwise, ranging from four-part analogy to "contamination," "blending," "recomposition," and folk-etymology, among others (Hock, 2003). The analogical mechanism expounded in this chapter is most closely related to what is called *four-part analogy* (e.g., *cat* : *cat-s* :: *cow* : *cow-s*, replacing earlier *kine*). Four-part analogy, though often posed diachronically, is essentially parallel to the problem of paradigm inference in acquisition: If a child has been given *cat*, *cats*, and *cow* as input and has never encountered the plural of *cow* but wants to produce it, its form has to be inferred. There is no way around that, and the only chance the child has at producing an intelligible form is to derive it from a productive pattern hypothesized on the basis of known forms. For example, we would expect a child in such a situation to hypothesize a "regular" form *cows*, at least until *kine* is attested. If *cows* gets a foothold in the population, that instance of regularization becomes analogical change.

Because paradigm saturations and inverse paradigm saturations tend to be very low in languages with even moderately large paradigms (Section 3.3), most morphological forms in

a language will never be attested to any given learner. This makes inference on the basis of productive patterns the norm rather than the exception in the acquisition of morphological paradigms. This chapter focuses on quantifying the conditions under which over-generalizations should have occurred in Proto-Germanic in order to test competing hypotheses for the innovation of lengthened *$\bar{e}$-grade. By committing to a mechanism for productivity learning, I am able to pit two hypotheses against each other and decide which one was more likely in a way that has not been feasible previously. Following that, I address conceptually how early learner interactions could bring productivity-driven innovations into local speech communities, building on the concept of cohorts introduced in Section 5.4.

The rest of this chapter is organized as follows: Section 6.1 summarizes the relevant facts regarding Proto-Germanic strong verbs and existing work on the lengthened *$\bar{e}$-grade before enumerating what would go into a satisfactory explanation for the change. A model should be able to account for how analogy proceeded from one strong verb class to another, why the analogy occurred in the direction it did, and why only the past stem analogized. Section 6.2 discussed how the Tolerance Principle is applied to Proto-Germanic data, and lays out how productivity, operationalized by the TP, forms the basis for a model of analogical change. Following that, Section 6.3 analyzes the possibility for analogy between Class IV and V to answer why analogy occurred from V to IV rather than vice-versa and why it did not continue onto other classes. Section 6.4 analyzes the predictions of the 'Eat' Analogy hypothesis, and Section 6.5 considers the Class VI Analogy. Finally, Section 6.6 ends with a discussion of the results of this study, the relative support for the 'Eat' Analogy over the Class VI Analogy, and a framework for a learner-driven account for the innovation and actuation of analogical change.

6.1. The Proto-Germanic Strong Verb

Proto-Germanic (PGmc) is the reconstructed common ancestor of the Germanic branch of the Indo-European family. It was spoken no more than 2,500 years ago (Ringe, 2017), likely in or near modern Denmark, and its reconstructed form is based on evidence from

archaic members of the family's three branches, West Germanic, North Germanic (together constituting Northwest Germanic), and East Germanic, as well as evidence from elsewhere in Indo-European. The Eastern branch provides evidence for conservative features otherwise missing from Northwest Germanic, but it has no living members, and the largest corpus of East Germanic text, Wulfila's Bible, is only a partial text based on a literal translation of the Greek Bible. North Germanic's early attestations are from Runic Norse, with much more substantial sources coming from Old Norse. Early West Germanic languages with significant attestation included Old High German and Old English.

The Proto-Germanic lexicon contained a large number of these strong verbs which were characterized by ablaut (cf. Modern English *swim* ∼ *swam* ∼ *swum* and *ride* ∼ *rode* ∼ *ridden*) as opposed to *weak verbs* which formed their preterites and past participles with with a dental suffix instead (cf. Modern English *jump* ∼ *jumped* ∼ *jumped* and *sleep* ∼ *slept* ∼ *slept*[1]). A Proto-Germanic strong verb exhibited four unique stems: a present stem, a past indicative singular stem (past.3sg), a default past stem (past), and a past participle stem (pptc). They are traditionally classified into seven classes (I-VII) according to which stem pattern they follow (Mailhammer, 2007). From a synchronic prospective, the strong verb system was highly regular, much more so than the descendent systems, in that each strong verb Class I-VI had a transparent phonological condition associated with it, there were few exceptions to those conditions, and most strong verbs patterned consistently with exactly one class (Mailhammer,2007, ch. 2, §3; Ringe, 2017, pp. 266-279). Class VII is unique because it is primarily etymological instead, with all of its members descended from reduplicated forms. Table 13 lays out a summary of the first six classes' ablaut patterns with examples.

[1]Many weak verbs have since developed secondary vowel alternations of their own.

Class		Pres.	Past.3sg	Past	Past Part.	
I	*-*īC*-	*b**ī**tanǫ	*b**ai**t	*b**i**tun	*b**i**tanaz	‘*bite*’
		*sn**ī**þanǫ	*sn**ai**þ	*sn**i**dun	*sn**i**danaz	‘*cut*’
II	*-*euC*-	*t**eu**hanǫ	*t**au**h	*t**u**gun	*t**u**ganaz	‘*pull*’
		*kl**eu**banǫ	*kl**au**b	*kl**u**bun	*kl**u**banaz	‘*split*’
III	*-*iNC*-	*f**i**nþanǫ	*f**a**nþ	*f**u**ndun	*f**u**ndanaz	‘*find*’
	*-*eRC*-	*h**e**lpanǫ	*h**a**lp	*h**u**lpun	*h**u**lpanaz	‘*help*’
	*-*eTC*-	*fl**e**htanǫ	*fl**a**ht	*fl**u**htun	*fl**u**htanaz	‘*plait*’
IV	*-*eR*-	*b**e**ranǫ	*b**a**r	*b**ē**run	*b**u**ranaz	‘*carry*’
		*k^{w}**e**manǫ	*k^{w}**a**m	*k^{w}**ē**mun	*k**u**manaz	‘*come*’
V	*-*eT*-	*g**e**banǫ	*g**a**b	*g**ē**bun	*g**e**banaz	‘*give*’
		*s**e**h^{w}anǫ	*s**a**h^{w}	*s**ē**gun	*s**e**wanaz	‘*see*’
VI	*-*aC*-	*f**a**ranǫ	*f**ō**r	*f**ō**run	*f**a**ranaz	‘*travel*’
		*hl**a**þanǫ	*hl**ō**þ	*hl**ō**dun	*hl**a**danaz	‘*load*’

Table 13: Summary of Proto-Germanic strong verbs Class I-VI ablaut patterns (collated from Mailhammer (2007) and Ringe (2017)).

It is clear from the attested members of the family that the core Indo-European system of verbal inflection had already been heavily reworked by the time of Proto-Germanic with the inherited aspect-based system collapsed and reshaped into a tense-based system instead (Ringe, 2017). Nevertheless, the origins of most (but not all) Germanic inflections are recoverable, and even the quintessentially Germanic strong verb system is transparently derived from PIE *ablaut* (root vowel mutation) patterns expanded into multiple phonologically conditioned classes over the course of many regular sound changes. Yet there is one significant exception to strong verb system's pattern of regular development: the past stem of two of the seven strong verb classes contains a long *$\bar{e}$ which does not correspond to the expected reflexes of any PIE ablaut pattern (Ringe, 2017, p.210; Mailhammer, 2007).

The diachronic development of the strong verb ablaut patterns from Proto-Indo-European is also fairly straightforward. As summarized in Table 14 the first two stems of Classes I-VI derive from the PIE *e*- and *o*-grades respectively, and the last two stems derive from *zero*-grades except for the past of Class IV and the past and past participle of Class V.[2]

[2]Indo-European verbs exhibited three ablaut grades, *e*-grade, *o*-grade, and *zero*-grade.

The Class VI stems were colored by adjacent laryngeals before they were lost but otherwise developed similarly (Mailhammer, 2007, ch. 2, §3.2.2). The remaining stems in Classes IV and V exhibit the aforementioned long *$\bar{e}$-grade of uncertain origin (Mailhammer, 2007, ch. 2, §3.1.8) and are the subject of this chapter.

Class	Pres.	Past.3sg	Past	Past Part.
I	*e*	*o*	zero	zero
II	*e*	*o*	zero	zero
III	*e*	*o*	zero	zero
IV	*e*	*o*	*ē*	zero
V	*e*	*o*	*ē*	*e*

Table 14: Origins of Proto-Germanic Class I-V strong verb ablaut.

Proto-Germanic weak verbs were divided into four classes (Weak I - Weak IV here) according to which suffix they expressed after the root. Weak I verbs had suffixes based on *-*ja*-, Weak II on *-*ō*-, Weak III on *-*ai*-, and Weak IV on *-*nō*-. As such, weak verb stems were usually phonologically distinct from strong verb stems in Proto-Germanic, unlike in its modern descendants (cf. strong *ride* ∼ *rode* ∼ *ridden* vs. weak *glide* ∼ *glided* ∼ *glided*). Additionally, weak verbs could be distinguished from strong by the presence of dental suffixes in the past and past participle, just as they can be today.

There were only a few exceptional strong verbs scattered throughout the system. Among the relevant classes, a few Class IV verbs had anomalous *-*uC*-[3] present stems, namely **knudanǫ* 'knead,' **trudanǫ* 'step on,' and **wulanǫ* 'be agitated' (Seebold, 1970; Ringe, 2017, p. 272), and there was a verb **brekanǫ* 'break' which fits the shape of a Class V verb but had a Class IV zero-grade past participle. There were two types of exceptions to Class V as well. First, three reconstructable verbs, with Weak I-like j-presents otherwise pattern with the class: **bidjanǫ* 'ask for,' **ligjanǫ* 'lie ,' and **sitjanǫ* 'sit' along with a single n-stem **fregnanǫ* 'ask about' (Ringe, 2017, pp. 273, 275). Second, there is a single reconstructable verb **etanǫ*, *$\bar{e}$t, *$\bar{e}$tun, **etanaz* '*eat*' which has a past.3sg stem in *$\bar{e}$. This is the only vowel-initial

[3] C – consonant; N – nasal; R – sonorant; T – obstruent

member of the class and the only member for which the *$\bar{e}$ is derivable by regular sound change from PIE (*h_1e-$h_1$ód- > *eēt- > *ēt-; *h_1e-h_1d-´ > *ēt-) (Ringe, 2017, p. 210). Members of Class III with *-*eTC*- roots are typically considered anomalous too because of their historical path of development, but they behave as regular members of the class in Proto-Germanic. Since Class III's *-*iNC*- can be projected back to **-*eNC*-, treating the *-*eTC*- members as a small but regular subclass allows for a more general root shape *-*eCC*- that captures all of Class III.

*6.1.1. Accounting for the Lengthened *ē-Grade*

The lengthened *$\bar{e}$-grade past has attracted considerable attention for over a century because it is a glaring exception to the otherwise regular historical derivation of the Proto-Germanic strong verb. It is widely believed that the *$\bar{e}$ spread from Class V to Class IV rather than vice-versa because of the presence of zero-grade Class IV preterite-presents and deverbal nouns in the daughters and the smaller relative size of Class IV (Matzel, 1970; Bammesberger, 1986; Mottausch, 2000; Ringe, 2017). There is no such consensus about the origin of *$\bar{e}$ within Class V however.

Many early explanations were purely phonological. As far back as Streitberg (1896), it was proposed that $\bar{e}$ was inserted into Class V past stems as a repair to forms like *$g^h eg^h b^h$- → **gb- which would have been rendered unpronouncable after reduplication was lost (Schumacher, 2000). A related alternative suggests that it was a kind of compensatory lengthening instead Hirt (1931). However, there is no particular phonological reason why *$\bar{e}$ should have been chosen as the repair vowel as opposed to some other one, and there are no parallels elsewhere in the language for the repair or the lengthening (Mailhammer, 2007, ch. 2, §3.1.8).

Most accounts attempt to explain the lengthened *$\bar{e}$-grade as some kind of morphological change instead. Matzel (1970) and Meid (1971) propose that it a a reflex of the Brugmann (1913) alternative PIE perfect, but most accounts appeal to analogy. Some have suggested

that the *$\bar{e}$ was imported from the nominal system (Bammesberger, 1994, 1996), or old aorists (Sverdrup, 1927; Prokosch, 1939; Peters, 1980). Others have argued that *$\bar{e}$ arrived in Class V via analogy with Class VI's long *$\bar{o}$ (Kuryłowicz and Mayrhofer, 1968; Bammesberger, 1986), and still others that it is the result of analogical spread from the single verb **etanǫ* 'eat' (Kortlandt, 1992; Schumacher, 1998; Mottausch, 2000; Ringe, 2017).

These accounts are not all created equal. Some of them can be dismissed because they clearly conflict with the evidence or because they grossly stress the limits of plausibility, but the rest of them seem intuitively reasonable. This highlights the problem with the current state of affairs in historical explanation. When it comes down to it, even the best of these accounts, the ones with the most historical evidence on their side are just-so stories that appeal to our intuitions as researchers. That is not to say that they are wrong — I suspect that one of these is right — just that the inductive approach that creates these hypotheses is not especially well-suited for hypothesis testing. While we will never be able to empirically validate historical explanations with the same level of rigor than we can synchronic ones, there is still a lot of ground to be gained.

6.1.2. Explicanda

We can never know for sure what went through the minds of Iron-Age people, but we can do a lot better than this, and that is why it is important to at least entertain a mechanism for analogical change. A mechanism forces us to make explicit predictions about each scenario it is applied to, so we can go one-by-one through the space of intuitively plausible hypotheses and scrutinize them individually. An effective mechanism should make predictions that comport with the detailed tendencies of analogy that have already been well described (e.g., Kuryłowicz's Laws 1945) as well as what is understood about the child language learners who we suppose are implementing the mechanism.

I use productivity learning as implemented through the Tolerance Principle to compare and contrast two of the prevailing analogy accounts, which I call the *'Eat' Analogy* which suggests

that the lengthened *$\bar{e}$-grade developed for 'eat' by regular sound change then spread from there to the rest of Class V, and what I refer to as the *Class VI Analogy* which proposes that the long vowel entered Class V by analogy with the long vowels of Class VI. Any satisfying account for either hypothesis has a number of questions to address. In either case, one must answer why the *$\bar{e}$-grade spread from Class V to Class IV. The rest are hypothesis-specific.

(24) **Both Hypotheses**

a. Why did the *$\bar{e}$-grade spread from Class V pasts to IV?

b. Why was the direction of spread not from Class IV to Class V?

c. Why did Class V's past participle stem not spread?

d. Why did the spread not involve Class III?

(25) **'Eat' Analogy Only**

a. Under what conditions, was 'eat' a sufficient basis to trigger the analogical spread of the past *$\bar{e}$ to Class V?

b. Why did the past.3sg *$\bar{e}$ not spread from 'eat' if the past *$\bar{e}$ did?

(26) **Class VI Analogy Only**

a. Why did analogy occur from Class VI to Class V and not vice-versa?

b. Is it possible for vowel length but not vowel quality to analogize, and if so why did that happen from Class VI to Class V but not Class V to Class VI?

In the next section, I lay out how the Tolerance Principle can be applied to Proto-Germanic data to model productivity and describe the basis of the productivity-based model of analogical change. Following that, I apply the analogy model to Classes IV and V to show why analogy proceeded from V to IV rather than vice-versa.

6.2. Applying the Tolerance Principle to Proto-Germanic

The obvious challenge in applying any kind of quantitative model to Proto-Germanic is the fact that it is prehistoric. Not only are there no extant speakers of the language, there are

no written records, and all that we know of the phonology and morphology of the language has been reconstructed through the comparative method. The known lexicon, at least the relevant part of it, must first be reckoned with our approximations of child lexicons if we are to apply methods from language acquisition to it. Reconstructed lexicons are necessarily incomplete, because only words that have left enough evidence in the daughter languages can be securely postulated, and in general, it is impossible to assess which items are missing from the reconstruction. Reconstructed lexicons also lack frequency information since if text corpora existed to perform counts over, the lexicons would not need to be reconstructed. Well-reconstructed lexicons provide two useful pieces of information though: rough type counts, and rough translations. Fortunately, type counts are all we need for the Tolerance Principle, and as shown in Chapter 3, the 258 securely reconstructable Proto-Germanic strong verbs actually constitute a plausible verbal lexicon for a three year old, both in terms of size and semantic contents. This is because items that are securely reconstructable are those which are attested in each daughter branch, and those which are likely to be attested in each daughter branch tend to be high frequency everyday items. These are exactly those items that tend to appear in child-directed speech. It is also worth noting that even though statistics about Proto-Germanic paradigm saturation cannot be gathered without corpora, we have every reason to believe that the same long-tailed distributions that describe attested modern and ancient languages should apply to prehistoric Proto-Germanic as well.

6.2.1. Synchronic Application to the Strong Verbs

Since the Tolerance Principle is a language-independent cognitive model of acquisition, it should apply equally well to ancient languages as it does to modern ones, and since the set of reconstructed Proto-Germanic strong verbs falls within the space of plausible child lexicons, it should apply here. Before moving on to the diachrony of the lengthened *$\bar{e}$-grade past, the TP is applied to the relevant part of the synchronic Proto-Germanic system in order to determine whether the traditional strong verb classes correspond to cognitively productive sets.

If the PGmc strong verb classes are 'real,' then there must be some unifying generalizations that can be made over them. As in the modern Germanic languages, these generalizations happen to be morphophonological rather than semantic. For the Proto-Germanic strong verb system, each of the traditional strong verb classes has a transparent phonotactic condition associated with its root, for example, *-*eCC*- (*-*iNC*- < **-*eNC*-) for Class III, *-*eR*- for Class IV and *-*eT*- for Class V. These are generalizations over which the Tolerance Principle can be calculated. The classes are enumerated in Table 15 along with their phonotactic generalizations, size in the reconstructed lexicon, which serve as N for the TP calculation, and tolerance thresholds. Classes IV and V have four members each with exceptional present stems, and Class V has one member with an irregular past.3sg stem, but since 4 and 5 are less than the thresholds for both classes (5.77 and 8.40), they do not pose a challenge for the productivity of either class. Classes I-III are internally productive for the same reason.

Class	**Generalization**	**N**	θ_N
I	*-*ī*C-	41	11.04
II	*-*euC*-	40	10.84
III	*-*eCC*-	52	13.16
IV	*-*eR*-	16	5.77
V	*-*eT*-	28	8.40
VI	*-*aC*-	29	8.61

Table 15: Class I-VI by phonotactic generalization, size, and tolerance threshold.

It is also conceivable that a Proto-Germanic child who was trying to figure out how the system was structured might try out phonotactic generalizations other than those which define the classes. For example, one could collapse Classes IV and V together as Class IV+V with a root shape *-*eC*-. It would be a reasonable thing to hypothesize, since none of the other classes are sensitive to the type of the post-vocalic consonant, after all, and three of the stems exhibit the same ablaut. However, if Class IV+V existed with a pptc stem *-*e*- like true Class V, all of the verbs in true Class IV would have exceptional past participles in *-*u*-, and vice-versa. The combined class would have $N = 44$ and $\theta = 11.63$, which is well under the size of either Class IV or V. As a result, the combined class would fail in the long

run regardless of which past participle were employed. No other super-class generalization is viable either.

6.2.2. Analogy as Over-Productivity

When considered as a whole, the Proto-Germanic strong verbal system is stable and productive. The classes can be clearly distinguished and derived by productive generalizations over their root phonotactics, so it might seem like there is no room for analogical change. But this is not the whole story. The key here is to keep in mind that young children start forming morphological generalizations well before their vocabularies have reached adult-like sizes.

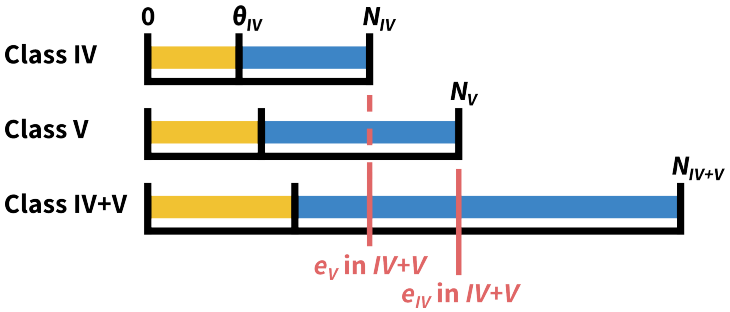
What kinds of generalizations could a child with a smaller vocabulary make? In general, the Tolerance Principle predicts that more generalizations should be possible over classes with small N than with larger N because the ratio between θ and N increases as N decreases. The higher that θ is relative to N, the greater proportion of exceptions it can tolerate. As a result, it is conceptually possible for a young child to acquire additional generalizations, including super-class generalizations that would be impossible for an adult.

Consider the Tolerance Principle number lines in Figure 33, which serve to illustrate this point on schematized versions of Classes IV, V, and super-class IV+V. In Figure 33a, Class IV and V are too large too permit the super-class generalization, as they would be if a learner knew most of the reconstructable members of each class. If Class IV+V were hypothesized with Class V's forms (henceforth, V→IV), Class IV's members would constitute too many exceptions (e_V in IV+V), and if it were hypothesized with Class IV's forms (IV→V), Class V's members would constitute even more exceptions (e_{IV} in IV+V). However, if both classes were small enough for a given learner, the super-class would actually be tenable as long as the current size of one class is smaller than the tolerance threshold of the super-class. This could happen if both classes were smaller (so θ would be proportionately larger and more permissive) as in Figure 33b, or if the learner happened to learn most of one class before

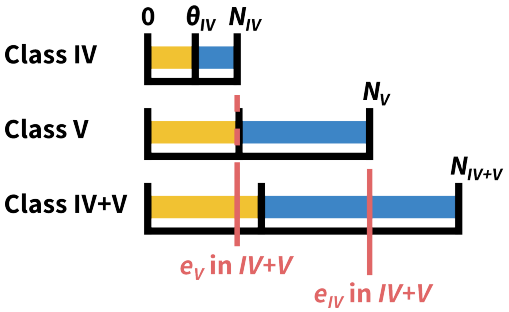
the other so that they were more unbalanced as in Figure 33c. In either of the provided examples, the learner could conclude that all the members of Classes IV and V were actually members of a single larger IV+V class with Class V's past participles and that the would-be Class IV verbs in IV+V simply had irregular past participles.

So, there are multiple opportunities for over-generalizations between classes while a learner's vocabulary is small. This provides the avenue for analogy when the forms of the subsumed class (Class IV here) have to be inferred, and they are generated as though they belonged in the other class (Class V here). Because average paradigm saturations are so low, as we can see in general and for German and Gothic in Chapter 3 Figures 9-11, it is more likely than not that some forms would have to be inferred. Every time this inference occurs, an over-generalization "error" is produced.

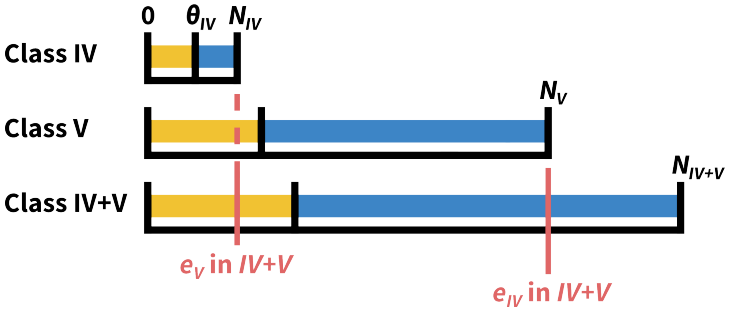
There are several possible super-class over-generalizations in the Proto-Germanic system. Class IV+V is most critical because it can be used to explain analogy of the lengthened *$\bar{e}$-grade from Class V to IV. A generalization *-[-hi]C- or *-$\breve{V}C$- could be employed to relate Class VI to IV and V if the Class VI was related to the origin of the lengthened *$\bar{e}$-grade. There also exists a potential generalization between *-*eC(C)*- Classes III, IV, and V, so a treatment for the distribution of lengthened *$\bar{e}$-grade that relies on these root generalizations should also account for its absence in Class III. While other super-class generalizations are conceivable, none of them is relevant going forward.



(a) "Large" Class IV and V



(b) "Small" Class IV and V



(c) Unbalanced Class IV and V

Figure 33: Number lines showing potential cases for over-generalization between similar verb classes. Classes IV and V share a possible root generalization *-eC- which could be applied to a super-class IV+V. a) No over-generalization possible. Both classes are too large and would serve as an intolerable number of exceptions to the other's patterns. b) Over-generalization is possible. Both classes are small, so the tolerance threshold is relatively permissive. Members of one class could be tolerated as exceptions to the other's patterns. c) Over-generalization is possible because one class is much larger than the other.

6.3. Analogy between Class IV and Class V

It is generally agreed that the lengthened *$\bar{e}$-grade was analogized from Class V to Class IV rather than vice-versa. In this section, I provide a perspective on why analogy proceeded in that direction by quantifying the likelihood of possible over-generalizations. A combined Class IV+V defined by the *-*eC*- root shape would contain 44 items and have a tolerance threshold of 11.62, so it could tolerate up to 11 exceptions before it failed to be productive. If it failed to be productive, then a child who hypothesized it would have to fall back on narrower Class IV and V generalizations.

However, if a child had, say, learned the past participles of only 5 Class IV verbs so far along with the past participles of just 9 Class V verbs (a plausible scenario given the relative sizes of these classes), then N for that child is $5 + 9 = 15$, and $\theta_N = 5.30$, so a combined V→IV holds up, while IV→V still would not. A child in this state of development would conclude that the combined Class IV+V forms past participles productively with *-*eCan*- and pasts with the lengthened *$\bar{e}$-grade albeit with a number of exceptions. This is analogy.

Since this analogy from Class V to IV is the result of a quantitative algorithm, it is possible to estimate roughly how common it was for Proto-Germanic learners. There are many learning states that a learner could land in to cause the analogy, but not all states are equally likely. For example, it would be unsurprising if a child had learned 5 Class IV verbs and 9 Class V verbs since Class IV is a little more than half the size of Class V, but it is also technically possible for a child to be in a position to postulate a Class IV+V according to IV's rule, for example having learned 6 Class IV verbs and 4 Class V verbs, though this seems unlikely.

Figure 34 plots the four possible states that a child could land in: separate Classes IV and V (cyan), combined Class IV+V with either rule possible (black), Class IV+V according to IV's rule (IV→V; yellow), or Class IV+V according to V's rule (V→IV; red). The x-axis is the number of Class V verbs learned so far for a given child, and the y-axis is the number of Class IV verbs learned, so that a child who has learned x Class V and y Class IV verbs falls

on cell (x, y) of the plot. Everyone starts out knowing zero verbs and eventually progresses towards an adult-size lexicon, so as visualized on these plots, a learner would progress from the bottom left knowing $(0, 0)$ verbs to the top right knowing $(28, 16)$. The bottom left of the plot is placed in the cyan zone because generalization is trivially impossible when less than two verbs are known.

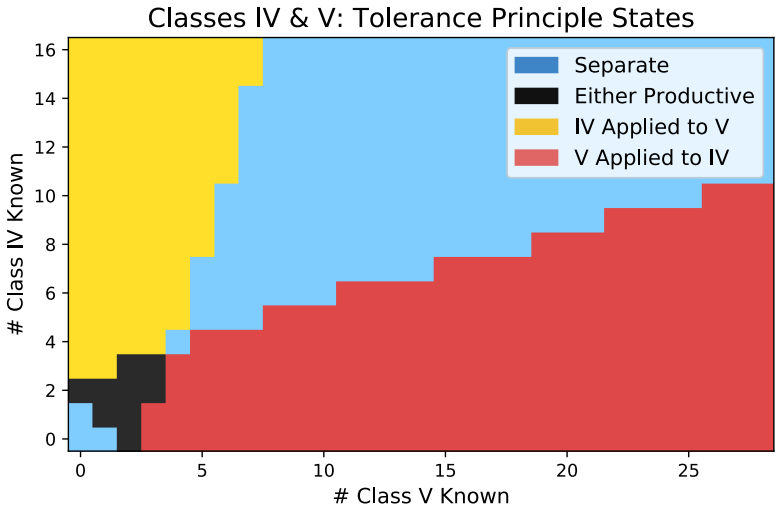


Figure 34: Tolerance Principle state space. Black: Either direction; Blue: Class IV and Class V as distinct classes; Red: V→IV; Yellow: IV→V.

A path roughly along the diagonal through this space, indicating that a child had learned members of both classes, is more likely than one which goes straight along one axis then the other. The likelihood of passing through each coordinate in the space can be formalized as well, which will permit us to calculate the probability of landing in each state for a given vocabulary size.

The problem of learning verbs from two classes is analogous to drawing two colors of models from an urn without replacement – whenever it is time to learn a new verb, either a Class IV marble or a Class V marble is drawn. This is modeled with a centralized hypergeometric

distribution.[4] Equation 6.1 gives the hypergeometric probability mass function which states how likely a given coordinate is given that the learner knows N verbs total so far from either class. N_{all} here is the total number of verbs in the corpus (|Class IV ∪ Class V| in this example) K_{all} is the total number of Class IV verbs (or Class V verbs; the function is symmetric), and K is the number of Class IV (or Class V) verbs learned so far. $N_{all} - K_{all}$ then is the size of the other class.

$$P(X = K) = f(K; N_{all}, K_{all}, N) = \frac{\binom{K_{all}}{K}\binom{N_{all}-K_{all}}{N-K}}{\binom{N_{all}}{N}} \tag{6.1}$$

Figure 35 visualizes the centralized hypergeometric distribution in the same coordinate space with dark colors signifying higher probability. Each top left to bottom right diagonal, a line of constant N, sums to 1. As expected, cells along the bottom left to top right diagonal are more likely than ones off to the edges. For example, 24.9% of learners who know 15 verbs know 5 from Class IV and 10 from Class V, but only 0.3% of those learners know 1 Class IV verb and 14 Class V verbs.

[4] A centralized (unweighted) distribution is a reasonable approximation as long as the *average* frequency rank of the members of each class is roughly equal. This is most likely the case for any two classes of non-trivial size unless the semantics of one lend them to more frequent use. Since none of the PGmc strong verb classes are defined by meaning, this is almost certainly a reasonable assumption. If this were not the case, the most likely path through the Tolerance Principle state space would have bowed out towards the top left or bottom right.

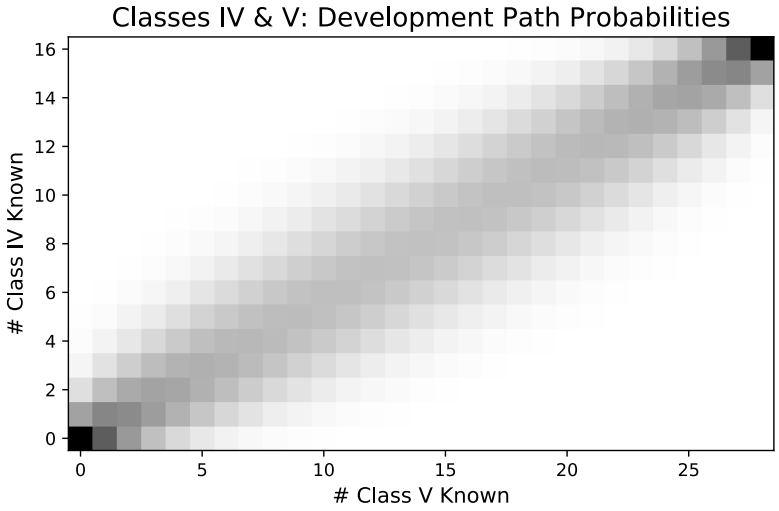


Figure 35: Development path probability visualization. Darker colors represent more likely states for each N.

When combined with Figure 34 this provides a visualization of how likely each kind of generalization is in Figure 36. Shades of red indicate cells where the learner analogizes Class V onto Class IV, the yellows are the cells where the analogy would have gone from Class IV to Class V instead, and the blacks are where either analogy was possible. The reds are by far the most likely kind of over-productivity, forming a large swash across the state space.

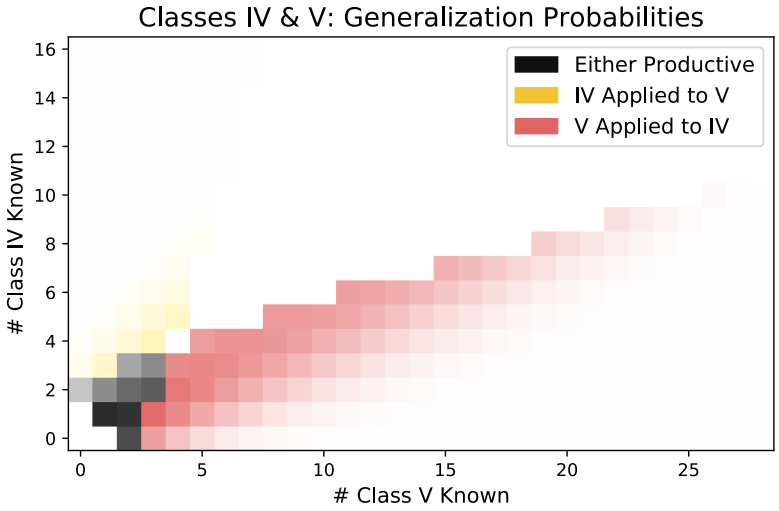


Figure 36: Tolerance Principle state likelihood space plot. Black: Either direction; Blue: Class IV and Class V as distinct classes; Red: V→IV; Yellow: IV→V.

Finally, summing up the probabilities corresponding to each color along each constant N diagonal yields the visualization in Figure 37, which is the likelihood of each Tolerance Principle state by vocabulary size N. Of the children who have learned, say, 10 Class IV and V verbs total, about 80% fill in unattested Class IV forms with Class V's inflections, about 10% go the other direction, and about 10% correctly hypothesize that Class IV and Class V are distinct. As they learn more verbs, the super-class over-generalizations become less and less likely, and the distinct synchronic system becomes dominant. By the time children have learned about 40 verbs, that is the only realistic option.

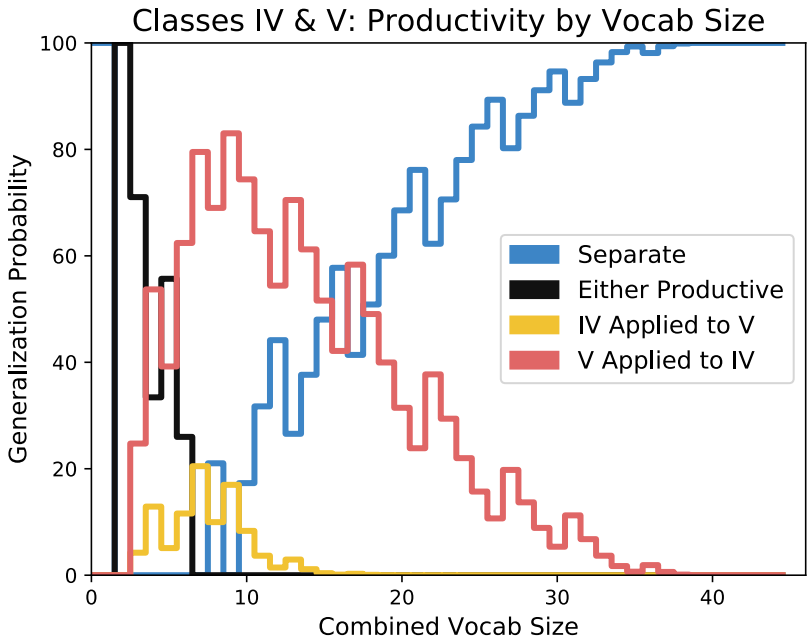


Figure 37: Tolerance Principle state likelihood by number of verbs learned. Black: Either direction; Blue: Class IV and Class V as distinct classes; Red: V→IV; Yellow: IV→V.

There are several takeaways from this plot. First, all learners are predicted to rapidly pass into and out of a phase in which either generalization is possible. Since this phase is only tenable very early on, it is unlikely that such children would even be verbal. Second, the V→IV generalization, the one that is considered to have occurred in Pre-Proto-Germanic, is always more likely than the other direction. It spends time as the dominant state, peaking at over 80% of learners, and remaining viable until most vocabulary have already been acquired. IV→V on the other hand never exceeds 20% and is only viable for a short period early on.

For illustrative purposes, if we assumed that the learning rate was constant and the population was well mixed, the areas under these curves would give a very rough estimate of what proportion of learners were in each state at any given time. 6.4%, all very young, could over-generalize in either direction, 2.2% also very young, were over-generalizing IV→V, 27.2%,

most young but a few older, were over-generalizing V→IV, and the rest, 64.3% skewing older, were not over-generalizing. The V→IV generalization corresponds to the direction of analogy that is believed to have occurred at some point in the prehistory of Proto-Germanic. Compared to the other direction, it was both much more common, and tenable late. The significance of late tenability will be discussed further in Section 6.6.2. For now, it suffices to say that only a generalization that is present in children who can communicate it to others has a chance at being transmitted to peers. There are two features of an over-generalizaton that increase its odds of being actuated in a speech community: if it is likely to occur at all, and if it is likely to occur relatively late.

The chance of analogizing between Class IV and Class V in one direction or the other was quite high (35.7%), suggesting that the status of the classes as distinct was a bit tenuous for learners.

If that was true, then the classes should have been somewhat permeable as individual items may have moved from one class to another as their past participles were guessed incorrectly. *Metaplasm*, the instability of the classes is a persistent state in attested historical Germanic, although most examples show IV→V analogy. My impression is that this would be due to the size of the classes in the attested languages, though this remains future work: It would be interesting to compute, say, whether Old High German Class IV was larger than its Class V. Consider the past participles Gothic *gabrukano* and Old English *brocen* (Modern *broken*. This may have happened in (Pre-)Proto-Germanic since it is attested in both East and West Germanic, motivating the reconstruction **brekanǫ* 'break.' There are also a few verbs in which the Class IV past participle vowel has analogized to the present in East and North Germanic, for example, Gothic *trudan* and Old Norse *troða* vs. Old English *treden* and Old High German *gitreten*. Finally, there is a large-scale movement of Old High German Class V verbs to Class IV (e.g., *gisprohhan* vs. Old English *sprecen*) and an instance of confusion in Beowulf 2981 *dropen* 'smitten' as opposed to the usual *drepen*.

6.3.1. Analogy between Classes III, IV and V

Class III shares an exclusive root generalization with Class IV+V based on an *-*eC(C)*- pattern, so it is conceivable that a child who succeeded with Class IV+V would try to build an even broader Class III+IV+V generalization. Since it seems like this change never happened – lengthened *$\bar{e}$-grades in Class III, and Class III's past *u never leveled over the *$\bar{e}$ of IV and V – it is important to account for why it did not while V→IV did.

A combined Class III+IV+V would contain $N = 96$ members with a tolerance threshold $\theta = 21.03$. The generalization leading to III→IV+V analogy fails because the 44 members of Class IV+V greatly exceed the threshold, and the generalization leading to IV+V→III fails for the same reason. Neither alternative paradigm for roots with *-*eC(C)*- succeeds, so a learner would have to fall back on the narrower generalizations. But of course, a learner would not have considered this hypothesis at $N = 96$ in the first place, because Class IV+V was not tenable when all Class IV and V members were known. Other combined classes III+IV and III+V were not possible because there is no root generalization that contains III with IV or V but not both.

It would be possible for some younger learners who could manage Class IV+V to entertain III+IV+V as well. But unlike for Class IV+V, the window in which this over-generalization was possible is quite small because of the larger sizes of the classes involved (cf. Figure 33). Calculating the TP state space for Classes III and IV+V quantify the point further.

Figure 38 visualizes the TP state likelihood space for Class III and Class IV+V with the probability calculations taken into account. Note that the swashes of red and yellow are much smaller in this plot than for Class IV and Class V in Figure 36. This already suggests that analogy was less likely here than it was between Classes IV and V.

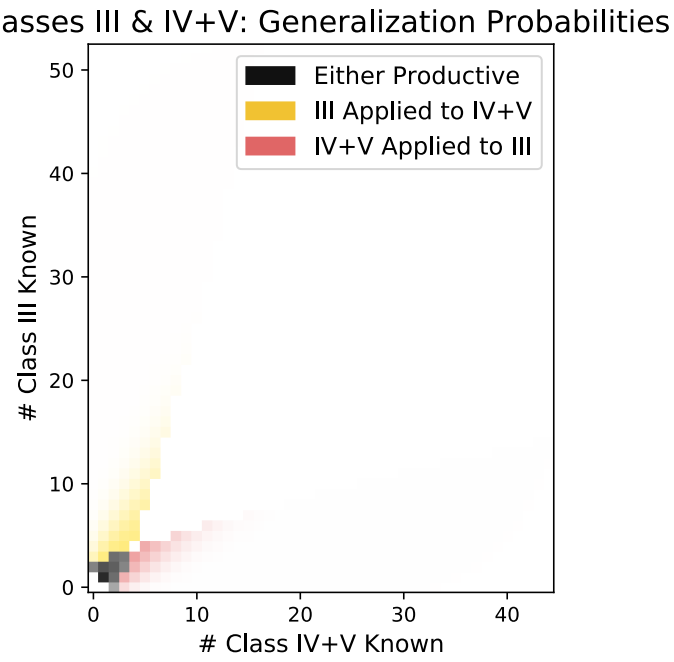


Figure 38: Likelihood of Tolerance Principle states for generalizations between Classes III and IV+V by number of verbs learned. Number of Class IV+V verbs learned on the x-axis and number of Class III verbs learned on the y-axis. Black: either generalization; Blue: distinct classes: Yellow: III→IV+V; Red: IV+V→III.

Figure 39 plots the probability of landing in each state by number of verbs learned so far and clarifies the reason why Class III was not involved in the analogy. The no-generalization grammar (blue) is dominant from a much earlier point compared to Figure 37, and all three over-generalization grammars are quite rare, all together summing to less than 10% of development time. Even more importantly, the over-generalization grammars are only feasible very early in development when less than roughly 20 verbs are known and the children are unlikely to be producing the inflections themselves, so an innovation was less likely to be symptomatic. It had little chance of entering the speech community.

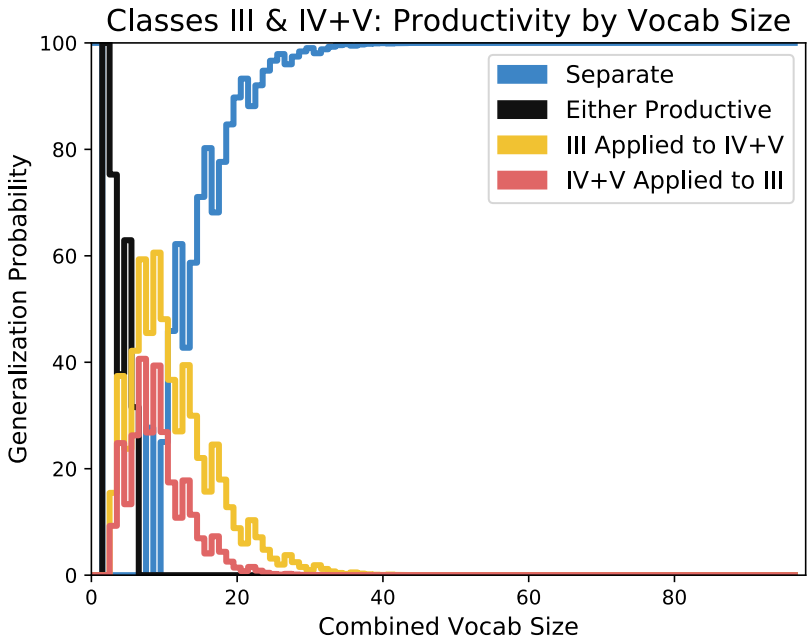


Figure 39: Likelihood of Tolerance Principle states for generalizations between Classes III and IV+V by number of verbs learned. Number of Class IV+V verbs learned on the x-axis and number of Class III verbs learned on the y-axis. Black: either generalization; Blue: distinct classes; Yellow: III→IV+V; Red: IV+V→III.

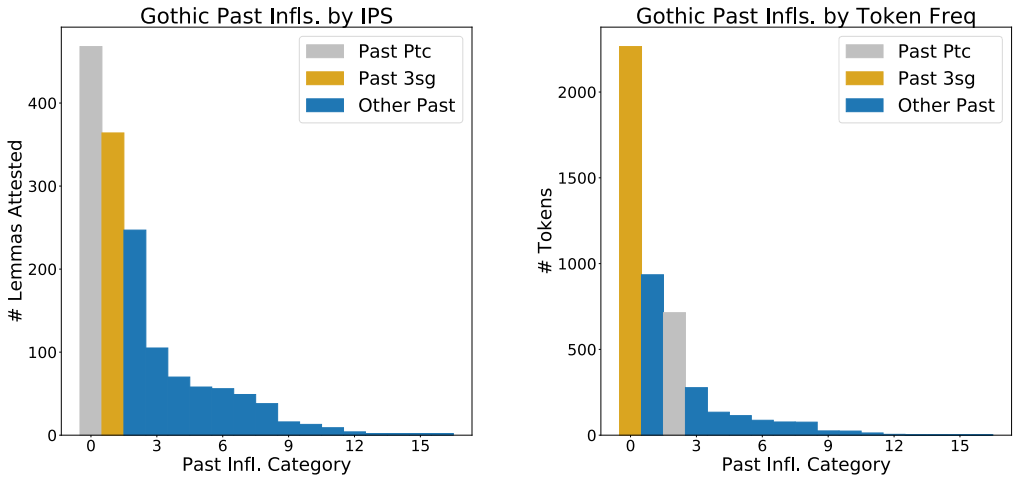
There is no hard constraint preventing the PGmc past stem vowel from analogizing to or from Class III, but it would have been very unlikely. The analogy model presented here based on the Tolerance Principle makes that clear. All over-generalizations except for the one that actually happened could only have been entertained by a small fraction of learners. Furthermore, only early learners who were likely not yet inflecting verbs in their speech were the only ones who would have entertained the other plausible over-generalization. And since analogical change requires child errors to gain a foothold in the population, if the errors were not embraced by learners who could speak with one another, those over-generalizations did not stand a chance.

6.3.2. The Past Stem vs. The Past Participle Stem

One issue left outstanding is why only the Class V past stem's lengthened *$\bar{e}$-grade spread to Class IV and not its *e*-grade past participle, since under the analogy model, the past participle could conceivably have spread in the same way that the past did and for the same reasons. I can only speculate on the reason from the perspective of inverse paradigm saturation.

An inflectional category with higher paradigm saturation is more likely to be attested in the input with any given lemma, and the more lemmas that are attested in a given category, the fewer of those forms need to be inferred. The productivity model of analogical change requires over-generalization during the inference step, so the higher a category's IPS, the less able it is to be analogized away. So, if the past participle had high IPS, then many Class IV pptcs would have been attested to the learner, and these could have been memorized as-is even under a V→IV over-generalization. Additionally, if past participles were more frequent on average than other inflectional categories, then a given inflected form would be more likely to be attested and to be attested early.

Unfortunately, we have no Proto-Germanic corpora over which we can estimate these metrics, so it is impossible to say for sure what the distribution of the past participle was in the language. However, we can evaluate its status in Gothi using the same morphologically tagged tagged dataset from Section 3.3. This comes with significant caveats, since IPS may vary by genre, and the text in the UD corpus is entirely drawn from translations of the Greek Bible, but it is a useful starting point. Figure 40 shows IPS and token frequency plots for Gothic past inflectional categories.



(a) IPS of Gothic Past Inflections

(b) Token Frequencies of Gothic Past Inflections

Figure 40: Inverse paradigm saturation and token frequency plots for Gothic past inflections. Colors correspond to Proto-Germanic strong verb stems. The highest IPS other past form is the past active indicative 1sg, and the most frequent is the past active indicative 3pl.

It turns out that the past participle has the highest IPS in the Gothic Bible, corresponding to nearly a third (468/1505) of past inflected verbs in the corpus. Additionally, it is the third most frequent past form in the corpus after the past 3sg and past 3pl. This lends credence to the notion that the pptc's higher IPS and token frequency protected it against analogical leveling.

The problems surrounding V→IV past stem analogy are common to the various analogical accounts for the distribution of the lengthened *$\bar{e}$-grade. So far, we can now conclude that V→IV analogy was far likelier to have occurred than the opposite direction, that further analogy with Class III was unlikely though possible, and that the past participle was less likely to be analogized than the past stem. In the next two sections, I address the specific predictions of the 'Eat' and Class VI Analogy hypothesis using the tools developed here.

6.4. The 'Eat' Analogy

The 'Eat' Analogy hypothesis proposes that the lengthened *$\bar{e}$-grade spread from **etanǫ* 'eat,' a single Class V verb, to the rest of Class V before it spread to Class IV. This section explores which assumptions would have to be made, given the analogy model adopted here, for the 'Eat' Analogy to be viable. It can then be compared to the assumptions behind the Class VI analogy.

The analogical extension of the lengthened *$\bar{e}$-grade from a single verb to the entire class must have required a propitious alignment of the stars. One helpful facilitator would have been the presence of additional now-lost vowel-initial Class V verbs besides **etanǫ* (Ringe, 2017). This would have made the task immensely easier, but their existence can be nothing more than speculation. Any path of change that is sufficient for the single vowel-initial verb would have been even more workable for more vowel-initial verbs, so the system as reconstructed with the single vowel-initial Class V verb can be seen as the floor or baseline scenario for the change. In reality, the situation was at least as favorable or perhaps more so.

Possibly the most crucial contributor to the analogical spread of the *$\bar{e}$-grade within Class V is hinted at in a single brief footnote in Mailhammer (2007, fn. 101) which describes a personal communication between that author and Theo Vennemann drawing attention to a few Class V verbs which rhymed with **etanǫ*. The significance of this observation was not elaborated on, but the presence of rhyming Class V verbs, namely **metanǫ* 'measure,' **fetanǫ* 'fall,' and **getanǫ* 'get, receive,' suggests a stepping off point for the change.

The Tolerance Principle itself has nothing to say about the actuation of a change from a single item, but the notion of successive levels of generalization is applicable here. While a pre-*$\bar{e}$-grade past Pre-PGmc. system could not straight away extend the *$\bar{e}$-grade past from one verb to all verbs with *-*eC*- or even *-*eT*- roots, the rhyming pattern *-*et*- is much more manageable. If the *$\bar{e}$-grade was able to spread to its three rhyming verbs, it would have

had basis it needed to extend to broader categories. At this point, the Tolerance Principle does come into play.

There is a whole range of possible generalizations between *-*et*- and *-*eT*-, for example, those roots whose final consonants are only voiceless plosives, voiceless obstruents, voiceless coronals, any coronal obstruents, or any plosives. Lengthened *$\bar{e}$-grade pasts conceivably could have been rendered as productive within any of these sub-classes of Class V on the way to productivity across all of Class V over succeeding cohorts of learners. I consider all these alternatives summarized in Table 16. Assuming that all Class V verbs except for 'eat' and its three rhymes formed pasts in *-*UT*-[5] at this stage, all members of the hypothetical sub-classes except for 'eat' and the three rhyming verbs would have constituted lexical exceptions.

Class	**N**	θ_N	**e=N-4**	**e=N-8**
***-e[-voi -cont -son]-**	7	3.60	3	-
***-e[-voi -son]-**	19	6.45	15	12
***-e[-voi COR]-**	11	4.58	7	-
***-e[-cont -son]-**	12	4.83	8	4
***-e[-son COR]-**	12	4.83	8	-

Table 16: Logically possible sub-classes of Class V between *-*et*- and *-*eT*- with initial exception counts. The column $e = N - 4$ indicates the number of exceptions at the *-*et*- stage, and $e = N - 8$ indicates the exceptions at the *-*e*[-voi -cont -son]- stage.

There are only four other regular verbs with voiceless plosives in the reconstructed Class V lexicon: **brekanǫ* 'break,' **lekanǫ* 'be leaky,' **rekanǫ* 'bank a fire,' and **wrekanǫ* 'drive out.'[6] These are exactly the three verbs that would serve as exceptions if a learner tried to create a voiceless plosive subclass of Class V with the *$\bar{e}$ past from 'eat' and those three verbs that rhyme with it. Since 3 is less than the tolerance threshold for 7, productivity among these verbs was feasible if a learner happened to acquire the pasts of the *-*et*- verbs before the past of the *-*ek*- verbs. Since this only had to happen once in history, and since

[5] where U is whatever ablaut grade existed before the *$\bar{e}$-grade

[6] Though **brekanǫ* had an unexpected past participle **brukanaz*. If it is excluded, the numbers in the rightmost column of Table 16 increase by one.

in all likelihood 'eat' and 'fall' were more common than 'be leaky' or 'drive out,' this is a sufficiently plausible scenario. Once the *$\bar{e}$ past reached productivity among those 7 verbs, it was within range of jumping from voiceless plosives to all plosives ($N = 12$, $e = 12-8 = 4$, $\theta_N = 4.83$) or picking up voiceless coronals, either of which put it closer to spreading to the rest of Class V.

There are a few additional factors that could have helped this process along. First, the verb 'eat' must have been quite common, almost certainly more common than than the other *-*et*- verbs if frequencies in modern corpora have any bearing. This means that even young learners were almost certainly familiar with the past stem for 'eat' even if they had not yet heard the pasts of the other verbs. The other rhyming verbs must have been nevertheless very common, so verbs with *$\bar{e}$-grade pasts would have formed an out-sized fraction of early learners' Class V lexicons after productivity extended to the 8 voiceless plosive verbs.

Second, and this is just speculation, the Pre-PGmc Class V past **-*UT*- stems may not have constituted a unified front against *$\bar{e}$. If the **U vowels arose as a repair for unpronounceable stems, they did not necessarily have to all share the same vowel. If, say, the vowel were **u in some contexts and **e in others, or if some speakers chose one and some the other, there may have been no default form in Pre-PGmc. Class V might not have been able form a productive past stem at all, instead resorting to lexical stem formation even before the *$\bar{e}$-grade past became viable. In that case, the *$\bar{e}$ past rule which was grounded productively in 'eat' and its phonological neighbors, would have had a substantial advantage.

The very first step in the expansion of the *$\bar{e}$-grade past from 'eat' to rhyming verbs remains speculative as well, but the Tolerance Principle suggests a path towards its rise in Class V. Additionally, the two or three mitigating factors described here, while not enough to form a satisfactory account of the change on their own, provide an additional nudge for learners working out the system according to the TP. Interestingly though, 'eat' had an *$\bar{e}$-grade past.3sg which did not spread. Once again, we can only speculate about why. Like with all change, it must have happened in the right place and at the right time if it gained a

foothold in the population. Slight perturbations to the system could easily prevent a rare once-in-a-generation occurrence from happening at all.

There are few such perturbations that could have put the *$\bar{e}$ past just over the edge but left the *$\bar{e}$ past.3sg behind. As discussed in Section 6.3.2, higher IPS and higher token frequency both protect against analogical replacement, IPS because more verbs are likely to be attested in a given inflectional category, and token frequency because a given verb is more likely to be attested in that category early on. It turns out that, at least in Gothic, the past.3sg had both a higher IPS and token frequency than any past stem inflectional category.

In summary, the 'Eat' Analogy is technically possible but extremely unlikely.

6.5. The Class VI Analogy

The Class VI Analogy hypothesis has an obvious advantage over the 'Eat' Analogy in that it does not require an inflectional pattern to claw its way up from a single verb to a whole class of verbs. Class VI is larger than Class V, which would make it a case of analogical leveling. On the other hand, it requires that the length, and only the length, of the Class VI past stem vowel analogize to Class V. This is enough for Mailhammer (2007) to reject the Class VI Analogy outright, but I entertain it anyway for now in order to work out what the analogy model has to say about it.

Before even attempting the Tolerance Principle, it should be noted that like 'eat,' the typical Class VI past.3sg stem contained the same long vowel as the past stem, so some explanation for why the past.3sg was not analogized has to be offered. The two analogy hypotheses are tied in that respect, or the 'Eat' Analogy might have a slight edge because it is easier to see how something might fail to spread as a matter of chance from a single verb than from a whole class of them.

Now, in order to calculation the likelihood of over-generalizaton from Class VI to V, a root

generalizaton needs to be present. Class V has a root shape *-eT-, and Class VI has *-aC-. The minimal root shape generalization between these is either *-$\breve{V}C$- with the short vowel to exclude Class I or *-[-hi]C-. Here, the model is already at an impasse: both of these generalizations would also include Class IV. It is not possible to define a root generalization that includes exactly Classes V and VI without also including IV, so the TP can never predict an analogical change that spread from Class VI to Class V then to Class IV. The best it can do is a spread from Class VI to Class IV+V, which creates a conflict with previous work that confidently shows that there was analogy from Class V to IV during this process.

Nevertheless, I apply the Tolerance Principle to Class IV+V and Class VI together for completeness. Table 17 shows the sizes of the classes under consideration. Class VI is slightly larger than Class V and is about two-thirds the size of the combined Class IV+V.

Class	**N**	θ_N
IV	16	5.77
V	28	8.40
VI	29	8.61
IV+V	44	11.62
IV+V+VI	73	17.01

Table 17: Class IV, V, and VI and combined class tolerance thresholds.

The results of the calculation are shown in Figure 41. According to this, analogy should have gone in the wrong direction if anything happened at all. The reverse analogy was tenable for speakers who knew fewer than 40 verbs in these classes, but it was only feasible with high probability for vocabularies about half that size. So even if we forgive the vowel length-only analogy which prior literature criticizes, the Class VI Analogy is not workable. Not only is VI→IV+V very improbable, it cannot account for a subsequent V→IV change. Even though it seems to offer a more intuitively reasonable story than the 'Eat' Analogy, it has actually has no advantages over it in practice when calculated out.

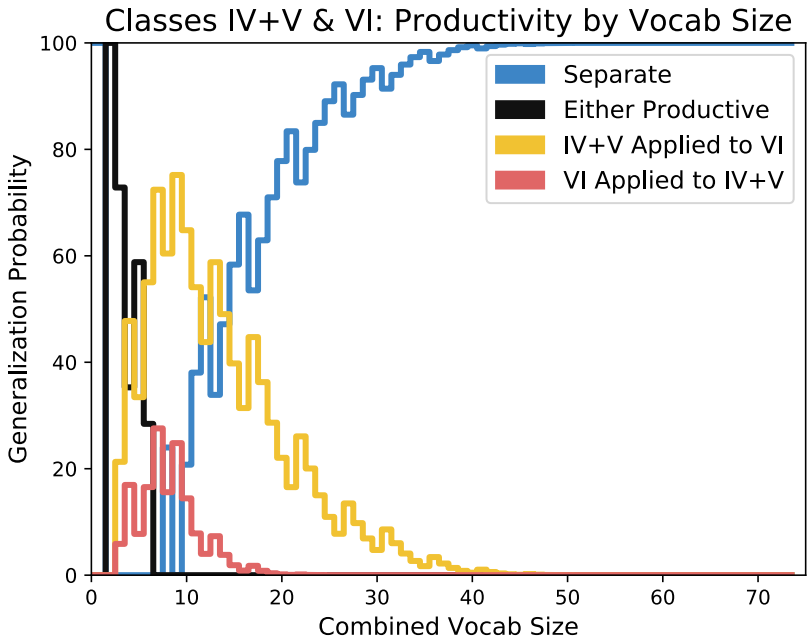


Figure 41: Likelihood of Tolerance Principle states for generalizations between Classes III and IV+V by number of verbs learned. Number of Class IV+V verbs learned on the x-axis and number of Class III verbs learned on the y-axis. Black: either generalization; Blue: distinct classes; Yellow: IV+V→IV+V+VI; Red: IV+V+VI→IV+V.

6.6. Discussion

The model introduced in this chapter for acquisition-driven analogical change provides new insights into the origin of the Proto-Germanic strong verbs' lengthened *$\bar{e}$-grade. In developing a model for analogical innovation based on the normal process of productivity learning, we can now differentiate between competing hypotheses with a degree of rigor that was previously unavailable. This quantitative analysis of the 'Eat' Analogy and Class VI Analogy hypotheses finds that the former is workable, but the latter is not possible. This work has implications for explanations of historical phenomena more broadly.

6.6.1. Concrete Mechanisms in Historical Explanation

It is the step of committing ourselves to a concrete acquisition mechanism that opens up new avenues for analysis. In Chapter 4, applying the Tolerance Principle to the acquisition of /aɪ/-raising brought together seemingly contradictory facts from competing accounts. Phonological acquisition in the face of even minimal variation explains the innovation of transparent /aɪ/-raising and its sparse distribution. In Chapter 5, connecting the innovation of the *to*-dative in Middle English to transient over-generalizations in modern child productions explains not only its rapid rise in historical corpora, but also its broad Middle English distribution, something dismissed as irrelevant in previous studies. It further challenges functional treatments that link the *to*-dative to the loss of overt case marking with an afunctional account that better accounts for the cross-linguistic facts.

In this chapter, adopting a quantitative model of productivity learning raises the bar for rigor in historical explanation. Several accounts for the lengthened *$\bar{e}$-grade have been proposed over the last century and a half, many of which are indeed consistent with the descriptive facts. If two treatments are consistent with the facts, and both are intuitively plausible, how can we distinguish between them? The productivity-based model adopted here forces us to favor one account over another. Under its conditions, the Class VI Analogy turns out to be not only unlikely, but actually incompatible with later analogy from Class V to IV, while the 'Eat' Analogy does seem to have been possible, albeit very unlikely.

I used the Tolerance Principle here because of its strengths as a model of productivity learning, but the real moral of this approach is not limited to that. If one commits to a reasonable acquisition model based on study of modern learners and speakers, it will make predictions that can be tested. Diachronic research will never have access to the full gamut of sources of evidence available to synchronic researchers, but applying what we know about the present to the past raises the level of rigor in historical hypothesis testing closer to what can be achieved in the present.

6.6.2. "Sibling-Induced" Change

The actuation of a change requires both its innovation and its entry into a speech community (Labov et al., 1972, p. 7). This gets at the Paradox of Language Change. Even if children are capable of innovation, through over-generalization for example, how can they be responsible for propagation? Children are famous for the impressive accuracy in acquisition, so we expect them to grow out of any novel patterns that they innovate. To be sure, children usually do grow out of their innovations, but even if they do not, who would pick up those innovations from a small child? I have hinted at a way out of the actuation Paradox a few times up to this point. Now is the time to lay it out.

There are a few ways out of the innovation problem. The first is to "blame the learner," that is, that learners sometimes introduce errors into the acquisition process despite receiving enough evidence to make the right choices. The second option is to "blame the environment" instead, that is, to drop the assumption that children learn in ideal single input source environments. We know that children receive input from multiple people who themselves may exhibit internal variation, and that change is formally *inevitable* in the face of even minimal variation in the input (Niyogi and Berwick, 1997). Further, the effect of the Poverty of the Stimulus is not to be discounted. Even the richest innate specifications to the language faculty render acquisition tractable, not trivial. Measures of sparsity such as paradigm saturation (Chan, 2008) show that much of what a child is tasked with learning is simply not available to them, and cases of divergent grammar outcomes under Abject Poverty suggest that learners really are operating at the edge of what is feasible.

The blame for child innovation in morphological systems can be pinned on input sparsity. The number and diversity of forms available to any given learner is unlikely to be enough to uniquely specify even a moderately large paradigm, so it is normal, over the course of development, for a learner to over-generalize and innovate new forms. The numbers work out this way for Proto-Germanic and also Latin in the upcoming chapter. If there is even a one in a million chance that some children never grow out of their over-generalizations and

transmit them to others, then changes will accrue over time.

Now what might prevent a child from growing out of an over-generalization? The most important thing to recognize here is that language transmission is not strictly generational (Manly, 1930; Weinreich et al., 1968; Roberts and Labov, 1995; Labov, 2001; Nardy et al., 2014). Other children transmit linguistic features to their peers. The impact of non-generational transmission is best conveyed with a thought experiment. Consider two Pre-Proto-Germanic children, Alice and Bob,[7] and say Alice is Bob's older sister. Alice is currently entertaining a grammar with the Class IV+V, so she is producing would-be Class IV verbs with lengthened *$\bar{e}$-grade past stems. How might little Bob react to Alice? Crucially, Bob may not even be able to recognize Alice's innovation since he will rarely (if ever, for some verbs) receive an adult's conservative token corresponding to one of Alice's innovations, and since Alice is mostly consistent with adults – she presumably can communicate with her parents – there is no reason to assume that Alice is acting oddly. Paradigm saturation is severe, and in a language with a paradigm the size of Proto-Germanic's, the average verb is only attested with one forms in even a million words (Section 3.3).

If Bob cannot recognize Alice's innovation, then he may adopt it. At the very least, he is receiving unambiguously innovated input, so he is more likely to make a similar innovation himself or extend Alice's, as in Chapter 5's study. If he does recognize it, he has options: he can adopt it categorically, adopt it as some kind of socially influenced variant, or reject it. Even young children do begin to orient towards their peers rather than their parents, a pattern that continues through adolescence (Labov, 1989; Roberts and Labov, 1995; Nardy et al., 2014). So as long as Bob does not find Alice's innovation egregiously ungrammatical, he may choose to adopt it if he values Alice's social prestige. If a three or four year-old Alice is cool to anyone, it will be two or three-year-old Bob. An example of this older-to-younger-sibling transmission process may be found in the Sankoff and Blondeau (2007) Montreal French /r/ study in which it is argued (§7.2) that the first cohorts of speakers to acquire

[7] or *Ermunahildiz and *Hrōþiwulfaz, if the reader prefers

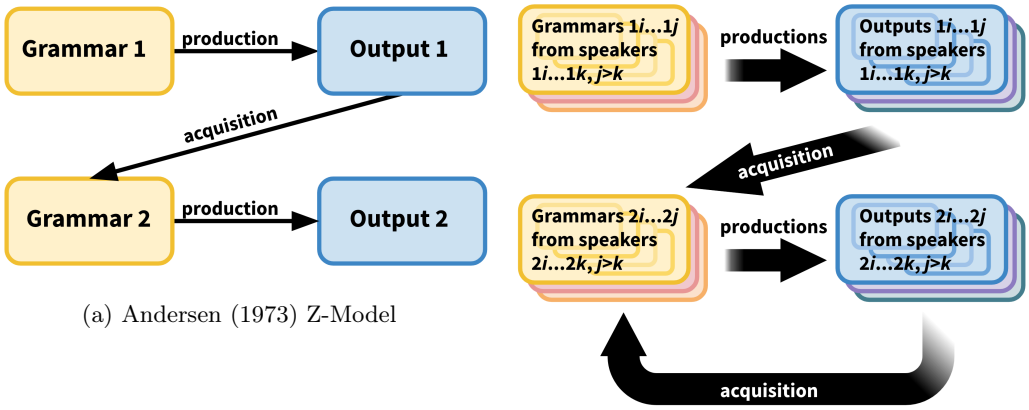
categorical [ʁ] (the innovative variant) probably acquired it from older siblings.[8]

Even if Bob learns the conservative form later, he may keep the innovative form as a sociolinguistic doublet. That depends on domain. In the case of morphology in particular, doublets are well-attested, suggesting that learners can accept multiple grammars. Within the history of English, these include the post-PGmc confusions between Classes IV and V, sustained variation in the weak verb system of Old and Middle English (Taylor, 1994), and fossils in the modern language such as *cloven-cleft*, *worked-wrought*, and ongoing variation including *dived-dove*, *brought-brang*, and *sneaked-snuck*. Since morphological doublets are so common, it seems most likely that Bob would internalize Alice's novel forms as well as the adults' once he heard them. When Bob matured, he would have two forms to choose from and might produce both around the children of the next generation. An Alice Jr. and Bob Jr. would receive both forms from the adults of the community as well as each other granting the innovative forms a foothold in the community.

Transmission of an innovation from slightly older children to their younger peers constitutes actuation. All that this "sibling-induced" framework of acquisition-driven change requires is that learners receive input from multiple individuals, and this, of course, is an entirely normal part of the acquisition process. Under this sibling-induced framework, over-generalizations that are more likely to be innovated are more likely to be transmitted. Perhaps more importantly (it remains to be seen), over-generalizations that are tenable relatively late in development are more likely to be transmitted as well. If Alice entertained an innovative hypothesis for a while but moved on before she was old enough to speak fluently or old enough for Bob to look up to her, then that hypothesis does not stand a chance at transmission. This is why the late tenability of the Class V→IV analogy (Figure 37) sets it apart from all of the failed analogies which are not tenable late. It is the only one that could realistically happen.

[8]Though in this case, the innovation was external to the community. The innovative variant was already standard in other parts of Quebec and the Francophone world more broadly.

Finally, Sibling-induced change reveals some insufficiencies in Andersen's Z-model of change: First, individuals may vary in their productions, both across their lifetimes, and across social settings, so learners may receive variation even from any given input source. Second, learners are embedded in speech communities, so they receive input from multiple individuals, input that is sure to contain at least minimal variation. Third, acquisition takes time, and immature learners can influence other learners. The updates to the Z-model yield the visualization in Figure 42. Like the traditional Z-model, the production and acquisition pathways continue on indefinitely.



(a) Andersen (1973) Z-Model

(b) "Sibling-Induced" Z-Model

Figure 42: The Z-model extended for "Sibling-Induced" Change. As with the classic Z-model, it continues indefinitely in a chain. Thick arrows indicate bundles of individual arrows, and these may also skip "generations."

Crucially, the loop back from the outputs of peers or the same or slightly older age cohorts breaks generational transmission and provides an avenue for innovations to enter the local speech community. This is actuation.

6.6.3. Proof-of-Concept for Sibling-Induced Change

The Alice and Bob thought experiment is just that, something we can reason through but not demonstrated in practice. The first step towards validating the sibling-induced Z-model

is some kind of proof-of-concept demonstration. That is how I conclude this discussion.

One of the key properties of analogical change is that productive patterns will tend to level unproductive ones. Nevertheless, irregular unproductive patterns do exist. Interestingly, these are not randomly scattered throughout the lexicon as one might expect if they were innovated stochastically by sound change, rather, morphological irregularity tends to occur among high token frequency items (e.g., Bybee, 1985). These irregulars are very often inherited rather than innovated, so from a diachronic perspective, their high frequency could be seen as protective against analogical leveling. This *Conserving Effect* of high frequency has been leveraged to explain the distribution of English strong verbs (Bybee, 1985) and Romance irregular past participles (Laurent, 1999, §6.6) among others (e.g., Bybee, 1995; Zuraw, 2003; Dahl, 2004; Diessel, 2007).

The correlation certainly exists – there is no doubting that – but the cause is more elusive. Proponents of usage-based theories (cited above) prefer a direct causal role for token frequency via entrenchment or some other mechanism, however, there are other options.

Consider the role of type attestation. Forms that are not actually present in the input have to be inferred, while those which are present can be memorized. If a learner infers a form (i.e., applies a productive pattern to it) rather than faithfully acquiring a parent's or older peer's irregular, that would be an example of innovative analogical leveling of the type described in this chapter and the next. Played out across populations over time, this could produce the Conserving Effect with reference to type frequency rather than token frequency.

From the study of paradigm saturation (Section 3.3, Chan, 2008), we know that most forms are likely *never* to be provided for a learner, so inference plays a major role in the acquisition process. If an item has low saturation, nearly all of its forms must be inferred, that is, they must be generated by some productive process. On the other hand, if an item has high saturation, most of its forms are attested and so can be memorized. A form that is attested to a given learner during the acquisition of morphology has the opportunity to be irregular,

but a low frequency one does not.

The obvious question is an invocation of the Paradox of Language Change: if this is a matter of type attestation, then why would one not grow out of it after eventually hearing the form from others? This can be accounted for in the framework sibling-induced change.

I carry out a simulation as proof of concept. The goal is to test whether type attestation together with sibling-induced change. Imagine a scenario in which some a class of verbs or other inflectible words contains 100 items, some fraction of which have an irregular inflectional category, say due to a sound change, that has gone to completion. These irregulars are uniformly scattered throughout the class.

This language is spoken by a local community of 100 members who vary in age, some of whom are children still acquiring the language. For every iteration of the simulation, a new learner is "born," the other community members are incremented in age, and the oldest member is culled. Only the youngest few members are "learners," and one graduates to maturity in each iteration. For every iteration, new and remaining learners each receive a sample of 10,000 inputs drawn from the class of items according to a Zipfian frequency distribution. To implement sibling-induced change, the inputs are drawn from all community members older than the learner.[9] Crucially, young learners receive input from older learners as well.

Each learner receives zero or more instances of each item on each iteration. These instances may be irregular or regularized, and the learner adopts the majority form for each item following Section 4.2 before applying the Tolerance Principle to decide whether a pattern can be productively employed to infer the forms of the unattested items. If an irregular form was not attested, it will be inferred to be regular if the TP predicted productivity of the regular pattern.[10] Token frequency information was not tracked by the learner and was not

[9]Samples were taken uniformly, inversely with age rank in one simulation, and by a Zipfian distribution by reverse age rank in another. Both produced similar and significant results. More input from younger community members resulted in more regularization.

[10]If the TP did not predict productivity, all unattested items were left as gaps. If an item was gapped for an existing speaker, it could not be sampled as part of a learner's input sample. Gaps were rare for mature speakers under the parameters presented.

factored into the inference of unattested forms in any direct way. This is a test of whether type attestation alone can produce the Conserving Effect.

The whole simulation was run for for 100 iterations, so that none of the initial community members were still present at the end, and the forms of the initially irregular items as learned by the youngest mature speaker were recorded. Items that were irregular for community members at the start of the simulation may or may not have been regularized by the time the youngest speaker acquired them. The simulation was repeated for 500 trials and the final outcomes averaged to get probabilities of regularization by frequency rank. Two experiments were carried out, one where 10 items were initially irregular and one were 20 were. These numbers were chosen to lie below the tolerance threshold for 100 ($\theta = 21.7$).

Figure 43 shows the regularizing outcome for each initially irregular item for both the 10-irregular and 20-irregular simulations, with x-axis indicating sampled token frequency rank and y-axis indicating rate at which the items remained irregular. Linear models predicting irregularity by token frequency ($p = 0.018$) or log token frequency find significant effects ($p = 5.32e - 08$). A likelihood ratio test of a mixed effect model of token rank, token frequency and number of initial irregulars as a random intercepts found a significant effect for token rank ($p = 4.693e-13$) but not token frequency ($p = 0.3979$). A mixed effect model with log token frequency instead finds the same pattern. The two are highly correlated with one another (-0.52).[11]

[11]Number of items = 100, Community size = 100, Number of learners per iteration = 3, items sampled per learner per iteration = 10000, item sampling was Zipfian, community sampling was proportional to the inverse of age rank.

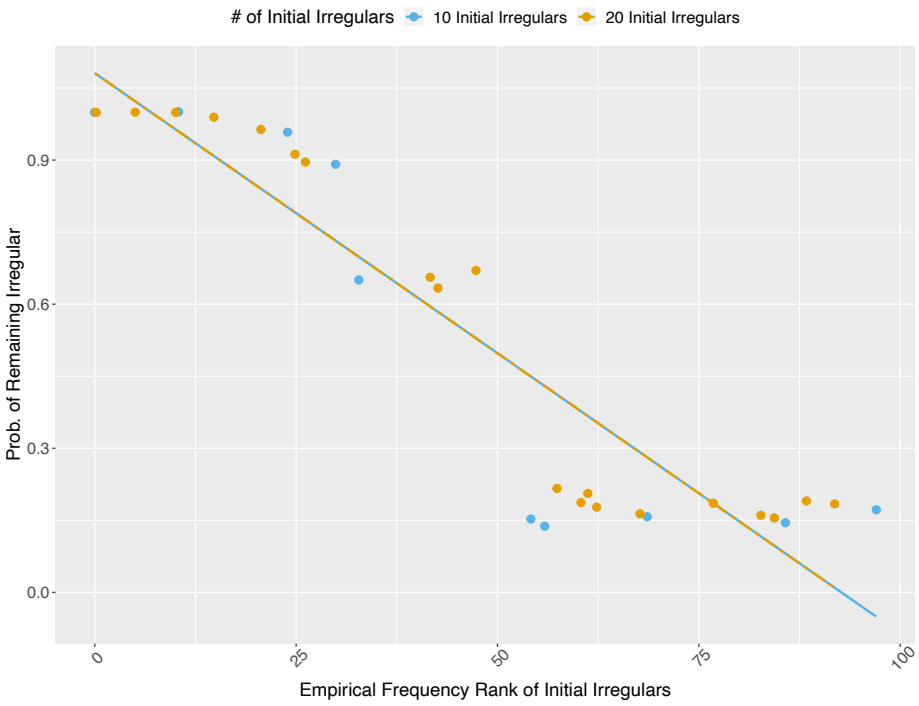


Figure 43: Relationship between token frequency rank and irregularity in a simulation of Sibling-Induced Change with Tolerance Principle learners and input sparsity.

These initial steps into the investigation of sibling-induced change show that, together with the Tolerance Principle and sparse type attestation, it is capable of reproducing the Conserving Effect of token frequency against analogical leveling. The task of fully characterizing the predictions of sibling-induced change and testing them against empirical data is left for is left to future work.

The next chapter builds on the lessons learned for Proto-Germanic to answer several questions about the Classical Latin inflectional and derivational morphology. A Tolerance Principle analysis of the past participles and so-called *t-deverbals* has implications for both diachrony in Late Latin and Romance and for theoretical treatments of the Classical system.

CHAPTER 7 : The Latin Past Participles and "*t*-Deverbals"

This final study builds upon the lessons learned in the previous four to discuss the synchronic productivity of the Classical Latin past participles and form correspondence between the past participles and so-called *t-deverbals*. This leads to a new productivity-based treatment of the form of the past participles and of the past participle-*t*-deverbal correspondence and has implications for the development of the past participle in Late Latin and Proto-Romance.

The forms of the Latin past participle are notoriously varied and difficult to predict. As such, they have inspired decades-worth of treatments in theoretical morphology and have even served as motivating case studies for some approaches (Matthews and Matthews, 1972; Lieber, 1980; Mel'čuk, 1982; Aronoff, 1994; Embick, 2000; Steriade, 2016). One reoccurring issue in prior studies is the role of regularity in the system: it is universally agreed that some forms are not predictable and have to be memorized or listed in some fashion. It is also clear that some generalizations exist in the system regarding theme vowels, root length, perfect stem formations, or other patterns, but it is far less obvious which of these patterns deserve status in the grammar.

The *t*-deverbals, a term coined in Steriade (2016), are a semantically heterogeneous set of deverbal agent (e.g., *doctor* 'teacher'), event (*dissertātiō* 'discourse'), result nouns (*cultus* 'cultivation'), and adverbs (*statim* 'immediately') among others which nevertheless faithfully share their form with their corresponding past participles (*doctus*, *dissertātus*, *cultus*, *status*). This pattern is reliable to the extent that even clearly irregular and suppletive past participles have corresponding *t*-deverbals. So, a theoretical treatment needs to account for regularities and irregularities in the form of the past participles on one hand, and the relationship between the past participles and their corresponding yet semantically diverse *t*-deverbals on the other.

Finally, there is a diachronic problem to be addressed in conjunction with the synchronic ones: along the way from Latin to Romance, the forms of the past participles were subject

to significant analogical change which removed most of the apparently irregular forms. This in itself would be unsurprising if it weren't for the fact that a relatively rare Classical past participle form, *-ūtus*, was extended at the expense of *-ītus* and bare *-tus*, the two most common forms in some of the conjugations (Weiss, 2009; Laurent, 1999).

The basic task that a theoretician faces in deciding what should be treated as regular or irregular is analogous to the one that children encounter when acquiring their native grammars, which suggests a learners' perspective approach to these problem. By grounding the split between regulars and irregulars in child language acquisition rather than by researchers' intuition, we gain new insights into the both the synchrony and diachrony of the system.

This chapter is a culmination of those that came before it. Building upon the /aɪ/-raising study in Chapter 4, I leverage a quantitative approach to child language acquisition to generate concrete predictions about learning and change. Following Chapter 3 and the *to*-dative study in Chapter 5, I use Latin data to model child learners whose input data may support innovative hypotheses. And following the morphological study in 6, productivity and input sparsity are implicated in analogical change.

Sections 7.1 and 7.2 begin with a synchronic and diachronic overview of the Latin verbal system with emphasis on the past participles and *t*-deverbals. Next, Section 7.3 summarizes influential and recent theoretical work on the past participles and the *t*-deverbal correspondence. It highlights how four theories in particular, Lieber (1980), Aronoff (1994), Embick (2000), and Steriade (2016), balance the regularity and irregularity and relate the past participle stem to the others. Following that, Section 7.4 discusses the Latin data set quantitatively analyzed here along with the results of a productivity analysis. The theoretical implications of this treatment are laid out in Section 7.5, and its diachronic predictions in Section 7.6. Finally, Section 7.7 discusses some higher-level implications of these results.

7.1. The Synchronic Classical Latin System

The Classical Latin verbal system was significantly reworked from Proto-Indo-European (Weiss, 2009), though to a lesser extent than Proto-Germanic. Most relevantly for this discussion were the development of a "perfect tense" which was used for present perfects and simple pasts and whose forms were supplied by the old Indo-European perfect and aorist, and a "past (perfect) participle" built on the Proto-Indo-European deverbal suffix *-*to*-. Inflectional patterns in the language were organized along *theme vowels* expressed on stems following the root.

Classical Latin was a formal literary register of urban Roman Latin, a dialect of the language traditionally spoken by the Latins, an Italic ethnic group who were native to the plains of Latium (including the city of Rome) south of the Tiber River in modern Lazio, Italy. The Classical standard came into being during the 1st century BC, and seems to be based closely on the Vulgar (spoken) Latin of the urban elite of the time, though it contains some archaisms (Clackson and Horrocks, 2011). The Vulgar Latin of the Classical period eventually developed into what is known as Late Latin and Proto-Romance, the common ancestor of the modern Romance languages. As such, the Romance languages express shared innovations from the Late Latin period which are absent in Classical Latin. The Late Latin of the 4th century and later, not the Vulgar Latin on which Classical Latin is based, is the last common ancestor of the Romance family.

The adoption of Classical Latin as a literary standard obscures dialectal variation that existed in the language. Many of the features that we associate with Late Latin and Romance, such as the loss of final /s/ and final nasals, monophthongization, and certain lexical items, are absent from Classical Latin but were actually present in Old Latin as evidenced through epigraphy (Clackson and Horrocks, 2011). It is likely that at least some of these features were present in Latin dialects throughout the Classical period, but were suppressed in writing in favor of the Classical standard only to resurface in Romance. Unfortunately, Classical Latin was treated as standard for so long that Late Latin and the earliest Romance varieties

are nearly unattested and must be reconstructed. Since we can be certain that dialectal variation existed and change was constantly unfolding behind the scenes, it is worth asking whether Classical sources can be investigated for our purposes. Fortunately, the answer turns out to be yes, because Latin verbal inflection, including that of the past participles, was fairly stable from Old to Classical Latin (Weiss, 2009; Laurent, 1999). In the relevant respects, Classical Latin was very similar to the spoken language in a certain time and place. We should, however, be cautious of Classical texts written by authors from Late Antiquity who were native speakers of Late rather than Classical-like Vulgar Latin.

7.1.1. The Conjugations, Principal Parts, and Past Participles

As far back as Priscian in the 6th century, Classical grammarians thought of verbs in terms of four forms or *principal parts* from which all the inflected forms of a verb could be determined. They classified the verbs themselves into four (and a half) classes or *conjugations* according according to the forms of their first two principal parts. This characterization is actually quite useful, and the four principal parts are still provided in Latin grammars and dictionaries in use today (e.g., Allen and Greenough, 1903; Glare, 2012). While they are meant as a purely descriptive tool and so may or may not be "real" in a cognitive sense, they actually do reflect meaningful patterns of Latin morphology: the conjugations relate to which *theme vowels* attach to the roots (Embick, 2000) (or at least the present stem (Aronoff, 1994)), and the four principal parts line up with three recognizable stems. The first two principal parts, the present active indicative first person singular and present active infinitive are associated with present stem (henceforth *present*), the third, the perfect active indicative first person singular is associated with the perfect stem (*perfect*), and the final, either the past participle or supine[1] is associated with the past participle stem (*pptc*). The theme vowels and stems provided in Table 18 figure prominently in theoretical accounts of the verbal system. Only a few verbs, including 'carry,' 'want,' and 'be' are so irregular or suppletive as not to figure

[1]The supine inflects like a fourth declension noun, is only accusative or ablative, and is very nearly always identical to the past participle, so they are usually handled together. The only exception that I am familiar with is *lavō*, *-āre* 'wash' with past participle *lautus* and supine *lavātum*

into this system.

Conj.	ThV	1st Pres.	2nd Pres.	3rd Perf.	4th PPtc	Meaning
1st	*ā*	*amō*	*amāre*	*amāvī*	*amātus*	'love'
		sonō	*sonāre*	*sonuī*	*sonitus*	'sound'
2nd	*ē*	*moneō*	*monēre*	*monuī*	*monitus*	'warn'
		maneō	*manēre*	*mānsī*	*mānsus*	'remain'
		teneō	*tenēre*	*tenuī*	*tentus*	'hold'
3rd	*e*	*legō*	*legere*	*lēgī*	*lēctus*	'choose'
		pellō	*pellere*	*pepulī*	*pulsus*	'push'
		tangō	*tangere*	*tetigī*	*tāctus*	'touch'
		iungō	*iungere*	*iunxī*	*iūnctus*	'join'
3rd *-iō*	*i*	*capiō*	*capere*	*cēpī*	*captus*	'take'
		faciō	*facere*	*fēcī*	*factus*	'make'
4th	*ī*	*audiō*	*audīre*	*audīvī*	*audītus*	'hear'
		exciō	*excīre*	*excīvī*	*excitus*	'summon'
		hauriō	*haurīre*	*hausī*	*haustus*	'drain'
Irregular		*ferō*	*ferre*	*tulī*	*lātus*	'carry'
		volō	*velle*	*voluī*	–	'want'
		sum	*esse*	*fuī*	–	'be'

Table 18: Example "regular" and "irregular" Latin verbs by conjugations and principal part with corresponding theme vowels and stems.

The verbs featured in Table 18 were chosen to emphasize the fact that, in general, the stems are not reliably predictable from one another. Merely sharing one or two stem forms, an infinitive in *-ere* or perfect in *-uī*,[2] for example, does not guarantee that the other stems will share their form too. In practice though, some patterns are more consistent than others. For example, the vast majority of 1st conjugation verbs pattern like *amō*[3] in all three stems, but third conjugation verbs vary widely. Considering the perfect stem alone, the third conjugation contains verbs with bare affixation (*bibō* ∼ *bibī* 'drink'), a vowel mutation (*agō* ∼ *ēgī* 'do'), reduplication (*tangō* ∼ *tetigī*), an *-s-* suffix (*scribō* ∼ *scripsī* 'write'), and an

[2] I do not normally segment endings into suffixes since most of the analysis here is agnostic to particular decompositions. That is not a claim that morphological decompositions not exist, for example *-ere* is reasonably *-e-re*, a ThV and infinitive suffix, and *-uī* is *-u-ī*, a morpheme carrying a perfect or perfective feature followed by first person singular person/number marking.

[3] The first principal part is commonly used as a citation form in Latin. The first principal part and infinite ending will be provided when it is necessary to unambiguously determine a verb's conjugation, e.g., *capiō, -ere*.

-u- suffix (*moloō* ∼ *moluī* 'grind'). A few third conjugation verbs are variably attested with up to three distinct perfects (*parcō* ∼ *pepercī*, *parsī*, *parcuī* 'spare') (Clackson and Horrocks, 2011). Deponent verbs lack a perfect stem, instead using periphrastic constructions with past participle forms.

There is even more variety in possible pptc forms, which inflect like 1st/2nd declension adectives in *-a*, *-us*, *-um*. A few unambiguously suppletive forms such as *lātus* aside, a verb has several options. Most 1st and 4th, and a few 2nd conjugation verbs (e.g., *flēre* ∼ *flētus* 'weep') exhibit their theme vowel followed by *-t-* and case marking , but some show a short *i* instead of the theme vowel. This short *i* is present in all conjugatons but is mostcommon in the 2nd and 3rd conjugations. There are also a few 2nd conjugation verbs with roots of the form *Cav-* or *Cov-* with past participles in *-autus* or *-ōtus* < **-outus* (e.g., *faveō* ∼ *fautus* 'favor,' *moveō* ∼ *mōtus* 'move') and 3rd conjugation verbs with roots ending in *u*[4] with past participles in *-ūtus* (e.g., *solvō* ∼ *solūtus* 'loosen').

Most other 2nd and 3rd conjugation verbs lack a vowel between the root and *-t-*, often with substantial effect on the form of the root. Most importantly, if the root ends in a coronal obstruent (*d*, *t*, *s*), then the final segment of the root and pptc *-t-* are replaced with *-s-*, either with compensatory lengthening of the root vowel (*videō* ∼ *vīsus* < Proto-Italic **wid-t-os* 'see'), or gemination *-ss-* (*sedeō* ∼ sup. *sessum* 'sit'). Most *s*-pptcs are phonologically predictable, but there are exceptions, for example in (*spargō* ∼ *sparsus* 'scatter,' and *pellō* ∼ *pulsus*). Bare past participles trigger a variety of other changes to the root, many of which are phonologically predictable such as the devoicing of final obstruents (e.g., *scribō* ∼ *scriptus*) or mid-raising of a low root vowel. Low vowel raising is typically seen in prefixed forms and corresponds to low-to-high raising in the present stem, for example, compare unprefixed *faciō* ∼ *factus* 'do, make,' to *con-ficiō* ∼ *con-fectus* 'finish.' This and related processes seems to have been automatic in Old Latin but may not be in Classical language (Weiss, 2009).

[4]Conventionally written <v> when it represents a glide and <u> when it represents a vowel.

The illustrative table that Laurent (1999) provides extending Aronoff (1994) (reproduced in Table 19) counts up the most frequent pptc forms from a large sample of verbs for each conjugation in order to estimate the regularity of the pptcs. They conclude that the 1st conjugation is overwhelmingly regular because the large majority of its verbs share a pptc form, the 2nd and 4th conjugations are predominately regular as well, while the 3rd conjugation is not. While useful at a glance, there are two problems with these conclusions. First, a simple majority of items sharing a pattern is not a well-motivated metric for regularity, and second, the 2nd conjugation count is actually the sum of two different patterns. I will come back to these issues in Section 7.4.

Conjugation	**# Verbs**	**# Regular**	**% Regular**	**% Ending**
1st	360	345	96%	*-ātus*
2nd	120	90	75%	*-itus/-tus*
3rd	170	60	35%	*-itus*
4th	60	40	67%	*-ītus*

Table 19: Percent of verbs by class exhibiting the most common pptc pattern. Reproduced from Laurent (1999, Table 1-1) with endings column added.

In addition to the variety and unpredictability of inflected forms, many verbs simply do not have past participles, either for semantic reasons (e.g., the copula, statives and inchoatives) or as apparently one-off gaps (e.g., there is no past participle for *bibō*, *-ere*, *bibī* 'drink' or *feriō*, *-īre*, *feriī* 'strike'). And while past participles are typically passive in meaning, deponent past participles are active (e.g., *locūtus* 'having spoken'), as are a handful of other verbs' including *iūrātus* 'having sworn.'

7.1.2. The "t-Deverbals"

Latin verbs have a significant amount of derivational forms in addition to their inflections. Of these, a number of these called "*t*-deverbals" (Steriade, 2016) are, according to the Priscian analysis, built on the past participle stem. What makes them interesting is that they are semantically heterogenous, including agent, event, and result nouns, adverbs, and future active participles, yet they share the form of the perfective, largely passive, past participle

stem, even if it is in some way unpredictable suppletive. Table 20 summarizes the *t*-deverbals and highlights the shared pptc stem. Of these, the agent and event(1) nouns and future participle (fptc) are particularly common and were certainly productive. The others are less common, a point which will be addressed further in Section 7.4.[5]

Type	Ending	Pres.	PPtc	Meaning	*t*-Deverbal	Meaning
Adverb	***-tim***	*stō*	***status***	'stood'	***statim***	'immediately'
Agent	***-tor***	*docēō*	***doctus***	'taught'	***doctor, is***	'teacher'
Event(1)	***-tiō***	*agō*	***actus***	'done'	***actiō, -nis***	'action'
Event(2)	***-tus***	*sūmō*	***sumptus***	'spent'	***sumptus, -ūs***	'expenditure'
FPtc	***-tūrus***	*currō*	***cursus***	'run'	***cursūrus, -a, -um***	'about to run'
Result	***-tūra***	*scribō*	***scriptus***	'written'	***scriptūra, -ae***	'writing'

Table 20: Example *t*-deverbals with corresponding past participles, Priscian stems highlighted.

The form correspondence between the pptc and the various *t*-deverbals is quite robust, although there are a few exceptions, such as *mortuus* 'dead,' but *moritūrus* 'about to die,' and *sonitus* 'sounded,' but *sonāturus* 'about to sound' (Laurent, 1999, pp. 18-19), or *favitor* 'favorer' which exists alongside *fautor* and pptc *fautus* (Steriade, 2016, (23)). Verbs without past paricipples may nevertheless have *t*-deverbals, for example *calitūrus* 'about to be warm' from stative *caleō* 'be warm' and *futūrus* 'about to be' from the copula *sum*. There are also *t*-forms derived from nominals. With no past participles to correspond with, these non-verbal *t*-deverbals vary greatly in their forms. *-tim*-Adverbs, for example, are attested with the theme vowels *ā*, and *ī*, bare, and are built either on the root or the genitive singular. Interestingly, these theme vowels do not necessarily correspond to the nominals' stem vowels, as summarized in Table 21. The form of *vicissim* is particularly strange because it seems to contain the genitive singular *vicis* rather than the stem/root, which would give **victim* or maybe **vixim*.[6] The vowel in *virītim* could be explained if it were also built on the genitive singular (*virī*), but this still leaves *turrītus* (Embick, 2000) (gen. sg. *turris*) unaccounted for.

[5]There are also some verbal derivatives, iteratives, intensives, frequentatives, and desideratives, which share their form with the pptc (Laurent, 1999, §§2.4, 2.10).

[6] *Vicis* is also unusual for lacking a nominative form.

Nominal	StemV	Meaning	Type	*t*-Denom.	Meaning
honos, -ris	∅	'honor'	**Abstract**	***honestās***	'integrity'
liber, -a, -um	***a, o***	'free'		***liber∅tās***	'freedom'
fūr, -is	∅	'thief'	**Adverb**	***furtim***	'stealthily'
paulus, -a, -um	***a, o***	'small'		***paulātim***	'bit by bit'
–, *vicis*	∅	'time, change'		***vicissim***	'in turn'
vir, -ī	***o***	'man'		***virītim***	'per man'
senex, senis	***i***	'old'	**Agent**	***senātor***	'senator'
barba, -ae	***a***	'beard'	**PPtc-like**	***barbātus***	'bearded'
turris, -is	***i***	'tower'		***turrītim***	'towered'

Table 21: Example non-verbal *t*-derivatives with corresponding noun (in the nominative and genitive singular) and stem vowel. Unexpected theme vowels are highlighted in red.

7.2. History of the Latin System

The basic organization of Latin verbal paradigm remained more or less static throughout the Old and Classical eras (roughly the 3rd century BC through the 2nd Century AD) following considerable changes from Core Indo-European (Weiss, 2009). The changes that have occurred since Latin, though perhaps less dramatic, are sometimes surprising and provide meaningful insights into the synchronic grammars of Classical and Late Latin (Laurent, 1999).

7.2.1. Indo-European to Latin

The past participle is descended from a Proto-Indo-European deverbal adjective ending in *-*to-* built on thematic stems with deverbal adjective cognates in Ancient Greek and Sanskrit (Weiss, 2009, ch. 28 §2) which was then incorporated into the verbal paradigm as a participle. Subsequent sound changes account for many of the differences between the past participle and other stems, such as devoicing before /t/ (e.g., *actus* vs. *agō*) and simplification of coronal obstruent clusters to /s/ (e.g., **dt* > *s visus* vs. *videō*). The *t*-deverbals have similar etymologies, all deriving from Proto-Indo-European suffixes beginning with **t* which attached to thematic or athematic stems (Weiss, 2009, ch. 29). Among these, *-tiō* may have been derived from an instrumental of verbal nouns, and *-tūra* may actually be built on **-to-*

itself. In a sense, the pptc was itself once a *t*-deverbal.

Many Latin verbs are themselves some sort of derived form: some 1st conjugation verbs are from *-*ie/o*- denominals, factitives from adjectives which are still productive in Classical Latin (e.g., *novāre* 'renew' ← *novus* 'new (adj.)'), frequentatives and intensives, also synchronically, and more (Weiss, 2009, ch. 36 §2). 2nd conjugation verbs derive from *-*eh*$_1$-(*ie/o*)- verbs (e.g., *plēre* as well as causative in *-*éie*- such that 2nd conjugation pptcs in -*itus* < *-*etos* (Ringe, 2017; Weiss, 2009). The derivations of the 3rd and 4th conjugations are similarly complex but are excluded from the present discussion. The rest of this chapter will not address changes that occurred before Classical Latin.

All of this means that most Latin pptcs are actually regular from a diachronic perspective even if they are synchronically unpredictable, since they were built on common endings and then subjected to regular sound changes. And *to a first approximation*, the *t*-deverbals and past participles are in form correspondence because their endings both began with *t and were subject to those same regular sound changes.

7.2.2. Late Latin to Romance

The Latin verbal system underwent some substantial changes in the transition to Romance, notably the loss of synthetic passives and the innovation of periphrastic perfects, futures, and conditionals. Additionally, perhaps facilitated by the collapse of vowel length and the merger of short *i* with *e* in most regions, there was also significant *metaplasm*, with many 2nd, 3rd, and 4th conjugation verbs shifting conjugations, often to the 4th (Laurent, 1999, §2.6). There was also an uptick in the coining of new derived intensive, frequentative, and iterative verbs which replaced older forms. These were consistently regular 1st conjugation verbs (e.g., *cantō*, -*āre*, -*āvī*, *ātus* 'sing' < intensive of *canō*, -*ere*, *cecinī*, *cantus* 'sing, recite') (Laurent, 1999, §§2.4, 2.10).

Most crucial for this study, there were significant changes to the forms of the past participle. Beginning in Late Latin, three past participle forms, *-*atu*, *-*itu*, and *-*utu* < -*ātus*, -*ītus*,

-*ūtus* began to spread at the expense of bare -*tus* and -*itus* despite the latters' frequency outside of the 1st conjugation (Laurent, 1999, §3).

The expansion of *-*atu* and *-*itu* can easily be accounted for in even a pre-theoretical account of analogical leveling. -*ātus* was already dominant among 1st conjugation verbs, so of course it would tend to spread to the remaining 1st conjugation verbs, and -*ītus* was the most common 4th conjugation form, so it is not too surprising that it would have spread to former 2nd and 3rd conjugation verbs as well in conjunction with the metaplasm that occurred.

The rise of *-*utu* is more perplexing. Why -*ūtus*, which only existed for about a dozen 3rd conjugation verbs, should have undergone analogical extension throughout the former 2nd and 3rd conjugations is unclear, particularly given the prevalence of bare -*tus* and -*itus* in those conjugations. Nevertheless, that is what happened. The map in Figure 44 illustrates the geographical extent of reflexes of -*ūtus* in modern Romance contrasted against reflexes of -*itus*. Strikingly, reflexes of -*ūtus* are present across Romance, both "Eastern" and "Western" except for Sardinian (red on the map). In most areas, the reflex is apparently productive for some class of verbs (dark blue), and in most languages, it is apparently the default for the former 2nd and 3rd conjugations (e.g., Italian -*ere* verbs: *vendere* ∼ *vend**uto***, French -*re*: *vendre* ∼ *vend**u***, Catalan -*re*/-*er*: *vendre* ∼ *ven**ut*** 'sell', contra Latin *vēn**itus*** 'on sale'). In others, it is apparently productive in smaller classes, hence Romanian *vând**ut***. There are no verbs with -*ūtus* past participles in modern Iberian, though they are attested both in Old Spanish and Old Portuguese (yellow; e.g., Old Portuguese *ven**udo***). Remmants of Iberian -*udo* can still be found outside the verbal system, for example Spanish *men**udo*** 'tiny' < Latin *min**ūtus*** 'small, diminished,' pan-Iberian *ag**udo*** 'sharp' < Latin *ac**ūtus*** 'sharpened,' and Portuguese *vend**udo*** 'something that is sold.'[7] All Iberian -*udo* past participles were eventually replaced with -*ido* (Laurent, 1999, §4.7). Finally, past participle reflexives of -*ūtus* are present but apparently unproductive in the Surselvan and Engadin dialects of

[7] I found the form *vendudo* with this definition in a modern Portuguese dictionary but no instances of it in use.

Romansch spoken in far-southern Switzerland (light blue) (Laurent, 1999, §4.3).

The distribution of reflexes of *-itus* in modern Romance is far more restricted, present in only Sardinian and Apulian in southern Italy (black stars). There is evidence that *-itus* was previously productive in other parts of southern Italy as well, for example, Lucanian *bippeto* and Neapolitan *vippeto* < Late Latin **bibitu* rather than Standard Italian *bevuto* (Laurent, 1999, §3.6). Its expression in Sardinian may be explained by the merger of /i/ and /iː/, which leveled the vowel quality in *-itus* and *-ītus*. Elsewhere in Romance, /i/ merged /eː/, and in many cases, /e/ (Loporcaro, 2015, §2.4).

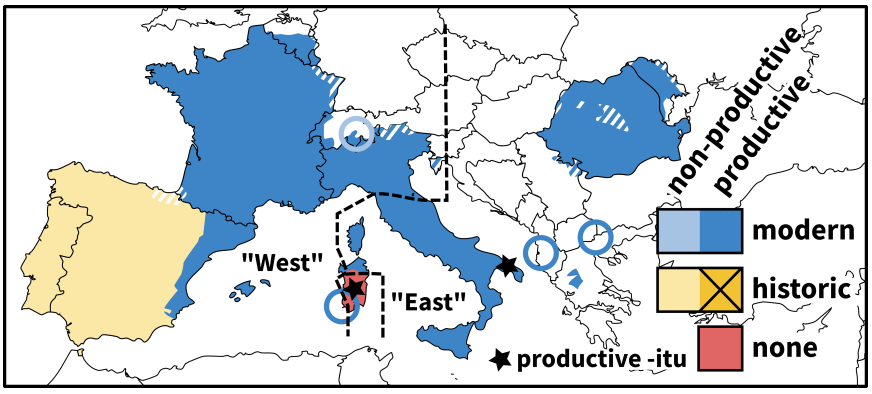


Figure 44: The distribution of past participle reflexes of *-ūtus* and *-itus* in modern Romance. Reflexes of *-ūtus* productively form some past participles productively in most regions (dark blue), form some apparently unproductively in dialects of Romansch (light blue), historically formed past participles in Iberia (light yellow), and never have in Sardinian (red). Only Sardinian and Apulian productively form past participles in reflexes of *-itus* (black stars) (compiled from Laurent, 1999, §3).

What reflexes of bare *-tus* remain have been relegated to irregular past participles in modern Romance or have been recast as adjectives. The irregulars are overwhelmingly high frequency items (e.g., Italian *fatto*, French *fait*, Spanish *hecho* < *factus* 'done, made' and Italian *detto*, French *dit*, Spanish *dicho* < *dictus* 'said') (Laurent, 1999, §6.6). The relationship between high frequency items and irregularity (or resistance to analogical leveling) is well-known (Bybee, 1985; Baayen, 1993, e.g.,) and predicted by any reasonable model of analogy. Those items that have been recast as adjectives result in several interesting doublets with regular past participles, for example regular Spanish *es despertado* 'he is being awoken' vs. *está*

despierto 'is awake' or regular *teñido* 'dyed' vs. *tinto* 'dark red like wine' < *tīnctus* 'dyed'.

The Romance reflexes of the *t*-deverbals largely followed the same path as the past participles. For the most part, they remained in form correspondence. What discrepancies did develop surround irregular past participles. Some inherited *t*-deverbals correspond as they did in Classical Latin (Spanish *escrito* 'written' ∼ *escritor* 'writer' < *scriptus* 'written' ∼ *scriptor* 'writer'), while others have been regularized despite retaining irregular past participles (Spanish *hecho* 'made, done' < *factus* but regular *hacedor* cf. infinitive *hacer*). However, not the existence of doublets. For example, the expected regular agent noun *escribidor* does exist in Spanish alongside *escritor*, but rather than 'writer,' it means 'scriptwriter.' According to Steriade (2016), only inherited irregulars can have irregular corresponding *t*-deverbals in Romance.

7.3. A Summary of Theoretical Accounts

The Latin past participles and their form correspondence with the *t*-deverbals have generated a significant amount of theoretical discussion. Not only is the form of the pptc stem convoluted, typically unpredictable from the other stems, but so is its semantics. If a past, perfective, usually passive, participle shares a stem with an active, imperfective, agent noun of all things, then what do its components mean? How many *-t-*[8] morphemes are there, and what semantics do they contribute? A related question, why does the form correspondence exist, is less interesting in my opinion, since it can be pinned primarily on diachrony. One does not necessarily need a synchronically active process to maintain the correspondence for most verbs.

7.3.1. Forms of the Past Participle

Setting the *t*-deverbals aside for now, we begin with the form of the past participle. In the most extreme analysis, one could propose that the pptcs need to be memorized in some way because they are so unpredictable over all. To the best of my knowledge, nobody has

[8] *-t-* should be read as *-t-/-s-*.

proposed that every Latin stem form be listed explicitly, but Lieber (1980) does propose a system without productivity. In that analysis, every root is associated with *morpholexical rules* that map the root onto its stem forms. There is a finite set of morpholexical rules, but none is productive, so each root needs to be associated with the correct rule on an individual basis.

A similar system could be implemented in any other theoretical framework that allows some kind of association between patterns and roots, however, other authors are not as pessimistic about the possibility for generalization. (Aronoff, 1994) disagrees with Lieber's 1980 contention that every verb's pptc pattern be listed on the basis of Table 19, which shows that some of the conjugations have a typical pptc pattern associated with them. If a verb can be placed in the correct conjugation, its pptc need not necessarily be listed. Adopting a lexeme-based treatment, he argues that the form of the pptc stem (referred to as the *third stem* to differentiate it from the part participle itself) for some verbs is based on the lexical representation of the root, while in others it is based on the form of the present or perfect stem via *realization rules*.

In Embick's 2000 DM analysis of the Latin perfect and pptc, stem forms are covered by underspecification of vocabulary items and *readjustment rules*, which can be defined to apply only in the presence of certain morphophonological conditions (i.e., productive given some condition), or in the presence of specific roots (i.e., listed, unproductive). For Embick (2000), the pptc *-t-* is the default realization of Asp (the aspect head) with pptc *-s-* a more specific vocabulary item that applies to a list of roots. Further readjustments can then account for the exact form of the past participle (27), though the question of what set of readjustment rules would yield the correct stem forms is left to the reader. Echoing Aronoff (1994), readjustments need not apply to listed roots if there are other generalizations that can be made.

(27) Realization of Asp (not raised to T) (Embick, 2000, (44), present participle vocabulary item given for exposition)

-nt- ⟷ [pres]

-s- ⟷ [ ] / ____ (List)

-nt- ⟷ [ ]

Steriade (2016) presents a system perhaps most opposite from Lieber (1980) in that most past participle forms are taken to be predictable. In the 1st and 4th conjugations, regular past participles are simply formed by appending the theme vowel and the pptc *-t-* morpheme. In the 2nd and 3rd conjugations, a regular pptc is either *-tus* or *-itus*, and the selection is made to achieve rhythmic correspondence with the form of the perfect. The intuition derives from the observation that the form of the pptc tends to contain the same number of syllables as the perfect. This can be accomplished with an OT analysis as in (28)

(28) Deriving the perfect participles corresponding to monosyllabic and disyllabic verbal perfects (Steriade, 2016, (16))

a. Monosyllabic (*scribō*, *-ere*, *scripsī*, *scriptus*)

Base [*scrip-s*]- Suffix: *-t*, *-it*	Dep V (perfect)
☞ a. [*scrip-t*]-*us*	
b. [*scrib-it*]-*us*	∗!

b. Disyllabic (*molō*, *-ere*, *moluī*, *molitus*)

Base [*mol-u*]- Suffix: *-t*, *-it*	Max V (perfect)
☞ a. [*mol-it*]-*us*	
b. [*mol-t*]-*us*	∗!

Max/Dep V (perfect): If two verb forms have the same lexical head and the same aspectual value, then each nucleus in the stem of one has a correspondent nucleus in the stem of the other.

Of the 325 2nd and 3rd conjugation verbs which Steriade (2016) considers, about 254 obey rhythmic correspondence, 37 are excluded as "archaisms" or due to "paradigmatic factors," and 34 of those which are analyzed do not obey, for a success rate of 78.2%. Violations are not

necessarily an issue for a violable constraint analysis, since they can result from competition among other unspecified constraints, though it would be helpful to have a sense of whether $\sim 22\%$ is a large or small number of exceptions. It should be noted, that $(34 + 37)/325$ exceptions does not pass a cursory application of the Tolerance Principle, which calls the learnability of such a system into question.

7.3.2. Forms of the t-Deverbals

In the Priscian account, the *t*-deverbals are quite literally built on the past participle stem: the case marking is removed from a given past participle, then an additional *t*-deverbal suffix is added (*-or-* for agents, *-ion-* for events, etc.) and case marking. Theoretical accounts in which the *t*-deverbals are literally based on the past participle are perhaps closest in spirit to this (Matthews and Matthews, 1972; Mel'čuk, 1982). In Mel'čuk, there is one *-t-*, which carries the meaning of the past participle. Any additional *t*-deverbal material after the *-t-* is *replacive*. It subtracts the past participle meaning and contributes its own. Mel'čuk's *t*-deverbals are literally built on the past participle.

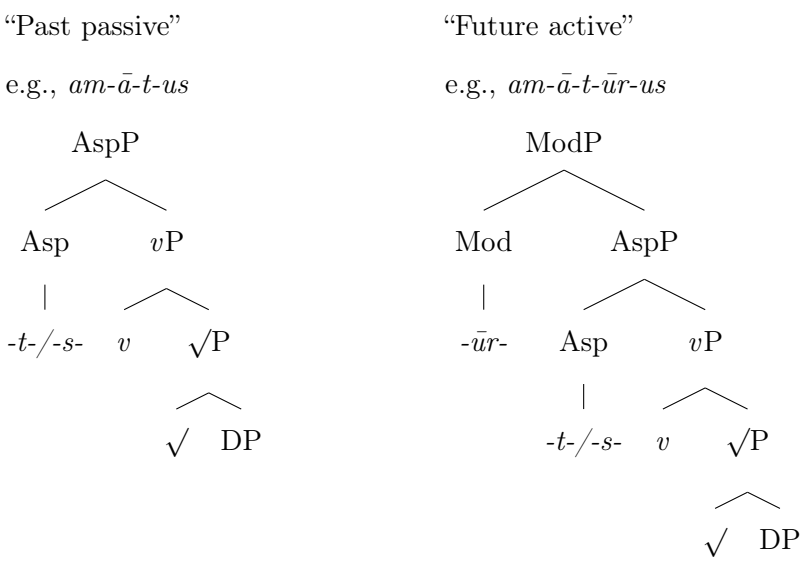
Aronoff argues instead that both the past participle and the *t*-deverbals are formed according to the third stem rather than one from the other. He notes that there are *t*-deverbals for verbs without past participles (discussed briefly in Section 7.1), so at least for those verbs, the *t*-deverbals cannot be based on the past participle. Also, since the supine is virtually always identical to the past participle, it is unclear why the past participle would be the base as opposed to the supine. Regarding the semantics of the pptc and *t*-deverbals, stems in this account are just phonological forms of a lexeme and are not associated with any particular inflectional or derivational meaning. One can conceive of a mapping between stems on one hand and inflectional and derivational categories on the other as visualized in Table 22. Under this analysis, the *-t-* does not actually carry a past participle meaning per se, so there is nothing to prevent both a past participle and agent from sharing the same meaning. It is a coincidence with diachronic origins.

Category	Present	Perfect	PPtc
Present	***ser****ō* 'I sow'		
Perfect		*sēvī* 'I sowed'	
PPtc			***sat****us* 'sown'
Agent			***sat****or* 'sower'
Event			***sat****iō* 'act of sowing'
Capacity	***ser****ibilis* 'sowable'		

Table 22: Stems (columns) and example categories (rows) for the verb *serō*, *-ere* 'sow, plant,' capacity adjective 'sowable,' agent noun 'sower,' and event noun 'sowing' illustrating stem-to-category mapping.

Embick addresses the past participle and the *t*-deverbals in his 2000 work on the morphosyntactic representation of the Latin perfect. He concurs with Aronoff, albeit in a very different framework, that the *-t-* of the pptc does not carry any meaning, which is why the *t*-deverbals are able to share that stem despite their unique semantics. In his account, *-t-* is just the default realization of the Aspect head (27), so the past participle and various *t*-deverbals need not agree in aspect or voice. The reason they share a form is because they share part of their structure (29), spell-out rules, and readjustments. This treatment can account for the presence of periphrastic passive and deponent perfects as well: both perfects share the structure in (29), however, the $\sqrt{}$-*v*-Asp complex which raises to T in synthetic perfects is blocked from doing so by the presence of a [pass] feature in passives and deponents. The form of the perfect stem is realized by spell-out rules for [perf] in the presence of T.

(29) Relevant structures and exponents for a pptc and fptc (Embick, 2000, (45)-(46))



Finally the Steriade (2016) *similarity-based syncretism* account for the form correspondence differs from other treatments in that the *-t-* of the past participle stem does have meaning as in (Mel'čuk, 1982) on the basis of pairs of capacity adjectives such as *duc-ibilis* 'can lead' or 'can be led' and *duct-ibilis* which only permits the passive meaning. Second, the underlying form of all *t*-deverbals consists of a linking *-i-*, *-t-*, and the appropriate suffix. Evidence for this comes from the observation that when a past participle and *t*-deverbals do not agree in form, the *t*-deverbal contains *-i-t-*[9], and *t*-derivatives from non-verbal sources also always contain *-i-t-*.[10] The underlying *-i-t-* *t*-deverbal is the brought into form correspondence with the help of a constraint $\textsc{Corr}_{\textsc{sim}}$ defined in (30) along with an examples.

(30) Similarity-based Syncretism (Steriade, 2016, (31), (32))

[9]This is common, but turns out to be false in general. Consider the forms of *sonitus* and *sonatūrus* listed in Section 7.1 and the presence of *t*-deverbals for which there is no corresponding past participle such as *iactūrus* 'about to lie' and *recasūrus* 'about to fall back.

[10]"always" (Steriade, 2016, §6.2.8.2). This turns out to be false as well. *ā*, and *ī* along with ∅ are attested in non-verbal *t*-derivatives (Section 7.1 (cf. Embick, 2000; Weiss, 2009)). Furthermore, several of the proposed examples cited in (Steriade, 2016, §6.2.8) turn out to be problematic. Forms that are only attested by late authors (*fīcitor* cf. *fīca* Nonius Marcelus fl. late 3rd c., *Imporcitor* contra *imporcātus* Servius Honoratus fl. late 4th c., *bibitor* (no pptc) Apollinarius Sidonius ob. 488, and *infenditor* contra *-fensus* in a 16th c. glossary attributed to 'Vulcanius') cannot be taken as evidence for underlying Classical *-i-t-*, because their language may have already lost vowel length distinctions and generalized **-itu* < *ītus*, and their Classical judgments were non-native. Other forms are likely not synchronically derived, for example, *iānitor* ← *iānus* (Weiss, 2009) (probably not *iānua* as claimed), *Imporcitor*, a proper noun 'God of Furrows,' and *adversitor* (actually, more often *advorsitor*) present questions as well.

a. CORR$_{SIM}$: For any pair of surface MinStems S_1, S_2, if (S_1, S_2 are lexically identical and (b) S_1, S_2 end in homorganic, [α sonorant] segments, then S_1, S_2 stand in correspondence.

b. Derivational suffixes attach to the root or the *infectum* (i.e., present) stem. A buffer *-i-* separates any stem-final C from a suffix-initial C.

Root *caed-*; PPtc: [*caes*]^{i}*-us*	CORR$_{SIM}$	MAX/DEPV, IDENT OO	(30B)
a. [[*caed-i-t*]^{i}*-or*]	∗!		
b. [[*caed-i-t*]^{j}*-or*]		∗! (s-d, i-∅, t-∅)	
☞ c. [[*caes*]^{i}*-or*]			∗

While instantiated under very different frameworks, each of these accounts provides a means for generating all and only the forms observed in Latin as a balance between general productive mechanisms and listing. Methodologically, they drawn the line between productivity and listing according to the authors' intuitions. I turn toward the problem of patterns and exceptions in the following section. Rather than relying on researchers' intuition and parsimony, I show that a principled analysis of the data, grounded in a model of acquisition, yields an unintuitive yet predictive division between regular and irregular forms.

7.4. Productivity of the Past Participles and *t*-Deverbals

An externally motivated treatment of listing and generalizations in the past participle and related forms stands both to improve existing theoretical accounts and to further disambiguate between their predictions. Since the task of delimiting generalizations which theoreticians face is analogous to the problem of generalization learning in child language acquisition, I take a learners' perspective approach to the issue. The first course of action is to adopt the Tolerance Principle's productivity-based definition of regularity and to define the system more rigorously in these terms. The goal here is to determine exactly which, if any, past participle stems could be predicted from the present or perfect stems and to relate this to the *t*-deverbals. In principle, a wide range of generalizations are possible, and a number of patterns identified in the literature (e.g., Weiss (2009, ch. 39)) are likely to be available to

learners. I calculate Tolerance Principle thresholds across the Classical lexicon for a wide variety of plausible hypothesis generalizations to determine which would be tenable for Latin learners.

To actually perform these calculations for learners with different vocabulary sizes, we need a frequency-sorted list of Latin verb lemmas and their principal parts. A dictionary lexicon does not work for these purposes because it over-weights rare vocabulary and does not provide frequency information (Baayen and Renouf, 1996). To get the largest sample possible, I scraped all Old and Classical texts from the Perseus Catalogue (Smith et al., 2000) online edition (which contains more text than can be downloaded), lemmatized, and POS-tagged them with scripts extended from the CLTK Python library,[11] sorted them by token frequency, and grouped them by root. Roots were sorted with prepositional prefixes removed (e.g., *faciō*: *faciō*, *conficiō*, *perficiō*, *officiō*, etc.), since these very rarely differ in their pptc patterns, modulo short vowel raising. Perseus does not mark vowel length, so length was restored and perfect and pptc stems acquired by merging the lemma list with the principal parts provided with vowel length in Latin Wiktionary.

What resulted is a list of 1,292 unique verb lemmas and their principal parts derived from about 3.5 million tokens of Latin text composed between the 3rd century BC and 2nd century AD by speakers of Latin varieties which should have been very similar to Classical Latin. This data set is analyzed along with the others in Chapter 3, where it is shown to meet the criteria for a reasonable CDS substitute. The process by which this frequency list was extracted renders the project feasible, but it comes with a few caveats. First, all automatic POS-tagging and lemmatization has the potential to introduce errors, though these do not appear significant in most cases on inspection of the data. The main problem comes from homophones such as the form *volō* 'want or fly,' which cannot easily be assigned to the correct lemma. The problem is exacerbated on Perseus data because it lacks vowel length annotation, which collapsed some forms such as the perfects *cecīdī* and *cecidī* which

[11] http://cltk.org/

correspond to *caedō* 'strike, kill' and *cadō* 'fall' respectively. Even so, homophony like this turns out to be uncommon in practice, so when these arose, the combined frequency count was applied to both verbs, and since the calculations here do not depend on the presence or absence of any particular verb, this decision does not affect the final calculations. Second, Wiktionary, which is necessary for restoring vowel length and collecting principal parts, is not perfect. About one hundred Wiktionary-provided principal parts were compared with entries in the Oxford Latin Dictionary (Glare, 2012) in order to confirm their accuracy. There were only a handful of discrepancies, all of which would have looked suspicious to any student of Latin. These were corrected, and all other suspicious entries were double checked. All in all, Latin Wiktionary is surprisingly accurate. Last, the process of removing derivational prepositional prefixes offers some implementational choices. As a matter of course, I removed every derivational prefix I could find with only a few exceptions that I did not find to be transparent, for example, *prōmō* 'produce, bring to light' was not decomposed into *prō+emō* (root: 'buy, acquire') and *vēndō* 'sell' was not decomposed into *vēn-+dō* (root: 'give').

We can now compare how the Perseus-derived verb list compares to Laurent's (Table 19) in terms of the sizes of the conjugations and the most common past participle forms. Table 23 summarizes the top 1,000 most frequent verb lemmas.[12] The 1st and 3rd conjugations are the largest, with the 1st conjugation accounting for over half of all verbs with past participles. They are also the most and least homogeneous respectively. *-itus* and *-tus* are the most common past participle forms in the 2nd and third conjugations and the second most common in the 1st and 4th. Most remaining verbs undergo some unpredictable stem mutation. The percents provided here are slightly lower than but in the same ballpark as those provided in Table 19.

[12]Verbs without recorded past participles, mostly inchoatives in *-escere*, were excluded from this and all subsequent calculations, which is why second column does not sum to 1,000.

Conj.	# Verbs	Most	freq	% Most	Next	Freq	% Total
1st	541	*-ātus*	528	97.6%	*-itus*	6	98.7%
2nd	65	*-itus*	25	38.5%	*-tus*	17	64.6%
3rd	215	*-tus*	80	37.2%	*-itus*	19	46.0%
4th	55	*-ītus*	34	61.8%	*-tus*	13	87.3%

Table 23: Percent of verbs exhibiting the two most common pptc forms in each class.

7.4.1. Productivity of the Past Participles

The morphological systems of the world leverage a wide variety of generalizations, and a child cannot know *a priori* which of these generalizations their language will actually use. This motivates a "see what sticks" approach to hypothesis evaluation constrained only by what are possible human language generalizations. To model this kind of learning procedure, I consider every plausible generalization found in the literature on Classical Latin (Aronoff, 1994; Weiss, 2009; Steriade, 2016) in addition to several that seemed plausible to me in order to give productivity its best shot. These generalizations are divided into four groups: traditional generalizations over present stem theme vowels, finer-grained generalizations over present stem forms, generalizations over perfect stem forms, and generalizations which require both the form of the present and perfect stem.

Generalizations are evaluated with the Tolerance Principle on the top n = 100, 500, and 1,000 most frequent verbs in the data set in order to model young learners, older, perhaps school age children, and young adults. At this stage, they can be formulated in a theory-independent fashion, for example, a generalization depending on the presence of the theme vowel *-ā-* in a structural representation can also be cast in terms of surface phonotactics by requiring the rightmost vowel in the stem to be /aː/. Phonologically predictable stem changes such as the devoicing of *b* in *scriptus* were not treated as exceptions. *s*-Pptcs were considered automatic if directly following a root-final *t*, *d*, or *s*, either resulting in a short vowel and geminate *s* as in *missus* or a short vowel and single *s* as in *ēsus*. Old Latin short vowel raising to *i* in open and *e* in closed syllables as in *-cipō* ∼ *-ceptus* (cf. *-capō* ∼ *-captus*)

was also treated as non-exceptional.

Table 24 presents the results of the theme vowel productivity calculations. It immediately stands out that only the 1st and 3rd-*iō* conjugations have productive pptc derivations at $n = 500$ and above. While the the majority of 4th conjugation verbs do indeed form their past participles in -*ītus* as Laurent observes, it is not quite enough to render that derivation productive according to the Tolerance Principle. Importantly, even though -*tus* and -*itus* are quite common among all non-1st conjugation verbs, their distribution is such that neither is productive according to these theme vowel generalizations. This stands in sharp contrast to the high predictability of Proto-Germanic strong verb classes (Chapter 6 Section 6.2) and calls into question the cognitive validity of the grouping. Following Aronoff (1994), one might conclude that the traditional classification is only valid for the present stems since it is not reliably maintained elsewhere.

Theme Vowel	**PPtc**	**Example**	**At 100?**	**At 500?**	**At 1000?**
(1st) *ā*	-*ātus*	*vocāre* ∼ *vocātus*	**yes**	**yes**	**yes**
(2nd) *ē*	-*itus*	*habēre* ∼ *habitus*	no	no	no
(2nd) *ē*	-*tus*	*docēre* ∼ *doctus*	no	no	no
(3rd non-*iō*) *e*	-*itus*	*reddere* ∼ *redditus*	no	no	no
(3rd non-*iō*) *e*	-*tus*	*scribere* ∼ *scriptus*	no	no	no
(3rd -*iō*) *i*	-*tus*	*capiō* ∼ *captus*	**yes**	**yes**	**yes**
(3rd) *i* or *e*	-*itus*	”	no	no	no
(3rd) *i* or *e*	-*tus*	”	no	no	no
(4th) *ī*	-*ītus*	*audīre* ∼ *audītus*	**yes**	marginal	no
(4th) *ī*	-*tus*	*venīre* ∼ *ventus*	**yes**	no	no

Table 24: Tolerability of past participle patterns by present stem theme vowel (corresponding to the traditional conjugations). “Marginal” cells indicate a calculation that was within 1 of the tolerance threshold in either direction.

Table 25 lists additional present stem-form productivity calculations. Few of these narrower patterns manage to achieve productivity either, and only the pattern followed by *faveō* and *moveō* remains clearly tolerable for large vocabulary sizes. Two of the hypothesized generalizations are right at the cusp of productivity in our data set: the pattern of *currō* ∼ *cursus* and that of *solvere* ∼ *solūtus*. In either case, if a certain child’s lexicon contained one

fewer exception, it would have been rendered productive. The latter is of particular interest because *-*utu* spread analogically in Late Latin.

Present	**PPtc**	**Example**	**At 100?**	**At 500?**	**At 1000?**
-*veō*	-*autus*/*ōtus*	*favēo* ∼ *fautus*	-	**yes**	**yes**
-[velar]*eō*	-*tus*	*doceō* ∼ *doctus*	-	no	no
-[not velar]*eō*	-*itus*	*debeō* ∼ *debitus*	marginal	no	no
-[not velar]*eō*	-*tus*	*teneō* ∼ *tentus*	no	no	no
-*vere*	-*ūtus*	*solvere* ∼ *solūtus*	**yes**	marginal	marginal
-{*ll*,*rr*}*ere*	-{*l*,*r*}*sus*	*currō* ∼ *cursus*	-	marginal	no
other 3rd	-*itus*	*reddere* ∼ *redditus*	no	no	no
other 3rd	-*tus*	*scribere* ∼ *scriptus*	no	no	no

Table 25: Tolerability of present form stem to past participle relationships.

If the learner hypothesized generalizations building the pptc on the perfect as in Table 26, then there would exist productive derivations for both -*ātus* and -*ītus* past participles along with the handful of -*ētus* that exist. This is because verbs with theme vowel-containing pptcs almost always express the theme vowel in the perfect as well. In the 1st and 4th conjugations, the verbs with exceptional pptcs tend to have -*uī* perfects, so -*āvī* and -*īvī* perfects serves as a more reliable evidence for the form of the pptcs than the presents. The only other pattern that reaches productivity is the -*Cs*--perfect to -*tus*. These are a major source for remnant exceptional past participles in modern Romance, including Spanish *escrito* 'written' < *scriptus*, which retained an *s*-perfect *escrisso* < *scrips*- in Old Spanish (Laurent, 1999, p. 301). Perhaps most surprisingly, there are no productive past participle patterns for -*uī* perfects, not even -*itus*, which is suggested as reasonably reliable in Aronoff (1994) and is a crux of the (Steriade, 2016) analysis.

In addition is a single generalization that improves in tolerability if both the form of the present and perfect stems are considered: third conjugation present in -*vō* + perfect in -*uī* to pptc in -*ūtus* (e.g., *volvō* 'roll' ∼ *voluī* ∼ *volūtus*) become firmly productive.

Perfect	PPtc	Example	At 100?	At 500?	At 1000?
-āv-	*ātus*	*amāvī* ∼ *amātus*	**yes**	**yes**	**yes**
-īv-	*ītus*	*dormīvī* ∼ *dormītus*	**yes**	**yes**	**yes**
-ēv-	*ētus*	*flēvī* ∼ *flētus*	**yes**	**yes**	marginal
-u-	*itus*	*valuī* ∼ *valitus*	no	no	no
-u-	*tus*	*tenuī* ∼ *tentus*	no	no	no
-[velar]*u-*	*tus*	*līquī* ∼ *līctus*	-	no	no
-[not velar]*u-*	*itus*	*dēbuī* ∼ *dēbitus*	no	no	no
-[not velar]*u-*	*tus*	*peruī* ∼ *pertus*	no	no	no
-s-	*tus*	*scripsī* ∼ *scriptus*	no	no	no
-Cs-	*tus*	*iūnxī* ∼ *iūnctus*	**yes**	**yes**	**yes**
bare	*tus*	*lēgī* ∼ *lēctus*	no	no	no

Table 26: Tolerability of perfect to past participle relationships.

Maybe the least intuitive conclusion that can be drawn here is that neither *-tus* nor *-itus* achieve broad productivity in Classical Latin despite their high frequencies and wide distributions. The former is only regular for 3rd-*iō* and *s*-perfect verbs, while the latter is never regular. Conceptually, the two endings compete with each other in the 2nd and 3rd conjugations and pushed each other above the tolerance threshold. Similarly, there is no broadly productive past participle that corresponds to *-uī* perfects despite its frequency in the language. The only subset of *-uī*-perfect verbs with a productive past participle are those with *-vere* presents and past participles in *-ūtus*. Overall, only *-ātus* and *-ītus* achieve productivity among large classes of Classical Latin verbs, and these are the only two endings that remain productive in every branch of modern Romance. In summary, there are more generalizations to be made than Lieber (1980) argued for, but still far fewer than could exist.

The fact that *-itus* and *-tus* are never productive except for narrow sets means that nearly all the past participles of 2nd and 3rd conjugation verbs must be memorized, which is consistent with the attested distribution of apparent paradigmatic gaps and a "no default" account (Yang, 2016; Gorman and Yang, 2019). In a no default account, gaps can potentially appear when there is no default pattern for speakers to fall back on when they have not memorized a specific form.

7.4.2. Productivity of the t-Deverbals

With a productivity analysis of the past participles complete, we can now begin to more fully explain the *t*-deverbal form correspondence. It is worth emphasizing that the correspondence is itself largely, but not exclusively, a diachronic accident: The past participle and *t*-deverbals are etymologically related, all containing the PIE denominalizer *-*to*-, and therefore were subject to the same sound changes. However, not all forms can be explained this way, such as clear instances of suppletion *offerō* ∼ *oblātus* ∼ *oblātiō*.

Acquisition is capable of bringing the rest into correspondence. Laid out conceptually, the forms of most *t*-deverbals need to be inferred by the learner because, due to the Poverty of the Stimulus, most are unlikely to be attested in the input, and most of the *t*-deverbals that are attested have a corresponding attested pptc. With these as evidence, the hypothesis "make the *t*-deverbals be like the pptc" is upheld more reliably than other options, so new forms are produced in correspondence with the pptc.

There are other possibilities which can be immediately dismissed. First, it could be that *t*-deverbals default to some form like Steriade's *-i-t-*, then exceptions are memorized. This however is a non-starter since global defaults for the *t*-deverbals are untenable for the same reason that they are for the past participles. Second, it could be the case that the form of the past participle is actually influenced by the *t*-deverbal rather than vice-versa. While this is technically possible, it could not have been the norm. A learner was much far more likely to hear a pptc and have to infer the corresponding *t*-deverbal than the opposite situation.

This is worked out quantitatively in Table 27 with forms collected from Perseus. This time, every inflected past participle and *t*-deverbal were collected[13] The first question is one of raw type frequency. How many *t*-deverbals are there relative to past participles? To determine this, the thousandth most frequent past participle was found, and then all *t*-deverbals with at least that frequency (= 35) were extracted. This resulted in the first two columns of the

[13]Result nouns were grouped with future participles because they have identical forms, and the relatively rare *-tus* event nouns were grouped with the past participles and supines.

table. These account for over three quarters of the forms. *-tiō̄* event nouns are the most common after the past participle, but they only account for about 13% of total types. The next question is how often a learner would have to infer the form of one or the other. To do this, the previous list was filtered to find the number of verbs that are only expressed in exactly one category. It turns out that very nearly 90% of these verbs are only expressed in the past participle and no *t*-deverbals, so for the overwhelming majority of verbs for which inference was necessary, it had to proceed from the past participle rather than to it.

Category	**# Freq ≥ 35**	**% Total**	**# Unique**	**% of Cat.**	**% of Uniq.**
PPtc	1006	75.9%	817	81.2%	89.6%
Adverb	18	1.4%	8	44.4%	0.9%
Agent	72	5.4%	20	27.7%	2.2%
Event	178	13.4%	54	30.3%	5.9%
FPtc	52	3.9%	13	25.0%	1.5%
Total	**1326**		**912**	**68.8%**	

Table 27: Past participle type frequencies relative to *t*-deverbals at least as frequent as the thousandth past participle. Many more verbs are attested as past participles than all *t*-deverbals combined. Most verb roots attested in a *t*-deverbal are also attested as a pptc.

Assuming that a given *t*-deverbal form has to be inferred, another option would be to draw its form either from the present or perfect stem then learn exceptional cases like *oblātiō̄*. This correspondence trivially holds for most *ā*-stem and *ī*-stem verbs since most of them have past participles in *-ātus* and *-ītus*, and these actually account for the majority of verbs. The problem is that there are too many exceptions for a learner to acquire this pattern more generally for the same reason that the pptc forms cannot be generalized in this way.

This leaves us with one option, to infer the form of the *t*-deverbal from a corresponding past participle. Referring back to Table 27, 414 verbs (1326−912) have an attested past participle and at least one *t*-deverbal in Perseus, and nearly all of these exhibit the correspondence. Any generalization model worth its salt should be able to learn the pptc ∼ *t*-deverbal correspondence from this data.

Thinking in terms of data sparsity and the Tolerance Principle provides solutions to some

wrinkles in this analysis. First, the presence of *t*-deverbals with verbs lacking past participles is entirely unremarkable in this account. While acquiring the language, the child finds no difference between a semantically arbitrary gapped past participle verb and one whose past participle exists but is just not yet attested, so there is nothing to prevent one from learning an attested *t*-deverbalfor such a verb. Furthermore, the fact that these pptcs remain gapped may suggest that inference did not operate from *t*-deverbal to past participle in practice. Second, the analysis allows for high frequency verbs whose past participles do not correspond to the *t*-deverbal as long as the learner can both from the input. As it turns out, *mortuus* 'dead' ∼ *moritūrus*, *sonitus* 'dead' ∼ *sonātūrus*, and *lautus* ∼ *lavātum* are all attested in this high frequency data set.

This analysis, summarized as "make the *t*-deverbals share a form with the pptc," is a comment on learning rather than representation since there are certainly several ways that this may be implemented in the grammar once the pattern is learned. The shared forms could clue in the learner to implement Embick (2000)'s shared structures (29) or motivate the association between the third stem and both the past participle and various *t*-deverbals in the Aronoff (1994) treatment. It also does not imply that a speaker necessarily construct the *t*-deverbals from the past participles on the fly in the Priscian sense. However, it does cast further doubt on the Steriade (2016) similarity-based syncretism analysis since there is no way to reliably generate *t*-deverbals in *-i-t-*. The following sections will elaborate further on both the synchronic and diachronic implications of this productivity-based treatment of the Latin verb.

7.5. Theoretical Implications

The primary synchronic contribution of this study is the principled analysis of regularity. Rather than relying on the researcher's intuition to decide the most parsimonious division between regular and irregular items, which leads to significant inconsistency between treatments (compare Lieber (1980), Aronoff (1994), Steriade (2016)), an externally motivated model of productivity is calculated empirically over a corpus of relevant data. The

conclusion is perhaps unintuitive. Regularities do exist, particularly in the 1st and 4th conjugations, but most patterns that one could glean from the data, such as the relationship between *-uī* perfects and *-itus* past participles, turn out to be insufficiently reliable.

Furthermore, the study shows that the form of the past participle can be more reliably predicted from the perfect stem than the present. From the present alone, one could form productive generalizations which account for 555 (1st conjugation + 3rd-*iō*) of the thousand verbs tested, while a generalization from the perfect adds an additional 132 verbs, mostly from the 3rd and 4th conjugations, an improvement of about 25%.

The first result, the high rate of unpredictable forms in the 2nd and 3rd conjugations, is largely orthogonal to the choice of theoretical framework. It is more of a clarification about which generalizations should be pursued, and which should be listed by whatever mechanism the framework provides.The analysis provides a particular challenge for Steriade (2016), which relies on relatively robust generalizations in the 2nd and 3rd conjugations that do not hold up (in addition to some problematic empirical claims, see Footnotes 9-10).

7.5.1. The Relationship between the Stems

In conjunction with the second result, that the most robust relationship between stems exists between the perfect and past participle, the high rate of unpredictability allows us to make some more concrete arguments. As Steriade (2016) notes, it would make more semantic sense if there were a relationship between the perfect and the past (perfect) participle. Even though the particular constraint-based analysis which it introduces turns out to work poorly, there does seem to be something to this. Along with the empirical results from the Tolerance Principle, there are several verbs whose pptcs can be shown to have been reworked in Pre-Latin on the basis of the perfect (e.g., *cernō*, *-ere*, ∼ *crēvī*, *crētus* 'separate, perceive,' with inherited pptc *certus* 'certain' attested as an adjective), and additional changes in Late Latin, such as the spread of *-*utu* past participles first among verbs with *-uī* perfects.[14]

[14]The most solid evidence for a relationship between the present and pptc comes from nasal infix verbs (Laurent, 1999; Weiss, 2009). These inherit a nasal infix in their present stems which is not expressed

A theoretical description needs to capture this generalization. For Aronoff (1994), this is just a matter of specifying the realization rule for the past participles to generate their forms from the perfects when possible. It is not more or less efficient than doing so from the present. For Embick (2000), it poses more of a problem.

Recall that in order to account for the periphrastic perfect and the form correspondence between the pptc and *t*-deverbals, the perfects, pptc and *t*-deverbals all share the structure in (29). The correct exponents for the perfect are spell-outs of the [perf] feature on the Aspect head raised T, and sometimes in the context of listed roots (Embick, 2000, (39)) along with whatever readjustments are necessary. The present can be accounted for in the same way. The *-t* or *-s-* of the past participles and *t*-deverbals is the default realization of Asp (27) which is selected when the $\sqrt{}$-*v*-Asp complex has not raised to T. Further readjustment rules (left to the reader) are then needed to alter the shape of the stems to yield the correct forms. This neatly captures the form correspondence because it does not depend on the presence of [perf], [pass], or any other features that are not shared between the past participles and every *t*-deverbal: they share the same structure and are subject to the same readjustment rules.

The issue with this approach is that it is not possible to write a readjustment rule that targets the pptc/*t*-deverbals and perfect stem without also targeting the present since the present and perfect share $\sqrt{}$-*v*-Asp raising to T, and [perf] is not shared between all the *t*-deverbals. Since the contexts are disjoint, it would require parallel rules to achieve a form relationship between the synthetic perfect and pptc. This misses the generalization

elsewhere (*fu**n**dō*, *-ere* ∼ *fūsī* ∼ *fūsus* 'pour')(Poultney, 1937). The infix has been reinterpreted as part of the root and extended to other stems in some verbs (*iu**n**gō*, *-ere* ∼ *iu**n**xī* ∼ *iū**n**ctus* 'join'), which is taken as evidence that the past participle is build on the present. However, the nasal infix is also present on the perfect stem, so it could be the case that the past participle is build on the perfect and the perfect on the present, and that the infix spread in multiple hops. There are several verbs with nasal infixes on the present and perfect but not pptc, which would support this idea (*fi**n**gō*, *-ere* ∼ *fī**n**xī* ∼ *fictus* 'fix'). For nasal infixes to support a relationship between the present and pptc, they would need to appear on the present and pptc but not perfect. There are at most two verbs like this, one of which has been attributed to a transcription error: *pu**n**gō*, *-ere* ∼ *pupugī* ∼ *pū**n**ctus* 'pierce' and *tu**n**dō*, *-ere* ∼ *tutudī* ∼ *tū**(n)**sus* 'beat'. It is possible that the past participle was built on the present for some verbs in Classical or Pre-Latin, but the evidence for it is not strong.

and therefore undercuts one of the treatment's strengths in contrast Aronoff, namely that for Aronoff, that a realization rule happens to connect two stems is essentially accidental, while for Embick, the shared stems are forced by shared structure. But if the readjustment rules that actually give the surface forms for each stem are parallel and just coincidentally identical for a large number of roots, then the account does not really achieve its goal of diminishing the arbitrariness. Furthermore, without shared rules between the perfect and past participle, it is harder to motivate the analogical changes that occurred in favor of shared stem forms between them, since the two could vary independently

7.6. Diachronic Predictions

To summarize, several key changes occurred in the relevant parts of the Latin verbal system on the way from Classical Latin to Romance. First, past participles in *-*atu*, *-*itu* < *ītus*, and *-*utu* expanded at the expense of short -*itus* and -*tus*. The expansion of *-*atu*, *-*itu* occurred throughout the entire Late Latin-speaking world and are clear instances of analogical leveling, while the expansion of *-*utu*, which occurred in most but not all regions (Figure 44), is an example of analogical extension. Related to this, the once common -*itus* is only productive in a couple regions of modern Italy, and reflexes of -*tus* are relegated to high frequency lexical exceptions.

Additionally, new intensives, iteratives, and frequentatives, all of which were regular 1st conjugation verbs, were already being coined in the Classical period and eventually replaced many of their bases in Romance, while at the same time, metaplasm between the 2nd, 3rd, and 4th conjugations caused the crossover of many verbs from one class to another.

The mechanism behind analogical leveling in favor of *-*atu* and *-*itu* is obvious. They were very frequent within their sphere of usage, and so tended to spread over time: *-*atu* through the remainder of the first conjugation and *-*itu* to the others. Reasoning through this in the productivity-based framework for analogical change developed in the previous chapter (Section 6.6.2), a child who had already worked out the productivity of -*ītus* could

over-generalize it to 1st conjugation verbs whose pptcs were not yet attested them, and this could lead to actuation of the analogy if another learner picked it up. The loss of vowel length distinctions in Late Latin combined with overall similar paradigms could even cause confusion between the 2nd through 4th conjugations.

There was no productive past participle form for the large majority of 2nd and 3rd conjugation verbs or -*uī* verbs, so their forms were at the mercy of attestation. Reflexes of -*itus* and -*tus* survive today among high frequency verbs because they were reliably attested in learners' input, but neither was widely productive. This suggests a pathway for *-*utu*'s extension: it was the only productive option for any subset of 2nd and 3rd conjugation verbs with -*uī* perfects. The forms of the perfect are the most reliable evidence for the forms of the past participles, and since it had no productive competitors, it was eventually extended to other forms. In a sense, it was "a big fish in a small pond." As with any instance of analogical extension, the conditions must have been just right for it to occur, which may explain the geographical variation in its expression. Any differences in regional lexicons could have pushed tolerance over the edge and prevented the extension. In Sardinia in particular, the merger of /i/ and /iː/ would have introduced many *-*itu* pptcs into the 2nd and 3rd conjugations and maybe have prevented -*ūtus* from taking off.

A related process may account for the collapse of the pptc ∼ *t*-deverbal correspondence in Romance, where it now appears to hold between the present and *t*-deverbals. To be clear, most Romance past participles themselves share a stem form with their presents because the are regular reflexes of *-*atu*, *-*itu*, and *-*utu*, so it is only exceptional past participles that do not correspond with their *t*-deverbals. This suggests the solution: even in Classical Latin, a correspondence trivally held between most 1st and 4th conjugation presents and their *t*-deverbals in -*āt*- and in -*īt*-. If the number of verbs with past participles like these grew while other past participle forms decreased in number, then learners could have eventually learned a broadly productive present ∼ *t*-deverbal correspondence. Though this hypothesis has not been worked out quantitatively for Late Latin, the rise of regular iteratives and

others at the expense of 2nd and 3rd conjugation verbs is exactly the change that would precipitate this.

7.7. Discussion

The progress made by the present treatment of the Latin past participles is made possible by combining evidence from several sources. The explanatory burden for the patterns we see is shared by representation, which is an expression of the innate human language faculty, language acquisition, and change, which together constitute *third factors* (Chomsky, 2005) relating to the generation and processing of evidence available to speakers who are in the process of instantiating their language faculties with their native language.

When developing an account which combines representation, learning, and change, one can offload some of the explanatory burden from the theory when observations from child development, psycholinguistics, historical records, or other sources can motivate. This results in simpler, more parsimonious theories. Taking a stronger stance, a shared approach yields better accounts. Acquisition and change happen, there is no way around that, and they are actually responsible for some of the patterns observed in modern and ancient attested languages. Furthermore, if our goal is to explain language as it is, and if some pattern actually derives from acquisition rather than representation, it would be incorrect to handle it primarily with representation without reference to acquisition. Delimiting the roles of several factors is not a trivial problem, and it is usually easier said than done, but it is not impossible. I would like to see as much care put into sorting out the causes of interesting linguistic problems as is put into their theoretical analysis.

Consider the alternative, in which all explanation is focused in one area. On one hand is a classic philological approach, which while excellent in collecting descriptive data, leaves something to be desired in terms of explanation. As exemplified in the Proto-Germanic case study, traditional approaches were able to conclude that analogical change drove the reconstructed patterns but lacked to tools to disambiguate them. On the other hand, is a

traditional theoretical approach in which representation is possible for generating all possible forms in the language while excluding all those that are not possible. In this chapter, for example, diachrony motivates a synchronic relationship between the perfect and pptc stems and acquisition establishes it. Theories that miss out on this generalization (Embick, 2000) are insufficient by their own standards. Other patterns should probably not be handled in representation in the first place. To take another example from Steriade (2016), the past participle of *vincō* 'conquer' is *victus*, and its agent is *victor*. The SBS account excludes logically possible alternatives like **vixor* explicitly in representation, but third factors provide another answer: The past participle and agent are in correspondence because of their etymologies. Both were prehistorically built on the same Indo-European nominalizer in **-to-* and were then subject to the same sound changes. Additionally, learners receiving a sample of Latin as input are extremely unlikely to ever hypothesize a form **vixor* because there is no evidence that would motivate it. This can be handled before theory, and it is probably more correct not to do so redundantly in representation.

CHAPTER 8 : Conclusions

It is quite common to invoke the process of child language acquisition when developing explanatory models for change, but treatments vary greatly in the extent to which they focus on the specifics of the acquisition process. The "learners' perspective" approach laid out here distinguishes itself by placing the child front and center. An account begins with how a learner is modeled to behave in response to data, and the consequences of that behavior are worked out to determine what kind of change it should drive in the long run and the implications that has for synchronic representation. In that sense, this is historical linguistics from the perspective of the child language learner.

I apply this learner's perspective to four studies: first, a treatment in Chapter 4 of the sporadic innovation of transparent /aɪ/-raising as phonological acquisition in a mixed canonical and non-raising environment, which accounts for the sparse distribution of transparent raising in places at the edges of the raising region, the position of lexical raising, and for the role of "monolingual" acquisition in language change. Second, a treatment of the Middle and Early Modern English *to*-dative in Chapter 5 which casts its innovation as a plausible reanalysis of recipient-like goal constructions and its spread as learner over-generalization. This accounts for the construction's "over-extension" and retreat without the faults of existing morphological erosion and borrowing accounts. Third, a model for acquisition-driven analogical change which provides new insights into the origin of the Proto-Germanic strong verbs' lengthened *$\bar{e}$-grade in Chapter 6 which distinguishes between the unlikely but possible 'Eat' Analogy account and the superficially plausible but impossible Class VI Analogy account. The "Sibling-Induced" Change framework introduced in that chapter allows us to reason about child-driven innovation and propagation. Finally, Chapter 7 discusses the Latin past participles and *t*-deverbals in terms of productivity to elucidate differences between competing theoretical accounts and to account for counter-intuitive changes which the system underwent in Romance.

The learners' perspective on change also sheds new light on the utility of historical corpora. Corpora are taken as tools for estimating child lexicons during the time of acquisition rather than evidence in their own right for grammaticality. This turns the conventional wisdom of corpus research on its head: it is the similarities rather than the differences between that are most interesting. Only the best-attested properties of the corpora are relevant for estimating child experience, and trimming the less attested items from the corpora removes most of their individual properties. Lexicons from different genres and slightly different eras can be collated in a way that would be problematic in most other circumstances.

8.1. Concrete Mechanisms in Historical Explanation

It is the step of committing ourselves to a concrete acquisition mechanism that opens up new avenues for analysis: in Chapter 4, applying the Tolerance Principle to the acquisition of /aɪ/-raising brought together seemingly contradictory facts from competing accounts. Phonological acquisition in the face of even minimal variation explains the innovation of transparent /aɪ/-raising and its sparse distribution. In Chapter 5, connecting the innovation of the *to*-dative in Middle English to transient over-generalizations in modern child productions through the Sufficiency Principle explains not only its rapid rise in historical corpora, but also its broad Middle English distribution, something dismissed as irrelevant in previous studies. It further challenges functional treatments that link the *to*-dative to the loss of overt case marking with an afunctional account that better accounts for the cross-linguistic facts.

In the study on Proto-Germanic in Chapter 6, adopting a quantitative model of productivity learning raised the bar for rigor in historical explanation. Several accounts for the lengthened *$\bar{e}$-grade have been proposed over the last century and a half, many of which are indeed consistent with the descriptive facts, but if two treatments are consistent with the facts, and both are intuitively plausible, how can we distinguish between them? The productivity-based model adopted here forces us to favor one account over another. Under its conditions, the Class VI Analogy turns out to be not only unlikely, but actually incompatible with later

analogy from Class V to IV, while the 'Eat' Analogy does seem to have been possible, albeit unlikely.

This concept is taken further in the Chapter 7 study of the Latin past participles. The replacement of frequent Classical *-itus* and *-tus* with reflexes of *-ūtus* has puzzled Romance linguists. Why would an ending that was infrequent in the language have replaced two of the most common endings, and why was it *-ūtus* as opposed to some other infrequent ending that spread? Calculating out the productivity of past participle stem formations reveals some counter-intuitive results, namely, that *-itus* and *-tus* are actually unproductive despite their frequencies while *-ūtus* is productive. Connecting analogical change to productivity, this provides a novel explanation for the fate of the Romance past participles.

I used the Tolerance and Sufficiency Principles in these studies because of their strengths as models of productivity learning, but the real moral of this approach is not limited to that. Committing to a reasonable acquisition model based on study of modern learners and speakers makes testable predictions and stands to improve our understanding of diachronic processes. Diachronic research will never have access to the full gamut of sources of evidence available to synchronic researchers, but applying what we know about the present to the past raises the level of rigor in historical hypothesis testing closer to what can be achieved in the present.

8.2. Sharing Explanatory Burden

Most generally, the greatest strength of the learners' perspective is that it presents a way for us to delimit explanatory roles in accounting for linguistic facts. The burden in explaining the patterns we observe in language is shared by change, representation, an expression of the innate human language faculty, and language acquisition, an expression of the language faculty, general cognitive processes, and a learner's immediate environment. Together, these constitute *third factors* (Chomsky, 2005) relating to the generation and processing of evidence available to speakers who are in the process of instantiating their language faculties

with their native language.

This results in simpler and more parsimonious theories. Taking a stronger stance, a shared approach yields *better* accounts as well. Acquisition and change happen, there is no way around that, and they are primarily responsible for some of the patterns observed in the languages of today. If some pattern is actually the result of acquisition rather than representation, it would be incorrect to handle it purely in representation without reference to acquisition. The same can be said for acquisition and change. Delimiting the roles of several factors is not a trivial problem, and it is usually easier said than done, but it is not impossible.

Contrast this approach with some alternatives in which all explanation is focused in one area: in the prior studies of the *to*-dative cited in Chapter 5, for example, specific theoretical assumptions carry all the explanatory burden for the development of the construction. Without engaging with acquisition, these accounts lack a means for explaining the over-generalization of the construction, and given the empirical issues with morphological erosion, they cannot explain its actuation either. Also consider the discussion surrounding the Latin past participle *t*-deverbal correspondence. On one hand, a purely theoretical account seeks to explain the complex semantic properties of the categories and describe why they share their form despite not sharing their meaning. However, diachronic evidence, once taken into account, shows that nothing much needs to be explained: they share their forms because they share their etymologies. That said, diachrony alone cannot explain all the relevant facts, and here is where acquisition and theory come in to complete the analysis. Traditional accounts of the Proto-Germanic lengthened *$\bar{e}$-grade, run into a similar problem because are disconnected from synchronic issues. Though they are strong in differentiating between descriptively adequate models, they struggle when description is not enough.

To take an even stronger stance, purely representational approach to change is actually insufficient in accounting for actuation in much the same way that atheoretical approaches are insufficient in accounting for competence. To the extent that representation is primarily

a function of our innate and universal human language faculty, it is a static "always on" factor (Hale, 1998; Walkden, 2017). Why, then, would it drive some change in one time and place and a different change or no change in another? It must be that dynamic third factors catalyze representational change.

I would like to end this section with a caveat. The particular balance that I strike in this work, which favors mechanisms of acquisition over theoretical analysis, effectively accounts for several historical phenomena in four languages, but it is not necessarily the optimal approach for any particular problem. I am in favor of tripartite treatments for linguistic phenomena in principle in whatever guise they take. What I want is for as much thought to be put into how acquisition, theory, and diachrony work together to create a pattern of language beforehand as is put into the theoretical analysis of the problem afterward. We should ask at what level the *t*-deverbals or the *to*-dative or any other phenomenon should be accounted for before diving headlong into them.

8.3. The Paradox of Language Change

The Paradox of Language Change is the disconnect between the claim that child language acquisition is among the primary drivers of language change and the observation that children are faithful learners of their native languages. Working through the Paradox requires a more rigorous understanding of what it means to be a driver of change. The component of change most relevant to the question has to be actuation, its initial innovation and propagation (Labov et al., 1972) because after actuation, the change runs loose in a community, and it no longer matters who its innovators were. Thus, resolving the Paradox requires answers for both innovation and initial propagation.

The pathway towards solving the Paradox becomes clear when we relax some classic simplifying assumptions about the learner and the learner's input, namely

8.3.1. The Paradox and Innovation

There are a few ways out of the innovation problem. The first is to "blame the learner," that is, that learners sometimes introduce errors into the acquisition process despite receiving enough evidence to make the right choices. The second option is to "blame the environment" instead, that is, to progress past the classic simplifying assumption that children learn in ideal single input source environments (cf. Chomsky and Halle, 1968). We know that children receive input from multiple people who themselves may exhibit internal variation, and that change is formally *inevitable* in the face of even minimal variation in the input (Niyogi and Berwick, 1997). Further, the effect of the Poverty of the Stimulus is not to be discounted. Even the richest innate specifications to the language faculty render acquisition tractable, not trivial. Measures of sparsity such as paradigm saturation (Section 3.3, Chan, 2008) show that much of what a child is tasked with learning is simply not available to them, and cases of divergent grammar outcomes under Abject Poverty suggest that learners really are operating at the edge of what is feasible.

The blame for child innovation in morphological systems can be pinned on input sparsity. The number and diversity of forms available to any given learner is unlikely to be enough to uniquely specify even a moderately large paradigm, so it is normal, over the course of development, for a learner to over-generalize and innovate new forms. The numbers work out this way for Proto-Germanic and also Latin. More generally, abject poverty is the condition where the input is so under-specified that learners may not all converge on the same grammar. In the case of the innovated *to*-dative, the surface word order of certain goal constructions supported a novel parse that was just as plausible as the intended meaning. If there is even a one in a million chance that some children never grow out of their innovations and transmit them to others, then changes will accrue over time.

To make matters worse, variation is an entirely normal part of the acquisition process, and this significantly exacerbates the problem of data sparsity. This is not only an issue of multilingual or multi-dialectal learning environments: even "monolingual" acquisition as

typically conceived is actually acquisition in the face of variation between minimally distinct varieties (Chapter 4). If data sparsity and variation are both unavoidable, then they can serve as critical components in a model of (even monolingual) acquisition-driven change (Sections 2.2.1 and 6.6.2).

8.3.2. The Paradox and Actuation

The actuation of a change requires both its innovation and its entry into a speech community (Labov et al., 1972, p. 7). Even granting that sparse input and variation contribute to innovation, how can children be responsible for propagation? In Sections 6.6.2-6.6.3, I lay out "Sibling-Induced' Change as a framework for reasoning about actuation in the context of the Paradox of Language Change.

The most important thing to recognize here is that language transmission is not strictly generational (Manly, 1930; Weinreich et al., 1968; Roberts and Labov, 1995; Labov, 2001; Nardy et al., 2014). Other children transmit linguistic features to their peers. The impact of non-generational transmission is best conveyed with a thought experiment. Consider two children, Alice and Bob, and say Alice is Bob's older sister. Alice is currently entertaining an innovative grammar and is occasionally producing novel forms. How might little Bob react to Alice? Crucially, Bob may not even be able to recognize Alice's innovation because of input sparsity. Whether a novel phonological form of a given word, a new inflection, or a new spell-out of a syntactic construction, there is a real chance that Alice's innovation is the only token he will hear, and since Alice is mostly consistent with adults – she can presumably communicate with her parents – there is no reason to assume that Alice is acting oddly.

If Bob cannot recognize Alice's innovation, then he may adopt it. At the very least, he is receiving unambiguously innovated input, so he is more likely to make a similar innovation himself or extend Alice's, as in Chapter 5's study. If he does recognize it, he has options: he can adopt it categorically, adopt it as some kind of socially influenced variant, or reject

it. Even young children do begin to orient towards their peers rather than their parents, a pattern that continues through adolescence (Labov, 1989; Roberts and Labov, 1995; Nardy et al., 2014). So as long as Bob does not find Alice's innovation egregiously ungrammatical, he may choose to adopt it if he values Alice's social prestige. If a three or four year-old Alice is cool to anyone, it will be two or three-year-old Bob.

The fundamentals of "sibling-induced" change are input sparsity and variation with transmission that is not purely generational. Figure 42b visualizes the framework as an extension of the Z-model. In Section 6.6.3, I provide a first validating simulation for the framework by reproducing a classic pattern in the distribution of morphological irregularity through analogical change. This *Conserving Effect* of high frequency is the observation that high frequency morphological forms are more likely to be irregular than low frequency ones (Bybee, 1985), which, cast in diachronic terms, is a resistance to analogical levelling. I derive the Conserving Effect in a simulation of sibling-induced Z-model transmission of a morphological paradigm in which irregulars are initially uniformly distributed. In the simulation, low frequency items are slowly levelled over several iterations of transmission until only high frequency irregulars remain.

8.4. Conclusion

It occurred to me a couple weeks before finishing this dissertation that I was thinking about the Paradox of Language Change during the season that I applied to graduate programs. So, while have I pursued other lines of study, I think it is fair to say that the research presented here is truly the culmination of my time as a PhD student. My hope for this work is that it will serve as a foundation for me and for others who are interested in taking a learners' perspective on language change and for those interested in the case studies that I have discussed. May it be an adequate initial response to Lehmann's (2013 §1.9.2) challenge, "*Advocates of change resulting from language acquisition must demonstrate when and how such patterns are adopted.*"

APPENDIX

A.1. Appendix to Chapter 3

English, Spanish, Latin, and Proto-Germanic lexicons that are reported in Chapter 3. Lexicons are presented as tables with items in the leftmost column sorted from most frequent to least frequent and reasonable matches presented in the right columns when available.

Brent	**Brown**
go	go
get	get
come	come
do	do
see	see
put	put
have	have
want	want
know	know
say	say
look	look
take	take
like	like
eat	eat
play	play
think	think
give	give
sit	sit
try	try
make	make
let	let
need	need
throw	throw
read	read
find	find
turn	turn
tell	tell
fall	fall
love	love
open	open
hear	hear
hold	hold
leave	leave
wash	wash
walk	walk
help	help
pull	pull
stand	stand

Brent	**Brown**
feel	feel
move	move
wait	wait
close	close
sing	sing
keep	keep
run	run
push	push
clean	clean
watch	watch
roll	roll
stop	stop
talk	talk
bring	bring
bite	bite
chew	chew
hurt	hurt
pick	pick
hit	hit
drink	drink
excuse	excuse
ready	ready
tire	tire
finish	finish
stick	stick
change	change
call	call
use	use
lay	lay
wear	wear
kiss	kiss
show	show
fix	fix
sleep	sleep
catch	catch
drop	drop
bless	bless
hang	hang

Brent	**Brown**
wipe	wipe
suppose	suppose
blow	blow
work	work
shake	shake
cry	cry
dry	dry
mess	mess
touch	touch
happen	happen
climb	climb
meow	
mean	mean
smell	smell
remember	remember
start	start
guess	guess
clap	
wet	wet
shut	shut
sound	sound
moo	
kick	kick
tickle	tickle
fuss	
jump	jump
peekaboo	
feed	feed
zoom	
crawl	
pant	pant
forget	forget
brush	brush
zip	
taste	taste
fly	fly
wave	
buy	buy

Table 28: Section 3.2 English comparisons in Brent token frequency rank order

Brent	Brown
hand	hand
fit	fit
cut	cut
reach	reach
check	check
lose	lose
miss	miss
hide	hide
listen	listen
build	build
break	break
knock	knock
ask	ask
spin	roll, turn, screw
back	
learn	learn
laugh	laugh
pat	
ride	ride
dump	dump
beat	beat
wake	wake
dirty	dirty
spit	
grab	take
bake	bake
scratch	scratch
hug	hug
write	write
dress	dress
bounce	bounce
bark	
understand	understand
bet	bet
excite	excite
press	press
wonder	wonder
swish	
comb	comb
purr	
hope	hope
dance	dance
grow	grow
slide	slide
pattycake	
tweet	
step	step
pack	pack
woof	

Brent	Brown
snap	snap
scare	scare
rain	rain
mix	mix
save	save
rock	rock
pour	pour
live	live
clink	
figure	
bang	bang
rub	
ring	ring
mark	
count	count
care	care
quit	stop
pet	
pay	pay
tear	tear
send	send
point	point
chase	chase
bump	bump
begin	begin
spill	spill
act	
thank	
rinse	wash
rid	
rejoice	
fill	fill
carry	carry
believe	believe
worry	
pinch	
follow	follow
color	color
share	share
fight	fight
drive	drive
cool	cool
wish	wish
swing	swing
cook	cook
stuff	fill
record	

Brent	Brown
belong	belong
suck	
set	set
rip	rip
lick	lick
cover	cover
attack	fight, hit, bang
stir	stir
smile	
lift	lift
ignore	
head	
fold	fold
warm	
stink	smell
side	
decide	
allow	let
sweep	wipe, brush
praise	
meet	meet
lean	
answer	
time	time
ding	
tip	tip
swim	swim
still	
pop	pop
hop	hop
whisper	
water	wet
splash	wet
speak	talk
lock	lock
poke	
interest	interest
hurry	hurry
holler	yell
visit	visit
tie	tie
march	march
hate	hate
caw	
bonk	
block	
bend	bend

Table 29: Section 3.2 English comparisons (continued)

Spanish	Translation	English
ir	'go'	go
tener	'have'	have
hacer	'do'	do
ver	'see'	see
querer	'want'	want
poner	'put'	put
decir	'say'	say
poder	'can'	
mirar	'look'	look
dar	'give'	give
caer	'fall'	fall
comer	'eat'	eat
tomar	'take'	take
jugar	'play'	play
saber	'know'	know
pasar	'pass'	cross
contar	'count, tell'	count, tell
dormir	'sleep'	sleep
esperar	'wait'	wait
meter	'put'	put
caber	'fit'	fit
llamar	'call'	call
venir	'come'	come
coger	'take'	take
dejar	'leave'	leave
llevar	'carry'	carry
parecer	'seem'	
salir	'leave'	leave
abrir	'open'	open
tirar	'throw'	throw
faltar	'be lacking'	
cerrar	'close'	close
montar	'climb on'	climb
gustar	'like'	like
correr	'run'	run
creer	'believe'	beve
pintar	'paint'	paint
quitar	'remove'	
subir	'go up'	
sacar	'bring out'	bring, show
sentar	'sit'	sit
traer	'carry'	carry
romper	'break'	break
entrar	'come in'	
acordar	'remind'	
bajar	'come down'	

Spanish	Translation	English
oír	'hear'	hear
tocar	'touch'	touch
quedar	'stay'	
volar	'fly'	fly
guardar	'guard'	save
echar	'throw'	throw
andar	'walk'	walk
nadar	'swim'	swim
llegar	'arrive'	come, reach
cantar	'sing'	sing
ayudar	'help'	help
buscar	'search'	
acabar	'finish'	finish
empezar	'begin'	begin
disparar	'shoot'	shoot
hablar	'speak'	
ganar	'win'	win
llorar	'cry'	cry
conducir	'drive'	drive
patinar	'skate'	
enseñar	'teach'	teach
cenar	'dine'	dine
levantar	'raise'	lift
pensar	'think'	think
mover	'move'	move
mamar	'suck'	
trabajar	'work'	work
saltar	'jump'	jump
agarrar	'grab'	
valer	'cost'	
sonar	'sound'	sound
sentir	'feel'	feel
pedir	'request'	ask
encontrar	'find'	find
preparar	'prepare'	make, build
seguir	'follow'	follow
esconder	'hide'	hide
volver	'turn'	turn
llover	'rain'	rain
desayunar	'breakfast'	eat, dine
comprar	'buy'	buy
despertar	'wake up'	wake
leer	'read'	read
vivir	'live'	live
terminar	'end'	
pegar	'hit adhere'	hit, stick, glue

Table 30: Section 3.2 Spanish comparisons with Brown sorted by Spanish token frequency rank

Spanish	Translation	English
colocar	'place'	put
cortar	'cut'	cut
morir	'die'	die
mojar	'moisten'	wet
fijar	'notice'	see, hear
perder	'lose'	lose
columpiar	'swing'	swing
escribir	'write'	write
beber	'drink'	drink
pillar	'catch'	catch
chocar	'collide'	knock, hit
vestir	'dress'	dress
deber	'must'	need
preguntar	'ask'	ask
dibujar	'draw'	draw
quemar	'burn'	burn
picar	'prick, dive'	
enganchar	'hook'	
merendar	'lunch'	eat, dine
colorar	'color'	color
nevar	'snow'	snow
platicar	'talk'	talk
curar	'cure'	
soplar	'blow'	blow
contentar	'satisfy'	
oler	'smell'	smell
matar	'kill'	kill
lavar	'wash'	wash
apagar	'turn off'	
tapar	'cover'	cover
igualar	'match'	match
doler	'hurt'	hurt
cambiar	'change'	change
cansar	'tire'	tire
aparcar	'park'	park
entender	'understand'	understand
servir	'serve'	
reír	'laugh'	laugh
morder	'bite'	bite
escuchar	'listen'	listen
sujetar	'subject'	
explicar	'explain'	teach
parar	'stop'	stop
bañar	'bathe'	clean, wash
soñar	'dream'	dream
necesitar	'need'	need
estudiar	'study'	learn
enfadar	'anger'	
unir	'unite'	
asustar	'frighten'	scare
aguar	'water down'	
soltar	'release'	leave
colgar	'hang'	hang
apretar	'tighten'	
pasear	'stroll'	walk
conocer	'know'	know
gritar	'shout'	yell
llenar	'fill'	fill
arreglar	'order'	
empujar	'push'	push
botar	'bounce'	bounce
limpiar	'clean'	clean
juntar	'assemble'	build
asomar	'show'	show
alar	'call'	call
regalar	'give present'	
probar	'test'	try
portar	'behave'	
pesar	'weigh'	
encender	'light'	
tumbar	'knock down'	knock, hit
partir	'divide'	split
largar	'go away'	
escapar	'escape'	
aprender	'learn'	learn
funcionar	'function'	work
ahogar	'drown'	
importar	'matter'	matter
casar	'marry'	marry
cuidar	'take care'	care
acostar	'put to bed'	
apuntar	'indicate'	
rodar	'roll'	roll
pisar	'press'	press
manchar	'stain'	dirty
intentar	'attempt'	try
caminar	'walk'	walk
piar	'chirp'	
durar	'endure'	
acercar	'bring near'	bring
tostar	'toast'	cook, burn
serrar	'saw'	saw

Table 31: Section 3.2 Spanish comparisons (continued)

Spanish	Translation	English
divertir	'entertain'	play
pelear	'fight'	fight
estrellar	'shatter'	, crack
destrozar	'wreck'	, break
cocinar	'cook'	cook
callar	'silence'	
calentar	'heat'	
luchar	'fight'	fight
lanzar	'hurl'	throw
castigar	'punish'	
bailar	'dance'	dance
ordenar	'order'	
freír	'fry'	cook
encerrar	'enclose'	close
crecer	'grow'	grow
mandar	'order'	
inventar	'invent'	
dañar	'harm'	hurt
amar	'love'	love
reñir	'quarrel'	fight
prender	'ignite'	
pescar	'fish'	fish
recomendar	'recommend'	
filmar	'film'	
contestar	'answer'	
usar	'use'	use
navegar	'navigate'	
hundir	'sink'	drop
estropear	'damage'	break,
enojar	'annoy'	bother
encajar	'encase'	
empatar	'tie with'	tie
aparecer	'appear'	
vengar	'avenge'	
tratar	'treat'	
preocupar	'preoccupy'	busy
ordeñar	'milk'	milk
molestar	'bother'	bother
encantar	'enchant'	
duchar	'take shower'	
chupar	'suck'	
arrancar	'snatch'	, take
soler	'accustom'	
hechar	'give'	give
grabar	'engrave'	scratch
regar	'water'	
espantar	'frighten'	scare
saludar	'greet'	meet
robar	'rob'	steal
pinchar	'pinch'	
elegir	'choose'	pick
cumplir	'accomplish'	
secar	'dry'	dry
regresar	'return'	
notar	'note'	
imaginar	'imagine'	think
fallar	'fail'	lose
chillar	'scream'	yell
bastar	'suffice'	
atacar	'attack'	fight
apetecer	'crave'	want
utilizar	'use'	se
sonreír	'smile'	
nacer	'be born'	
balancear	'balance'	
acompañar	'accompany'	
trepar	'scramble'	mix
girar	'revolve'	turn
galopar	'gallop'	run
bucear	'plunge'	
aterrizar	'land'	
atar	'fasten'	tie, stick, glue
vigilar	'watch over'	
titular	'greet'	meet
suceder	'happen'	happen
salvar	'save'	save
roncar	'snore'	
ocupar	'occupy'	busy
cazar	'hunt'	

Table 32: Section 3.2 Spanish comparisons (continued)

Proto-Germanic	Translation	English	Spanish
*bīdaną	'wait (for)'	wait	*esperar*
*bītaną	'bite'	bite	*morder*
*diganą	'knead'	rub, stir	
*drībaną	'drive'	drive	*conducir*
*gī-n-	'yawn, gape'		
*glītaną	'glitter'		
*grīpaną	'grasp'	hold	*agarrar*, *tener*
*hnīwaną	'bow'		
*kīnaną	'sprout'		*crecer*
*klībaną	'cling'	stick/hold	*enganchar*, *pegar*
*kli-n-	'smear'	wipe	
*bi-lībaną	'remain'		*quedar*
*līhwaną	'lend'		*dar*, *regalar*
*līþaną	'go'	go	*ir*
*mīganą	'urinate'	pee	
*mīþaną	'change->avoid'	change	*cambiar*
*nīpaną	'get dark'		
*rīdaną	'ride'	ride	*gallopar*, *conducir*
*rīfaną	'tear'	tear, rip	*romper*
*rīsaną	'rise'		*subir*, *montar*
*sīhwaną	'filter'		*partir*
*skīnaną	'shine'		*enceder*
*skītaną	'defecate'		
*skrītaną	'tear'	tear, rip	*romer*
*slīdaną	'slide'	slide	
*slīkaną	'slink'		
*smītaną	'fling'	throw	*tirar*, *echar*
*snīþaną	'cut'	cut	*cortar*
*snīwaną	'snow'	snow	*nevar*
*spīwaną	'split'	cut, crack, snap	*cortar*, *partir*
*stīganą	'go up/down'		*subir*, *montar*
*stikaną	'pierce'		*picar*
*strīkaną	'stroke'		
*swībaną	'cease'	stop	*parar*, *terminar*, *acabar*
*tīhaną	'make solemn decl.'		
*wiganą	'fight'	fight	*luchar*, *pelear*, *reñir*
*wīkaną	'yield'	stop, wait	*esperar*, *parar*
*wīpaną	'encircle'		*encajar*
*wītaną	'admonish'		*castigar*
*wlītaną	'look'	look	*mirar*
*wrītaną	'scratch'	scratch	*rascar*, *grabar*
*beudaną	'offer'		*dar*, *regalar*
*beuganą	'bend'	bend	
*blewwaną	'beat'	hit, beat, spank	*pegar*
*brewwaną	'brew'		

Table 33: Section 3.2 Proto-Germanic comparisons

Proto-Germanic	Translation	English	Spanish
**dreuganą*	'be retainer'		*servir*
**(bi-)dreuganą*	'deceive'	fool,trick	
**dreusaną*	'fall'	fall	*caer*
**fleuganą*	'fly'	fly	*volar*
**fleuhaną*	'flee'		*escapar*
**fleutaną*	'flow'		*correr*
**freusaną*	'freeze'		
**geutaną*	'pour'	pour	*regar*
**heufaną*	'lament'	cry	*llorar*
**hleutaną*	'cast lots'		
**hneupaną*	'pluck'		*arrancar*
**keusaną*	'test'	try	*probar*
**kewwaną*	'chew'	chew	*comer*, *morder*
**kleubaną*	'split'	cut, crack, snap	*cortar*, *partir*
**leudaną*	'grow'	grow	*crecer*
**leuganą*	'tell a lie'		
**leukaną*	'pluck'		*arrancar*
**fra-leusaną*	'lose'	lose	*perder*
**neutaną*	'use/enjoy'	use, like	*usar*
**reudaną*	'redden'		
**reufaną*	'tear'	tear, rip	*partir*
**reutaną*	'weep'	cry	*llorar*
**seukaną*	'be sick'		*enfermar*
**seuþaną*	'boil'	cook, bake	*cocinar*, *calentar*
**skeubaną*	'push/shove'	push	*empujar*
**sleupaną*	'slip'	slip, slide	*resbalar*
**smeuganą*	'creep/bend'	bend	
**smeukaną*	'smoke'	smoke	*fumar*
**spreutaną*	'sprout'		*crecer*
**teuhaną*	'lead/pull'	pull	*conducir*
**þeutaną*	'resound loudly'	yell, bang	*sonar*, *chillar*, *gritar*
**þreutaną*	'tire out'	tire	*cansar*
**lūkaną*	'close'	close	*cerrar*
**lūtaną*	'bow'		
**sūpaną*	'slurp'	drink	*chupar*
**þrūtanaz*	'swollen'		*crecer*
**bindaną*	'tie'	tie	*empatar*, *enganchar*, *pegar*
**brinnaną*	'burn'	burn	*quemar*
**drinkaną*	'drink'	drink	*beber*
**finþaną*	'find'	find	*encontrar*
**-ginnaną*	'begin'	start, begin	*empezar*
**grindaną*	'grind'		
**hinþaną*	'seize'	take	*tomar*, *coger*, *agarrar*
**linnaną*	'go away'	leave	*largar*, *dejar*
**rinnaną*	'run/flow'	run	*correr*
**singwaną*	'sing'	sing	*cantar*

Table 34: Section 3.2 Proto-Germanic comparisons (continued)

Proto-Germanic	Translation	English	Spanish
**sinkwanǫ*	‘sink’		*hundir*
**fra-slindanǫ*	‘swallow’	swallow	*comer*
**spinnanǫ*	‘spin’	turn	*volver*, *revolvar*
**stinkwanǫ*	‘knock’	knock, hit, bump	*tumbar*, *chocar*
**swimmanǫ*	‘swim’	swim	*nadar*
**þinhanǫ*	‘thrive’	win	*cumplir*, *contentar*
**þinsanǫ*	‘pull’	pull	*traer*
**þrinhanǫ*	‘press’	push, press	*pisar*
**þwinganǫ*	‘force’		
**windanǫ*	‘wind/wrap’	wind, wrap	*empatar*
**winnanǫ*	‘struggle’	fight	
**wringanǫ*	‘twist’	turn, roll, screw	*torcer*, *volver*, *rodar*
**belganǫ*	‘swell’		*crecer*
**berganǫ*	‘hide/keep’	keep, hide	*esconder*
**derbanǫ*	‘exert oneself’		
**felhanǫ*	‘penetrate’		
**fertanǫ*	‘fart’		
**geldanǫ*	‘pay (for)’	pay	*comprar*
**helpanǫ*	‘help’	help	*ayudar*
**hwerbanǫ*	‘turn’	turn, roll	*volver*, *rodar*
**kerbanǫ*	‘cut/carve’	cut	*cartar*, *grabar*
**melkanǫ*	‘milk’	milk	*ordeñar*
**meltanǫ*	‘melt’	melt	*calentar*
**skerfanǫ*	‘gnaw’	chew	
**smertanǫ*	‘hurt’	hurt	*doler*, *dañar*
**snerpanǫ*	‘contract/shrivel’		*apretar*
**sterkanǫ*	‘congeal’		
**swellanǫ*	‘swell’		*crecer*
**sweltanǫ*	‘die’	die	*morir*
**swerbanǫ*	‘wipe off’	wipe	*limpiar*, *lavar*
**þersanǫ*	‘dry out’	dry	*secar*
**weltanǫ*	‘roll’	roll, turn	*volver*, *rodar*
**welwanǫ*	‘rob’	take, steal	*agarrar*, *coger*
**werpanǫ*	‘throw’	throw	*tirar*, *echar*
**werþanǫ*	‘become’	become	*cumplir*
**fehtanǫ*	‘fight’	fight	*luchar*, *pelear*, *reñir*
**flehtanǫ*	‘plait’		
**þreskanǫ*	‘thresh’		
**wreskwanǫ*	‘grow’	grow	*crecer*
**mur-n-*	‘mourn’		
**spurnanǫ*	‘trample’		
**beranǫ*	‘carry’	carry	*llevar*, *traer*
**brekanǫ*	‘break’	break, snap, crack, rip	*romper*
**bremanǫ*	‘roar/bellow’		
**dwelanǫ*	‘be confused’		
**helanǫ*	‘hide’	hide	*esconder*

Table 35: Section 3.2 Proto-Germanic comparisons (continued)

Proto-Germanic	Translation	English	Spanish
**knudaną*	'knead'		
**kwemaną*	'come'	come	*venir*, *llegar*
**nemaną*	'take'	take	*tomar*, *coger*, *agarrar*
**skeraną*	'cut/shear'	cut	*cortar*
**snewaną*	'hurry'	run, hurry	*correr*
**stelaną*	'seal'	close, glue	*encerrar*, *encajar*
**stenaną*	'groan'		
**temaną*	'fit'	fit	*caber*
**teraną*	'tear'	tear, rip	*arrancar*
**trudaną*	'step on'	step, kick	
**wulaną*	'seethe/be agitated'		*enfadar*
**etaną*	'eat'	eat	*comer*
**ga-fehaną*	'rejoice'		*gustar*
**fetaną*	'fall'	fall	*caer*
**gebaną*	'give'	give	*dar*
**bi-getaną*	'get'	get	
**hlefaną*	'steal'	take, steal	*agarrar coger*, *arrancar*
**hreþaną*	'sift'		
**jehaną*	'affirm'		
**jesaną*	'ferment'		
**kweþaną*	'say'	say	*decir*, *contar*
**lekaną*	'be leaky'		
**lesaną*	'gather'		
**metaną*	'measure'		*contar*
**ga-nesaną*	'survive'	live	*vivir*
**rekaną*	'bank a fire'		
**sehwaną*	'see'	see	*ver*
**swefaną*	'fall asleep/sleep'	sleep	*dormir*, *acostar*
**treganą*	'grive'	cry	*llorar*
**webaną*	'weave'	sew	
**wedaną*	'join'	tie	*emputar*, *pegar*
**weganą*	'move'	move	*mover*
**wesanǫ1*	'be/remain'	be, wait	*quedar*
**wesanǫ2*	'feast'	eat, dine	*comer*, *merendar*, *desayunar*
**wrekaną*	'drive out'		*quitar*
**bidjaną*	'ask for'	ask	*preguntar*, *pedir*
**ligjaną*	'lie'	lay	
**sitjaną*	'sit'	sit	*sentar*
**fregnaną*	'ask/find out about'	ask	*volver*, *rodar*
**akaną*	'drive'	drive	*conducir*
**alaną*	'nourish/raise child'		
**ananą*	'breathe'	blow	
**ga-dabaną*	'happen to'	happen	*suceder*
**drabaną*	'break up?'	break, cut, rip	*romper*, *parti* , *estrellar*
**draganą*	'haul'	pull, carry	*traer*
**faraną*	'journey'	go, walk	*,conducir*
**flahaną*	'skip'	hop	*saltar*, *botar*

Table 36: Section 3.2 Proto-Germanic comparisons (continued)

Proto-Germanic	**Translation**	**English**	**Spanish**
**galaną*	'sting'		*picar*
**grabaną*	'dig'	dig	
**hlaþaną*	'load'	fill	*cargar*
**kalaną*	'be cold'		
**lahaną*	'reproach'		*castigar*
**malaną*	'grind'		
**sakaną*	'dispute'		*luchar*
**skabaną*	'shave'	shave	
**slahaną*	'hit/kill'	hit, kill, punch	*pegar*, *matar*
**þwahaną*	'wash'	wash	*lavar*, *limpiar*
**wadaną*	'walk/wade'	walk	*andar*, *caminar*, *pasear*
**fraþjaną*	'understand'	think, understand	*endender*, *saber*
**habjaną*	'lift'	lift	*levantar*
**hlahjaną*	'laugh'	laugh	*reir*
**kwabjaną*	'extinguish'		*aguar*, *terminar*
**sabjaną*	'notice'	see, hear	*fijar*
**skapjaną*	'create'	make	*juntar*, *preparar*
**skaþjaną*	'harm'	hurt	*dañar*
**swarjaną*	'swear'		
**wahsijaną*	'grow'	grow	*crecer*
**standaną*	'stand'	stand	
**brēaną*	'smell'	smell	*oler*
**fēaną*	'blame'		
**grētaną*	'weep'	cry	*llorar*
**hwētaną*	'show/run into'	show	*asomar*
**knēaną*	'recognize/know'	know	*conocer*
**lēaną*	'rebuke'		
**lētaną*	'leave/let'	leave, let	*dejar*, *salir*
**nēaną*	'sew'	sew	*coser*
**rēdaną*	'advise'		*explicar*
**sēaną*	'sow'		
**tēkaną*	'touch'	touch	*tocar*
**wēaną*	'blow (wind)'	blow	
**blēsaną*	'blow'	blow	*soplar*
**slēpaną*	'sleep'	sleep	*dormir*
**blōaną*	'bloom'		
**blōtaną*	'worship'	bless	
**flōaną*	'flow'		
**flōkaną*	'clap/beat'	beat, snap	
**hwōpaną*	'boast'		
**hwōsaną*	'cough'		
**rōaną*	'row'		
**spōaną*	'succeed'	win	*ganar*
**aikaną*	'acknowledge'	mind	*notar*, *conocer*, *fijar*
**aistaną*	'respect'		
**flaih...*	'speak friendly'		
**fraisaną*	'try'	try	*intentar*
**haitaną*	'call/name/command'	call, tell, name	*llamar*, *alar*

Table 37: Section 3.2 Proto-Germanic comparisons (continued)

Proto-Germanic	**Translation**	**English**	**Spanish**
**laikaną*	'play'	play	*jugar*
**maitaną*	'cut off'	cut	*cortar*, *partir*, , *quitar*
**skaiþaną*	'separate'	cut, snap	*separar*, *cortar*, *partir*
**taisaną*	'pluck'		*arrancar*
**aukaną*	'increase'	grow	
**ausaną*	'draw (water)'		
**hlaupaną*	'leap'	jump, hop	*saltar*
**stautaną*	'knock/shove'	knock, hit, push, bump	*tumbar*, *chocar*
**falþaną*	'fold'	fold, bend	
**haldaną*	'keep'	keep, hold	*cuidar*, *tener*
**saltaną*	'salt'		
**spaldaną*	'split'	cut, crack	*partir*
**staldaną*	'possess'	have, hold, keep	*tener*
**waldaną*	'control'		*conducir*
**walkaną*	'rock/full (cloth)'	rock, clean	*lavar*, *limpiar*
**blandaną*	'mix'	mix	*trepar*
**fanhaną*	'seize'	take, steal	*agarrar*, *coger*
**ganganą*	'go'	go, walk	*ir*
**hanhaną*	'hang'	hang	*colgar*
**bnūaną*	'rub'		
**hrōpijaną*	'cry out/call'	cry, call	*alar*, *gritar*
**wōpijaną*	'cry out'	cry, call	*alar*, *gritar*
**arjaną*	'plow'		
**bringaną*	'bring'	bring	*sacar*, *acercar*
**brūkaną*	'use'	use	*usar*
**būaną*	'dwell'	live	*vivir*

Table 38: Section 3.2 Proto-Germanic comparisons (continued)

Latin	Translation	English	Spanish
edō[1]	'eat'	eat	*comer*
dicō	'say, mention, talk, declare'	say	*decir*
faciō	'do, make'	do	*hacer*
uideō	'see look, observe, seem'	see	*ver, entender*
habeō	'have, hold, own'	have	*tener*
dō	'give, offer, render, yield'	give	*dar*
uolō	'wish, want, mean, intend'	want	*querer*
ferō	'bear, carry, support'	bring	*llevar*
ueniō	'come, approach'	come	*venir*
agō	'do, act, accomplish, drive'	do	*hacer, cumplir*
fīō	'happen, become'	happen	*pasar, suceder*
scrībō	'write'	write	*escribir*
putō	'clean, arrange, ponder'	clean, think, peel	*limpiar, pensar, ordenar*
sciō	'can, know, understand'	know	*poder, saber, entender*
uocō	'call, summon, name'	call	*llamar*
mittō	'send, release'	send	*echar*
legō	'choose gather, read'	pick, read	*elegir*
dēbeō	'have, owe'	have, need	*necesitar*
iubeō	'command, order'		*mandar*
quaerō	'seek, ask, lack, want'	ask, want	*preguntar, pedir, faltar*
crēdō	'believe, trust, confide'	believe	*creer*
capiō	'capture, seize, take'	take	*tomar*
audiō	'hear, listen, attend, obey'	hear	*oir, escuchar*
petō	'ask, seek, desire, attack'	ask	*preguntar, pedir, querer*
accipiō	'receive, accept'	get	*coger*
teneō	'hold, have, grasp'	have, hold	*tener, sujetar*
sinō	'let, permit, suffer, put, lay'	let, put	*dejar*
neō	'spin, weave, entwine'		
pōnō	'place, lay, ordain, pitch'	put	*colocar, meter*
sequor	'follow, go'	follow	*seguir, ir*
reor	'reckon, think'	think	*pensar*
dūcō	'lead, guide, draw, pull'	pull	*conducir*
alō	'feed, maintain, develop'	feed	
nōscō	'know, recognize'	know	*saber, conocer*
suō	'sew, stich, join'	sew, glue, stick	*atar, empatar*
moueō	'move, disturb, shake'	move	*mover*
pariō	'beget'		
relinquō	'abandon, leave'	leave	*dejar*
uincō	'win, conquer'	win	*ganar, suceder*
serō	'sow, found'		*crear*

Table 39: Section 3.2 Latin comparisons sorted by Latin token frequency rank

[1]The rank for *edō* is over-estimated because some of its perfect forms are homophonous with 'to be'

Latin	Translation	English	Spanish
referō	'bear, bring, drive'	bring, drive	*sacar, acercar, regresar*
soleō	'tend, accustomed'		*soler*
uīuō	'live, survive, reside'	live	*vivir*
amō	'love, like, enjoy'	like	*amar, gustar*
for	'speak, talk, say'	talk	*habar*
stō	'stand, stay, remain'	stand	*quedar*
appellō	'address, call'	call	*llamar*
cadō	'fall, die, cease, happen'	fall	*caer, morir*
patior	'suffer, endure'		*durar*
parō	'prepare, arrange'		*preparar, ordenar*
reddō	'return, restore, provide'	give, share	*dar, regalar, hechar*
inueniō	'find, discover'	find	*econtrar*
aiō	'say, assert, affirm'	say	*decir*
gerō	'carry, wear, possess, wage'	carry, wear	*llevar, traer, vestir*
loquor	'say, speak, tell, talk'	say	*decir, hablar*
morior	'die, decay'	die	*morir*
intellegō	'understand, comprehend'	understand	*entender*
trādō	'deliver, transmit, surrender'	give	*dar, contar*
tueor	'look, watch, guard, protect'	keep, look	*mirar, guardar*
ualeō	'strong, be well, be worth'		
negō	'deny, refuse, reject, prevent'		
doceō	'teach, show'	teach	*enseñar, asomar*
lābor	'slip, glide, flow'	slip, slide	
cōlō	'till, inhabit, protect, nuture'	care	*salvar, servir*
fugiō	'flee, escape, speed, hasten'	run	*largar, escapar, correr*
nāscor	'born, arise, proceed, grow'	grow	*nacer*
rogō	'ask, enquire, request'	ask	*preguntar, pedir*
efficiō	'make, effect, complete'	make	*hacer, cumplir*
fallō	'deceive, trick, cheat'	trick	
cōgō	'collect, assemble, gather'		*juntar*
sapiō	'taste, discern, skill'	taste	
coepiō	'begin, commence, initiate'	begin	*empezar*
timeō	'fear, be afraid'	frighten	
cōgnōscō	'learn, know, recognize'	learn	*aprendar*
addō	'place, lay bring, add'	put, lay, bring	*juntar, colocar*
laudō	'praise'		
liceō	'fetch, have value'		
praestō	'stand, excel'	stand, win	*suceder*
sūmō	'take, assume, seize, begin'	take, begin	*tomar, empezar*
cūrō	'arrange, attend, care'	care	*juntar, cuidar, ordenar*
rapiō	'rob, abduct, rape'	take, steal	*agarrar, robar*
seruō	'keep, protect, guard, save'	keep, save	*salvar, cuidar*
ōrō	'orate, plead, beg'	say, ask	*predir*
cōnstō	'stand, agree fit'	stand, write, fit	*caber*
afferō	'carry, bring, conduct'	carry, bring	*conducir*
audeō	'dare, venture, risk'		*intentar*

Table 40: Section 3.2 Latin comparisons (continued)

Latin	Translation	English	Spanish
trahō	'drag, trail'	pull	*traer*
regō	'rule, govern, guide, manage'		*arreglar, mandar*
iūdicō	'judge, decide, condemn'		
nesciō	'not know'		
respondeō	'reply, answer, be present'		*contestar, aparecer*
mereō	'earn, merit, obtain'	win, get	*ganar*
recipiō	'take, receive, accept'	take, get	*tomar*
placeō	'please, satisfy'	like	*entendar, contentar*
probō	'approve, test, prove'	try	*probar*
tollō	'raise, lift, remove, destroy'	lift	*levantar, romper*
orior	'rise, appear'		*aparecer*
armō	'furnish, mobilize, provoke'		
conueniō	'convene, meet, fit'	meet, fit	*juntar, caber, atacar*
cernō	'separate, discern, perceive'	understand	*partir*
maneō	'stay, endure'	wait	*durar*
exīstimō	'think, estimate, consider'	think	*pensar*
mūtō	'remove, move, transform'	change	*mover, cambiar*
cōnsulō	'consult, think, regard'	ask, think	*pensar*
cēnseō	'think, judge, count'	think, count	*pensar*
cēdō	'go, move, proceed, happen'	move, happen	*ir, mover*
terō	'rub, tread'		
uītō	'avoid, evade, shun'		
taceō	'silent, omit'		
misceō	'mix, confuse'	mix	
accēdō	'approach, reach, agree, enter'	come, reach	*entrar*
contineō	'hold, keep, contain'	hold, keep	*cerrar, encerrar*
mīrō	'wonder, marvel'	wonder	
recingō	'loosen, undo, refasten'		*soltar*
soluō	'loosen, release, explain'		*explicar, soltar*
intrō	'enter'		*entrar*
tegō	'cover, clothe, protect'	cover	*guardar, tapar*
spectō	'watch, observe'	watch	*mirar*
pāreō	'appear, obey, submit'		*cumplir*
iaceō	'lie, linger, stop'	stop, lay	*parar*
turbō	'disturb, unsettle, upset'	bother	*molestar*
caedō	'cut, fell, strike, defeat'	cut, hit	*pegar, tumbar*
ostendō	'expose, exhibit, show'	show	*asomar*
uetō	'forbid, oppose, veto'		
premō	'press, pursue'	chase	*pisar*
perdō	'destroy, ruin, lose'	break	*romper*
prōficiō	'progress, benefit, help'	reach, help	*ayudar*
locō	'place, set, arrange, lend'	put	*colocar, meter*
cupiō	'desire'	want	*querer*
exerceō	'work, harass, oversee'	work, bother	*vigilar, trabajar*
currō	'run, hurry'	run	*correr*
reperiō	'find, learn, realize'	find, learn	*aprender*
praecipiō	'take, command, inform'	take, tell	*mandar*

Table 41: Section 3.2 Latin comparisons (continued)

Latin	Translation	English	Spanish
arbitror	'witness, believe, think,'	see, think, believe	*creer*
excipiō	'rescue, receive, understand'	take, understand	*entender*
fundō	'pour, shed, scatter, wet'	pour	*aguar*
pugnō	'fight, contend, struggle'	fight	*pelear*
sedeō	'sit, remain, sink, stay'	sit	*sentar*
iungō	'join'	stick, glue	*atar, empatar*
dēfendō	'defeat, guard, protect'	lose	*guardar*
mētior	'measure, estimate, distribute'	guess, share	
aperiō	'uncover, open'	open	*abrir*
gignō	'beget, produce, cause'		
caueō	'beware, avoid, prevent'		
generō	'beget, produce, descend'		*bajar*
gaudeō	'rejoice'		
indō	'put, set, insert, introduce'	put	*colocar, meter*
frangō	'break, shatter'	break	*romper*
dīligō	'esteem, love'	love	*amar*
cōnstituō	'establish, confirm, decide'		
perueniō	'come, arrive, reach'	come, reach	*venir*
adiciō	'throw, fling'	throw	*lanzar*
cōgitō	'think, consider, ponder'	think	*pensar*
cōnferō	'bring, collect, unite, join'	bring, carry, glue	*pisar*
dubitō	'doubt, ponder, consider'	wonder	*sentir*
nōminō	'name, call, nominate'	name, call	*llamar*
pateō	'open, expose'	open	*abrir*
dōnō	'give, grant, forgive'	give	*dar*
lateō	'hide, retire'	hide	*esconder*
bibō	'drink'	drink	*beber*
noceō	'injure, hurt, damage'	hurt, break	*dañar*
contingō	'touch, contact, reach, attain'	touch	*tocar*
nō	'swim, float'	swim	*nadar*
spērō	'hope, expect, anticipate'	hope	*esperar*
certō	'fight, wrestle, contend, compete'	fight	*pelear*
concēdō	'depart, retire, withdraw'	leave	*dejar*
tangō	'touch, grasp, reach, arrive, attain'	touch	*tocar*
incipiō	'begin'	begin	*empezar*
discō	'learn, study, practice'	learn	*aprender*
prōdō	'give, put, bear, report, record'	give, put, bring	*dar, creer*
inpōnō	'place, set, lay, establish'	put, lay	*colocar, meter*
auferō	'take, carry, remove, withdraw'	take, carry	*partir*
ēueniō	'happen, occur'	happen	*pasar*
careō	'lack'		*faltar*
pertineō	'extend, reach, matter'	pull, belong	
incīdō	'fall, drop, attack'	fall, drop	*caer, ataca*
occupō	'occupy, fill, seize'	fill, take	
augeō	'increase, spread, expand'		
pellō	'push, drive, hurl, strike'	push	*lanzar, empujar*

Table 42: Section 3.2 Latin comparisons (continued)

Latin	Translation	English	Spanish
mandō	'order, commit'		*mandar*
cieō	'move, act, shake, summon'	move, shake, stir	*mover, hacer*
fruor	'enjoy, engage'	like	*encantar*
canō	'sing, recite, sound'	sing	*cantar, gritar*
damnō	'disapprove, reject, punish'		*castigar*
perficiō	'complete, perfect, accomplish'	finish	*cumplir*
uersō	'turn, whirl'	turn	*girar*
errō	'wander, rove, err'	bumble	
metuō	'fear, afraid'		
labōrō	'work, strive, suffer'	work, bother	*trabajar*
āmittō	'lose'	lose	
optō	'select, desire'	pick, want	*elegir, querer*
queror	'complain, lament'		
dīuidō	'divide, separate, distribute'	break, snap, share	*partir*
sustineō	'hold, keep, support, sustain'	hold, keep	*sujetar*
contendō	'hurry, stretch, fight, demand, ask'		*pelear, preguntar*
iuuō	'help, save, gratify'	help, save	*ayudar*
ignōscō	'forgive'		
prōmittō	'send, promise'	send	*echar*
nouō	'renew alter, change'	change	*cambiar*
aspiciō	'look, regard, observe, notice'	look	*notar*
accidō	'fall, descend, happen, occur'	fall	*caer, bajar*
cōmō	'bring, form, care, dress'	do, bring	*hacer, vestir*
sonō	'resound, speak, call, cry'	sound, yell, call	*gritar*
lībō	'taste, sip, sprinkle, spill'	taste, spill	
aduenīō	'arrive'		*llegar*
numerō	'count, pay, reckon'	count, pay	*contar*
uereor	'revere, fear'		
dēcernō	'decide, settle, determine'		
exigō	'drive out, expel, demand'		*mandar*
praebeō	'provide, grant, service, show'		
dēserō	'leave, desert, abandon'	leave	*dejar*
postulō	'demand, ask, desire, need'	ask, want, need	*querer, pedir*
flōreō	'bloom, prosper'		
appāreō	'appear, be visible, serve'		*servir*
līberō	'free, release, absolve'		*soltar*
ārdeō	'burn'	burn	*quemar*
uertō	'turn, revolve, exchange'	turn, share	*girar*
īnstituō	'establish, found, build'	build	
circō	'traverse, wander'	bumble, cross	
spargō	'scatter, strew, sprinkle'		
dēcoquō	'boil, diminish, repress, concoct'	bake, cook, make	*cocinar*
dūrō	'harden, last, endure'		*durar*
imperō	'command, govern, demand'		*mandar*
iūrō	'vow, swear'		
ruō	'hurry, rush, collapse, fail, fall'	hurry	*fallar*

Table 43: Section 3.2 Latin comparisons (continued)

Latin	Translation	English	Spanish
occīdō	'fall, set, perish, die, ruin'	fall, die, break	*caer, morir*
emō	'buy, purchase, acquire'	buy, shop	*comprar*
serpō	'creep, crawl'		
fleō	'weep, grieve'	cry	*llorar*
oleō	'smell'	smell	*oler*
prōpōnō	'set forth, declare'		
uulgō	'broadcast, publish, issue'		
cōnsequor	'follow, accompany, reach'	follow, reach	*seguir, accompañar*
ōrnō	'equip, prepare, adorn'		
committō	'commit, begin'	begin	*empezar*
exspectō	'wait, expect'	wait	*esperar*
pleō	'fill, fulfill'	fill, take	*llenar, contentar*
uerrō	'sweep, brush, clean'	wipe, brush	*limpiar*
uoluō	'roll, tumble'	roll	*rodar*
ēripiō	'rescue, snatch, escape'	take, steal	*salvar*
rumpō	'break, tear, split'	snap, saw, break	*partir*
patrō	'accomplish'	finish, win	*cumplir*
afficiō	'handle, attack, affect'		
moneō	'warn, advise, remind'		
differō	'carry, spread, distract'	carry	*partir*
sacrō	'declare, dedicate, worship'	say, bless	
discēdō	'leave, depart'	leave	*dejar*
indicō	'indicate, point, show, reveal'	point	*puntar*
lūdō	'play, practice, amuse, tease'	play	*jugar*
fateor	'admit, own, show, indicate'	show, point	*notar*
adhibeō	'extend, call, invite'		
lēuō	'elevate, lighten, relieve'	lift	*levantar*
dēsīderō	'want, wish, miss, lack'	want	*faltar*
dīmittō	'send, dismiss'	send	*echar*
mentior	'lie, deceive, pretend'		
dīuertō	'separate, divert, visit'	visit	*partir*
rēgnō	'rule, govern'		
signō	'mark, sign, seal, stamp'		*grabar*
plācō	'appease, pacify'		
expōnō	'place, put, lay'	put, lay	*colocar, meter*

Table 44: Section 3.2 Latin comparisons (continued)

A.2. Appendix to Chapter 5

Verbs sorted by Levin (1993) classification. Directional-to are marked with an asterisk. Bold are PPCME2-only. Italic are PPCEME-only.

- TRANSFER OF MESSAGE

  answer, confirm, preach, read, show, teach, **leren** 'teach2,' tell, write*, **scriben*** 'write2'

- GIVE

  give*, pay*, *render*, serve, sell*, lend*

- THROWING

  cast*, *pass*, *throw*

- FUTURE HAVING

  *advance**, assign*, bequeath*, *extend**, grant*, bitaken* 'grant2,' **lenen*** 'grant3,' **unnen*** 'grant4,' leave, offer*, **beoren*** 'offer2,' owe, promise, **biheten** 'promise2,' *win*, yield*

- SAY

  admit, *communicate*, *confess*, *mention*, *propose*, *repeat*, *report*, *reveal*, say*, speak*

- CARRY

  *carry**, *pull**

- MANNER OF SPEAKING

  cry, sing, *utter*

- PUTTING SPECIFIED DIR

  adjoin, *affix**, fasten, **teiten** 'fasten3,' join, lift*, **reren*** 'lift2,' raise*, **hewen*** 'raise2'

- LATINATE

  *address**, *administer*, *ascribe*, *attribute**, *compare*, *confine*, *convey**, **couple**, deliver*, direct*, *disclose*, *dispatch**, *distribute**, *explain*, *expose*, *express*, *forfeit*, minister, proffer*, *recommend*, *reduce*, *refer*, *relate*, restore, *return**, *sacrifice*, signify, *subject*, submit*, translate*,

- DUB

  fatten 'anoint2,' call, **clepen** 'call2,' make, name

- SEND

send*

- DO ONLY

  ask, bear, beget, *cost.* deny, forbid, forgive, *save*, *spare*, *vouchsafe*, *wish*

- BRING/TAKE

  bring*, **nimen*** ‘nimen,’ take*, **underfongen*** ‘take2’

- APPOINT

  appoint, *allow*, **condempne** ‘judge2,’ ordain, *permit*, **want**

- FULFILLING

  *commit**, convert, entrust*, *gain*, *impart**, present*, *provide**, *serve*

- BILL

  afford, *tender spare*

- DRIVE

  *chase**, *drive**, *draw** ‘drive2’

- DECLARE

  assume, declare, find, **think**, *warrant*

BIBLIOGRAPHY

T. A. Åfarli and E. H. Fjøsne. Weak dative case in norwegian dialect syntax. *Studia Linguistica*, 66(2):75–93, 2012.

J. Aguado-Orea and J. M. Pine. Comparing different models of the development of verb inflection in early child Spanish. *PLOS One*, 10(3):e0119613, 2015. doi: 10.1371/journal.pone.0119613.

J. Aitchison. *Language Change: Progress or Decay?* Cambridge University Press, 2001.

C. Allen et al. The development of a new passive in English. In *Time over Matter: Diachronic Perspectives on Morphosyntax*. CSLI Publications, 2001.

C. L. Allen. *Case marking and reanalysis: Grammatical relations from Old to Early Modern English*. Clarendon Press, 1995.

C. L. Allen. Case syncretism and word order change. In A. van Kemenade and L. Bettelou, editors, *The handbook of the history of English*, pages 201–223. Blackwell, Oxford, 2006.

J. H. Allen and J. B. Greenough. New latin grammar. 1903.

B. Ambridge, J. M. Pine, C. F. Rowland, and C. R. Young. The effect of verb semantic class and verb frequency (entrenchment) on childrenâĂŹs and adultsâĂŹ graded judgements of argument-structure overgeneralization errors. *Cognition*, 106(1):87–129, 2008.

B. Ambridge, J. M. Pine, C. F. Rowland, and F. Chang. The roles of verb semantics, entrenchment, and morphophonology in the retreat from dative argument-structure overgeneralization errors. *Language*, 88(1):45–81, 2012.

H. Andersen. Abductive and deductive change. *Language*, pages 765–793, 1973. doi: 10.2307/412063.

J. M. Anderson. *Structural aspects of language change*, volume 13. Longman, 1973.

S. R. Anderson. On mechanisms by which languages become ergative. *Mechanisms of syntactic change*, 2:317–364, 1977.

J. M. Anglin. Vocabulary development: A morphological analysis. *Monographs of the society for research in child development*, 50:v–186, 1993.

M. Aronoff. *Word formation in generative grammar*. MIT Press, Cambridge, MA, 1976. doi: 10.2307/412595.

M. Aronoff. *Morphology by itself: Stems and inflectional classes*. Number 22. MIT press, 1994.

S. Arunachalam. Argument structure: Relationships between theory and acquisition. In *Cognitive science perspectives on verb representation and processing*, pages 259–280. Springer, 2015.

S. Arunachalam and S. R. Waxman. LetâĂŹs see a boy and a balloon: Argument labels and syntactic frame in verb learning. *Language Acquisition*, 22(2):117–131, 2015.

H. Baayen. On frequency, transparency and productivity. In *Yearbook of morphology 1992*, pages 181–208. Springer, Dordrecht, 1993. doi: 10.1007/978-94-017-3710-4_7.

R. H. Baayen and A. Renouf. Chronicling the times: Productive lexical innovations in an english newspaper. *Language*, pages 69–96, 1996.

H. A. Bacovcin. *Parameterising Germanic Ditransitive Variation*. 2017.

A. Baker, D. Archangeli, and J. Mielke. Variability in american english s-retraction suggests a solution to the actuation problem. *Language variation and change*, 23(3):347–374, 2011.

C. L. Baker. Syntactic theory and the projection problem. *Linguistic Inquiry*, 10(4):533–581, 1979.

C. L. Baker. Learning and cognition: The acquisition of argument structure. *Language*, 68 (2):402–413, 1992.

M. C. Baker. Incorporation: A theory of grammatical function changing. 1988.

A. Bammesberger. *Der Aufbau des Germanischen Verbalsystems*, volume 1. C. Winter, Universitätsverlag, 1986.

A. Bammesberger. *Dehnstufe und Reduplikation im UrGermanischen*. Universitaetsverlag Winter, 1994.

A. Bammesberger. The preterite of Germanic strong verbs in classes four and five. *NOWELE. North-Western European Language Evolution*, 27(1):33–43, 1996.

N. S. Baron. *Language acquisition and historical change*. North-Holland Publishing Company, Amsterdam, 1977.

M. Baroni. 39 distributions in text. *Corpus Linguistics: An International Handbook Volume*, 2:803–822, 2005.

J. Barðdal, K. E. Kristoffersen, and A. Sveen. West scandinavian ditransitives as a family of constructions: With a special attention to the norwegian "v-refl-np'construction". *Linguistics*, 49(1):53–104, 2011.

K. Bech and G. Walkden. English is (still) a west germanic language. *Nordic Journal of Linguistics*, 39(1):65–100, 2016.

H. Behrens. The input–output relationship in first language acquisition. *Language and cognitive processes*, 21(1-3):2–24, 2006.

M. D. Berger. New York City and the ante-bellum South: The maritime connection. *Word*, 31(1):47–53, 1980.

J. Berko. The child's learning of English morphology. *Word*, 14(2-3):150–177, 1958. doi: 10.1080/00437956.1958.11659661.

K. Berkson, S. Davis, and A. Strickler. What does incipient /ay/-raising look like?: A response to Josef Fruehwald. *Language*, 93(3):e181–e191, 2017.

R. C. Berwick, R. C. Berwick, and R. S. Berwick. *The acquisition of syntactic knowledge*, volume 16. MIT press, 1985.

T. Biberauer and G. Walkden. *Syntax over time: lexical, morphological, and information-structural interactions*, volume 15. Oxford Studies in Diachronic a, 2015.

A. Biggs. A new case for a-movement in northwest british English. In *West Coast Conference on Formal Linguistics (WCCFL)*, volume 32, pages 218–227, 2015.

J. D. Bobaljik and H. Thráinsson. Two heads arenâĂŹt always better than one. *Syntax*, 1 (1):37–71, 1998.

M. Bowerman. *Early syntactic development: A cross-linguistic study with special reference to Finnish*, volume 11. CUP Archive, 1973.

M. Bowerman. How do children avoid constructing an overly general grammar in the absence of feedback about what is not a sentence?. *Papers and Reports on Child Language Development*, 22:23–35, 1983.

M. Bowerman and W. Croft. The acquisition of the English causative alternation. *Crosslinguistic perspectives on argument structure: Implications for learnability*, pages 279–307, 2008.

M. D. S. Braine and P. J. Brooks. Verb argument structure and the problem of avoiding an overgeneral grammar. *Beyond names for things: Young children's acquisition of verbs*, page 353, 1995.

M. R. Brent and J. M. Siskind. The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2):33–44, 2001. doi: 10.1016/s0010-0277(01)00122-6.

R. Brown. *A first language: The early stages.* Harvard University Press, Cambridge, MA, 1973. doi: 10.4159/harvard.9780674732469.

R. Brown and U. Bellugi. Three processes in the child's acquisition of syntax. *Harvard educational review*, 34(2):133–151, 1964.

B. Bruening. Double object constructions disguised as prepositional datives. *Linguistic Inquiry*, 41(2):287–305, 2010.

K. Brugmann. Zu den ablautverhältnissen der sogenannten starken verba des Germanischen. *IndoGermanische Forschungen*, 32:179, 1913.

J. Bybee. Regular morphology and the lexicon. *Language and cognitive processes*, 10(5): 425–455, 1995.

J. Bybee. *Language, usage and cognition*. Cambridge University Press, 2010.

J. L. Bybee. *Morphology: A study of the relation between meaning and form*. John Benjamins, Amsterdam, 1985. doi: 10.1075/tsl.9.

T. Cameron-Faulkner, E. Lieven, and M. Tomasello. A construction based analysis of child directed speech. *Cognitive Science*, 27(6):843–873, 2003.

L. Campbell. *Historical linguistics*. Edinburgh University Press, 2008.

J. K. Chambers. Canadian raising. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 18(2):113–135, 1973.

E. Chan. *Structures and distributions in morphological learning*. PhD thesis, University of Pennsylvania, Philadelphia, PA, 2008.

N. Chomsky. *Syntactic structures*. Walter de Gruyter, 1958.

N. Chomsky. A review of bf skinner's verbal behavior. *Language*, 35(1):26–58, 1959.

N. Chomsky. Aspects of the theory of syntax. page 3, 1965.

N. Chomsky. *The logical structure of linguistic theory*. Springer, 1975.

N. Chomsky. On cognitive structures and their development: A reply to piaget. *Philosophy of mind: Classical problems/contemporary issues*, pages 751–755, 1980.

N. Chomsky. *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group, 1986.

N. Chomsky. Three factors in language design. *Linguistic inquiry*, 36(1):1–22, 2005.

N. Chomsky and M. Halle. The sound pattern of english. 1968.

J. Clackson and G. Horrocks. *The Blackwell history of the Latin language*. John Wiley & Sons, 2011.

H. Clahsen. The representation of participles in the German mental lexicon: evidence for the dual-mechanism model. In *Yearbook of morphology 1996*, pages 73–95. Springer, 1997.

E. Conwell and K. Demuth. Early syntactic productivity: Evidence from dative shift. *Cognition*, 103(2):163–179, 2007.

A. Cournane. In defense of the child innovator. In E. Mathieu and R. Truswell, editors, *Micro Change and Macro Change in Diachronic Syntax*, pages 10–24. Oxford University Press, Oxford, 2017. doi: 10.1093/oso/9780198747840.003.0002.

W. Croft. *Explaining language change: An evolutionary approach*. Pearson Education, London, 2000.

H. Cuyckens and M. Verspoor. *On the road to to*. LINCOM Europa, Munich, 1998.

E. Dąbrowska. Learning a morphological system without a default: The polish genitive. *Journal of child language*, 28(3):545–574, 2001.

Ö. Dahl. *The growth and maintenance of linguistic complexity*. John Benjamins, 2004.

Ö. Dahl. *Testing the assumption of complexity invariance: The case of Elfdalian and Swedish*. na, 2009.

J. Dailey-O'Cain. Canadian raising in a midwestern us city. *Language Variation and Change*, 9(1):107–120, 1997.

P. S. Dale and L. Fenson. Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28(1):125–127, 1996.

M. Davies. The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(2):159–190, 2009. doi: 10.1075/ijcl.14.2.02dav.

L. De Cuypere. The Old English to-dative construction. *English Language & Linguistics*, 19(1):1–26, 2015.

V. Demonte. Dative alternation in spanish. *Probus*, 7(1):5–30, 1995.

H. Diessel. Frequency effects in language acquisition, language use, and diachronic change. *New ideas in psychology*, 25(2):108–127, 2007.

H. Diessel. Language change and language acquisition. In A. Bergs and L. J. Brinton, editors, *Historical linguistics of English: An international handbook*, volume 2, pages 1599–1613. De Gruyter Mouton, Berlin, 2012. doi: 10.1515/9783110251609.1599.

B. E. Dresher and A. Lahiri. Romance loanwords and stress shift in English: A quantitative approach, 2015.

P. Durkin. *Borrowed words: A history of loanwords in English*. Oxford University Press, 2014.

P. Eckert. *Jocks and burnouts: Social categories and identity in the high school*. Teachers college press, 1989.

J. Elman. Generalization, simple recurrent networks, and the emergence of structure. In *Proceedings of the twentieth annual conference of the Cognitive Science Society*, pages 543–548, Mahwah, NJ, 1998. Lawrence Erlbaum.

J. Elter. The rise of the *to*-dative: a language-contact approach to a phenomenon of structural language change. Master's thesis, Universität Mannheim, 2018.

D. Embick. Features, syntax, and categories in the latin perfect. *Linguistic inquiry*, 31(2): 185–230, 2000.

J. E. Emonds and J. T. Faarlund. English: The language of the vikings. 2014.

C. Fellbaum. Examining the constraints on the benefactive alternation by using the world wide web as a corpus. *Linguistic evidence: Empirical, theoretical and computational perspectives*, pages 209–240, 2005.

S. Finley. Cognitive and linguistic biases in morphology learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 9(5):e1467, 2018. doi: 10.1002/wcs.1467.

J. D. Fodor. Why learn lexical rules. In *10th Annual Boston University Conference on Language Development*, 1985.

L. Friedman. *The St. Louis corridor: Mixing, competing, and retreating dialects*. PhD thesis, University of Pennsylvania, 2014.

J. Fruehwald. *The phonological influence on phonetic change*. PhD thesis, University of Pennsylvania, 2013.

J. Fruehwald. The early influence of phonology on a phonetic change. *Language*, 92(2): 376–410, 2016.

J. Fruehwald. Response to Berkson, Davis, & Strickler, 'What does incipient /ay/-raising look like?'. *Language*, 93(3):e192–e197, 2017.

J. Fruehwald, J. Gress-Wright, and J. Wallenberg. Phonological rule change: The constant rate effect. In *Proceedings of NELS*, volume 40, 2009.

P. Garbacz. *Word Order in Övdalian: A Study in Variation and Change.*, volume 70. Centre for Languages and Literature, Lund University, 2010.

P. G. Glare. *Oxford latin dictionary*. Oxford univ. press, 2012.

A. E. Goldberg. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press, Chicago, IL, 1995.

J. C. Goodman, P. S. Dale, and P. Li. Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35(3):515–531, 2008. doi: 10.1017/s0305000907008641.

K. Gorman and C. Yang. When nobody wins. In *Competition in Inflection and Word-Formation*, pages 169–193. Springer, 2019.

J. Grimshaw. *Argument structure.* the MIT Press, 1990.

J. Gropen, S. Pinker, M. Hollander, R. Goldberg, and R. Wilson. The learnability and acquisition of the dative alternation in English. *Language*, pages 203–257, 1989. doi: 10.2307/415332.

G. R. Guy and S. Boyd. The development of a morphological class. *Language variation and change*, 2(1):1–18, 1990.

M. Hale. Diachronic syntax. *Syntax*, 1(1):1–18, 1998.

W. S. Hall, W. E. Nagy, and R. L. Linn. *Spoken words, effects of situation and social group on oral word usage and frequency.* Lawrence Erlbaum, 1984.

M. Halle. Phonology in generative grammar. *Word*, 18(1-3):54–72, 1962. doi: 10.1080/00437956.1962.11659765.

C.-h. Han, J. Lidz, and J. Musolino. V-raising and grammar competition in korean: Evidence from negation and quantifier scope. *Linguistic Inquiry*, 38(1):1–47, 2007.

H. Harley. If you have, you can give. In *Proceedings of the 15th West Coast Conference on Formal Linguistics*, pages 193–207. CSLI Publications Stanford, CA, 1997.

H. Harley. Possession and the double object construction. *Linguistic variation yearbook*, 2 (1):31–70, 2002.

B. Hart and T. R. Risley. *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing, Baltimore, MD, 1995. doi: 10.5860/choice.33-3435.

B. Hart and T. R. Risley. The early catastrophe: The 30 million word gap by age 3. *American Educator*, 27(1):4–9, 2003.

B. Heine, H. Bernd, T. Kuteva, et al. *World lexicon of grammaticalization.* Cambridge University Press, 2002.

K. Hess Zimmermann. *El desarrollo lingüístico en los años escolares: análisis de narraciones infantiles (Linguistic development during the school years: Analysis of child narrations).* PhD thesis, El Colegio de México, Mexico City, 2003.

C. Heycock and J. Wallenberg. How variational acquisition drives syntactic change: The loss of verb movement in Scandinavian. *Journal of Comparative Germanic Linguistics*, 16 (127), 2013.

H. Hirt. Handbuch des Urgermanischen. t. 1, Laut- und Akzentlehre (Handbook of protogermanic 1, phonology and accent). *Indogermanische Bibliothek. 1. Abt., 1. Reihe, Sammlung Indogermanischer Lehr- und Handbücher. Grammatiken*, 1931.

H. H. Hock. *Principles of historical linguistics*. Walter de Gruyter, Berlin, 1991.

H. H. Hock. Analogical change. *The handbook of historical linguistics*, pages 441–460, 2003.

D. Howes. Zipf's law and miller's random-monkey model. *The American Journal of Psychology*, 81(2):269–272, 1968.

C. L. Hudson Kam and E. L. Newport. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*, 1(2):151–195, 2005.

R. Jackendoff. On larson's treatment of the double object construction. *Linguistic inquiry*, 21(3):427–456, 1990.

R. Jackendoff. *Semantic structures*, volume 18. The MIT Press, 1992.

L. D. Jarmulowicz. English derivational suffix frequency and children's stress judgments. *Brain and Language*, 81(1-3):192–204, 2002. doi: 10.1006/brln.2001.2517.

F. Jelinek. *Statistical methods for speech recognition*. MIT press, 1997.

J. S. Johnson and E. L. Newport. 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, 1989.

M. Joos. A phonological dilemma in Canadian English. *Language*, 18:141–144, 1942.

C. L. H. Kam and E. L. Newport. Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, 59(1):30–66, 2009.

P. Kiparsky. *Phonological change*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1965.

P. Kiparsky. 'Elsewhere' in phonology. In S. R. Anderson and P. Kiparsky, editors, *A festschrift for Morris Halle*, pages 93–106. New York: Holt, Rinehart & Winston, 1973.

C. Kirov and R. Cotterell. Recurrent neural networks in linguistic theory: Revisiting Pinker and Prince (1988) and the past tense debate. *Transactions of the Association for Computational Linguistics*, 6:651–665, 2018. doi: 10.1162/tacl_a_00247.

E.-S. Ko. Acquisition of vowel duration in children speaking American English. In *Eighth Annual Conference of the International Speech Communication Association*, 2007.

J. Kodner. Estimating child linguistic experience from historical corpora. *Glossa: A Journal of General Linguistics*, 4(1):122, 2019. doi: http://doi.org/10.5334/gjgl.926.

J. Kodner and C. Richter. Transparent /ai/-raising as a contact phenomenon. In R. Purse and Y. White, editors, *Penn Working Papers in Linguistics: Selected Papers from NWAV*, volume 25, pages 61–70, 2020.

W. Koopman. The order of dative and accusative objects in old English. *Ms., University of Amsterdam*, 1994.

F. Kortlandt. The germanic fifth class of strong verbs. *NOWELE. North-Western European Language Evolution*, 1992.

A. Kroch. Morphosyntactic variation. In *Proceedings of the thirtieth annual meeting of the Chicago Linguistic Society*, volume 2, pages 180–201, Chicago, IL, 1994. Chicago Linguistics Society.

A. Kroch. Modeling language change and language acquisition. In *Expansion of an LSA Institute forum lecture*, 2005.

A. Kroch and A. Taylor. The penn-helsinki parsed corpus of middle english (ppcme2). department of linguistics, university of pennsylvania. cd-rom. *Department of Linguistics, University of Pennsylvania, CD-ROM*, second edition, release 4 ((http://www.ling.upenn.edu/ppche-release-2016/PPCME2-RELEASE-4), 2000.

A. Kroch, B. Santorini, and L. Delfs. The penn-helsinki parsed corpus of early modern english (ppceme). department of linguistics, university of pennsylvania. cd-rom. *Department of Linguistics, University of Pennsylvania, CD-ROM*, first edition, release 3 (http://www.ling.upenn.edu/ppche-release-2016/PPCEME-RELEASE-3), 2004.

A. S. Kroch. Reflexes of grammar in patterns of language change. *Language variation and change*, 1(3):199–244, 1989.

A. S. Kroch. Syntactic change. In M. Baltin and C. Collins, editors, *The Handbook of Contemporary Syntactic Theory*, pages 699–729. Blackwell, Hoboken, NJ, 2001. doi: 10.1002/9780470756416.ch22.

P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044):606–608, 1992.

J. Kuryłowicz. La nature des procès dits «analogiques». *Acta linguistica*, 5(1):15–37, 1945.

J. Kuryłowicz and M. Mayrhofer. *Indogermanische Grammatik*, volume 2. C. Winter, 1968.

W. Labov. The social motivation of a sound change. *Word*, 19(3):273–309, 1963.

W. Labov. Some principles of linguistic methodology. *Language in society*, 1(1):97–120, 1972. doi: 10.1017/s0047404500006576.

W. Labov. Building on empirical foundations. *Perspectives on historical linguistics*, 24:17, 1982.

W. Labov. 12. the overestimation of functionalism. *Functionalism in linguistics*, 20:311, 1987.

W. Labov. The child as linguistic historian. *Language Variation and Change*, 1(1):85–97, 1989. doi: 10.1017/s0954394500000120.

W. Labov. *Principles of linguistic change, volume 1: Internal factors*. John Wiley & Sons, 1994.

W. Labov. *Principles of linguistic change, volume 2: Social factors*. John Wiley & Sons, 2001.

W. Labov. Transmission and diffusion. *Language*, 83(2):344–387, 2007.

W. Labov and I. Rosenfelder. The Philadelphia neighborhood corpus. *Philadelphia: University of Pennsylvania.*, 2011. Online: http://fave.ling.upenn.edu/pnc.html.

W. Labov, M. Yaeger, and R. Steiner. *A quantitative study of sound change in progress*, volume 1. US Regional Survey, 1972.

W. Labov, S. Ash, and C. Boberg. *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter, 2005.

R. K. Larson. On the double object construction. *Linguistic inquiry*, 19(3):335–391, 1988.

R. Laurent. *Past participles from Latin to Romance*, volume 133. Univ of California Press, 1999.

J. A. Legate and C. Yang. Assessing child and adult grammar. *Rich languages from poor inputs*, pages 168–182, 2013.

W. P. Lehmann. *Historical linguistics: an introduction*. Routledge, 2013.

B. Levin. *English verb classes and alternations: A preliminary investigation*. University of Chicago Press, Chicago, IL, 1993.

B. Levin. Dative verbs a crosslinguistic perspective. *Lingvisticæ Investigationes*, 31(2): 285–312, 2008.

H. Levin and M. Novak. Frequencies of Latinate and Germanic words in English as determinants of formality. *Discourse Processes*, 14(3):389–398, 1991. doi: 10.1080/01638539109544792.

H. Levin, S. Long, and C. A. Schaffer. The formality of the Latinate lexicon in English. *Language and Speech*, 24(2):161–171, 1981. doi: 10.1177/002383098102400204.

J. D. Lewis and J. L. Elman. Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In *Proceedings of the 26th Annual Conference on Language Development*, 2001.

P. Li, X. Zhao, and B. Mac Whinney. Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31(4):581–612, 2007.

R. Lieber. *On the organization of the lexicon*. PhD thesis, Massachusetts Institute of Technology, 1980.

D. Lightfoot. *The development of language: Acquisition, change, and evolution*. Wiley-Blackwell, Hoboken, NJ, 1999.

D. W. Lightfoot. *Principles of diachronic syntax*. Cambridge University Press, Cambridge, 1979.

C. Lignos and C. Yang. Morphology and language acquisition. *Cambridge handbook of morphology*, pages 765–791, 2018.

C. Lignos, E. Chan, C. Yang, and M. P. Marcus. Evidence for a morphological acquisition model from development data. In *Proceedings of the 34th Annual Boston University Conference on Language Development*, volume 2, pages 269–280, 2010.

M. Loporcaro. *Vowel length from Latin to Romance*, volume 10. OUP Oxford, 2015.

B. Lundquist. Double object constructions: active verbs. *Nordic Atlas of Linguistic Structures (NALS)*, 2014.

L. MacKenzie and G. Bailey. Our dialects: Mapping variation in English in the uk, 2016. URL `http://projects.alc.manchester.ac.uk/ukdialectmaps/syntactic-variation/give-it-me/`.

B. MacWhinney. The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, pages 1–123, 1978. doi: 10.2307/1166047.

B. MacWhinney. *The CHILDES language project: Tools for analyzing talk*. Lawrence Erlbaum, Mahwah, NJ, 1991.

B. MacWhinney. *The CHILDES project: The database*, volume 2. Psychology Press, Abingdon-on-Thames, 2000. doi: 10.1162/coli.2000.26.4.657.

R. Mailhammer. *The Germanic strong verbs: Foundations and development of a new system*, volume 183. Walter de Gruyter, 2007.

J. M. Manly. From generation to generation. In N. Bøgholm, A. Brusendorff, and C. Bodelson, editors, *A Grammatical Miscellany Offered to otto Jespersen on His 70th Birthday*, pages 287–289. Levin & Munksgaard, Copenhagen, 1930.

M. Maratsos. More overregularizations after all: new data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen & Xu. *Journal of Child Language*, 27(1):183–212, 2000. doi: 10.1017/s0305000999004067.

P. H. Matthews and P. Matthews. *Inflectional morphology: A theoretical study based on aspects of Latin verb conjugation*, volume 6. CUP Archive, 1972.

K. Matzel. Zum system der starken verben des Germanischen. *Anjali, Papers on Indology and Buddhism: A Felicitation Volume Presented to Oliver Hector de Alwis Wijesekera on his Sixtieth Birthday*, pages 173–181, 1970.

I. Mazurkewich and L. White. The acquisition of the dative alternation: Unlearning overgeneralizations. *Cognition*, 16(3):261–283, 1984.

T. McFadden. The rise of the to-dative in middle english. *Syntactic effects of morphological change*, pages 107–123, 2002.

J. McWhorter. What happened to english? *Diachronica*, 19(2):217–272, 2002.

W. Meid. *Das germanische Praeteritum*, volume 3. Druck H. Kowatsch, 1971.

J. M. Meisel. Bilingual language acquisition and theories of diachronic change: Bilingualism as cause and effect of grammatical change. *Bilingualism: Language and Cognition*, 14(2): 121–145, 2011. doi: 10.1017/s1366728910000143.

I. A. Mel'čuk. *Towards a language of linguistics: A system of formal notions for theoretical morphology*, volume 44. Fink, 1982.

D. Michelioudakis. The evolution of inherent case in the diachrony of greek. *Syntax over time. Lexical, morphological, and information–structural interactions*, pages 197âĂŞ–216, 2015.

J. Milroy and L. Milroy. Linguistic change, social network and speaker innovation. *Journal of linguistics*, 21(2):339–384, 1985.

T. H. Mintz. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1):91–117, 2003.

B. Mitchell. Old english grammar i and ii, 1985.

E. Moreton and E. R. Thomas. Origins of Canadian Raising in voiceless-coda effects: a case study in phonologization. *Laboratory phonology*, 9:37–64, 2007.

K.-H. Mottausch. Das präteritum der 4. und 5. starken verbklassen im Germanischen. *NOWELE. North-Western European Language Evolution*, 36(1):45–58, 2000.

W. S. Murray and K. I. Forster. Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, 111(3):721–756, 2004.

W. E. Nagy and R. C. Anderson. How many words are there in printed school English? *Reading Research Quarterly*, pages 304–330, 1984. doi: 10.2307/747823.

L. R. Naigles. Form is easy, meaning is hard: Resolving a paradox in early child language. *Cognition*, 86(2):157–199, 2002.

A. Nardy, J.-P. Chevrot, and S. Barbu. Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study. *Language Variation and Change*, 26 (3):273–301, 2014.

P. Nation and R. Waring. Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, 14:6–19, 1997.

E. Newport, H. Gleitman, and L. Gleitman. Mother, id rather do it myself: Some effects and non-effects of maternal speech style. 1977.

E. L. Newport. Children and adults as language learners: Rules, variation, and maturational change. *Topics in cognitive science*, 2019.

E. L. Newport. Children and adults as language learners: Rules, variation, and maturational change. *Topics in Cognitive Science*, 12(1):153–169, 2020.

J. e. a. Nivre. Universal dependencies 2.3, 2018. URL `http://hdl.handle.net/11234/1-2895`. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

P. Niyogi and R. C. Berwick. A language learning model for finite parameter spaces. *Cognition*, 61(1):161–193, 1996. doi: 10.1016/s0010-0277(96)00718-4.

P. Niyogi and R. C. Berwick. A dynamical systems model for language change. *Complex Systems*, 11(3):161–204, 1997.

P. Niyogi and R. C. Berwick. The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences*, 106(25):10124–10129, 2009.

J. Nycz. New contrast acquisition: methodological issues and theoretical implications. *English Language & Linguistics*, 17(2):325–357, 2013.

R. T. Oehrle. *The grammatical status of the English dative alternation.* PhD thesis, Massachusetts Institute of Technology, 1976.

J. Ohala, C. S. Masek, R. A. Hendrick, and M. F. Miller. The listener as a source of sound change. *parasession on language and behavior (Chicago Linguistics Society, Chicago 1981)*, 1981.

H. Paul. *Prinzipien der sprachgeschichte*. Tubingen. Niemeyer, 1880.

L. Pearl and J. Sprouse. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, 20(1):23–68, 2013.

J. E. Pegg and J. F. Werker. Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America*, 102(6):3742–3753, 1997.

A. Perfors, J. B. Tenenbaum, and T. Regier. The learnability of abstract syntactic principles. *Cognition*, 118(3):306–338, 2011.

D. M. Pesetsky. *Zero syntax: Experiencers and cascades*. Number 27. MIT press, 1996.

M. Peters. *Untersuchungen zur Vertretung der indoGermanischen Laryngale im Griechischen*. Verlag d. Österr. Akad. d. Wiss., 1980.

J. B. Pierrehumbert. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, 46(2-3):115–154, 2003a.

J. B. Pierrehumbert. On frequency, transparency and productivity. In *Probabilistic Linguistics*, pages 177–228. MIT Press, Cambridge, MA, 2003b.

S. Pinker. Learnability and cognition: The acquisition of argument structure, 1989.

S. Pinker and A. Prince. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1-2):73–193, 1988. doi: 10.1016/0010-0277(88)90032-7.

S. Pinker and A. Prince. Regular and irregular morphology and the psychological status of rules of grammar. *The reality of linguistic rules*, 321:51, 1994. doi: 10.1075/slcs.26.21pin.

S. Pinker and M. T. Ullman. The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11):456–463, 2002. doi: 10.1016/s1364-6613(02)01990-3.

S. Pinker, D. S. Lebeaux, and L. A. Frost. Productivity and constraints in the acquisition of the passive. *Cognition*, 26(3):195–267, 1987. doi: 10.1016/s0010-0277(87)80001-x.

S. Pintzuk. *Phrase structures in competition: Variation and change in Old English word order*. 1991.

C. Polo. Double objects and morphological triggers for syntactic case. *Syntactic effects of morphological change*, pages 124–142, 2002.

J. W. Poultney. The distribution of the nasal-infixing presents. *Language*, 13(3):163–176, 1937.

E. Prokosch. *A comparative Germanic grammar*. Linguistic Society of America, University of Pennsylvania, 1939.

H. Ragnarsdóttir, H. G. Simonsen, and K. Plunkett. The acquisition of past tense morphology in Icelandic and Norwegian children: An experimental study. *Journal of Child Language*, 26(3):577–618, 1999. doi: 10.1017/s0305000999003918.

M. Ramscar. The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, 45(1):45–94, 2002. doi: 10.1016/s0010-0285(02)00001-4.

M. Rappaport Hovav and B. Levin. The English dative alternation: The case for verb sensitivity. *Journal of Linguistics*, 44(01):129–167, 2008. doi: 10.1017/s0022226707004975.

R. Rimac and B. L. Smith. Acoustic characteristics of flap productions by American English-speaking children and adults: implications concerning the development of speech motor control. *Journal of Phonetics*, 12(4):387–396, 1984.

D. Ringe. *From Proto-Indo-European to Proto-Germanic*, volume 1. Oxford University Press, Oxford, 2017.

J. Roberts and W. Labov. Learning to talk philadelphian: Acquisition of short a by preschool children. *Language variation and change*, 7(1):101–112, 1995.

J. L. Roberts. Acquisition of variable rules:(-t, d) deletion and (ing) production in preschool children. 1994.

B. W. Rohrbacher. *The Germanic VO languages and the full paradigm: a theory of V to I raising*. PhD thesis, University of Massachusetts, 1994.

S. Romero, A. Santos, and D. Pellicer. *The construction of communicative competence in Mexican Spanish speaking children (6 months to 7 years)*. University of the Americas, Mexico City, 1992.

D. E. Rumelhart and J. L. McClelland. *On learning the past tenses of English verbs*. MIT Press, Cambridge, MA, 1986.

D. Sankoff. Variable rules. *Sociolinguistics: An international handbook of the science of language and society*, 2:984–997, 1988.

G. Sankoff and H. Blondeau. Language change across the lifespan:/r/in montreal french. *Language*, 83(3):560–588, 2007.

I. M. Schlesinger. Production of utterances and language acquisition. *The ontogenesis of grammar*, pages 63–101, 1971.

K. Schuler, C. Yang, and E. Newport. Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. In *The 38th Cognitive Society Annual Meeting*, pages 2321–26, Mahwah, NJ, 2016. Lawrence Erlbaum.

K. D. Schuler. *The acquisition of productive rules in child and adult language learners*. PhD thesis, Georgetown University, 2017.

S. Schumacher. Eine alte crux, eine neue hypothese: gotisch iddja, altenglisch ēode. *Die Sprache*, 40(2):179–201, 1998.

S. Schumacher. Langvokalische perfekta. *indoGermanischen Einzelsprachen und ihr grundsprachlicher Hintergrund. In Gerhard Meiser & Olav Hackstein (eds.), Sprachkontakt und Sprachwandel. Akten der XI. Fachtagung der IndoGermanischen Gesellschaft, Halle an der Saale*, 17(23):591–626, 2000.

E. Seebold. *Vergleichendes und etymologisches Wörterbuch der germanischen starken Verben (Comparative and Etymological dictionary of the Germanic strong verbs)*. Walter de Gruyter, Berlin, 1970. doi: 10.1515/9783110821956.

G. Seiler. *Präpositionale Dativmarkierung im Oberdeutschen*. Number 124. Franz Steiner Verlag, 2003.

J. L. Singleton and E. L. Newport. When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive psychology*, 49(4):370–407, 2004.

D. A. Smith, J. A. Rydberg-Cox, and G. R. Crane. The Perseus Project: A digital library for the humanities. *Literary and Linguistic Computing*, 15(1):15–25, 2000. doi: 10.1093/llc/15.1.15.

B. Sneller, J. Fruehwald, and C. Yang. Using the Tolerance Principle to predict phonological change. *Language Variation and Change*, 31(1):1–20, 2019.

J. Y. Song, M. Sundara, and K. Demuth. Phonological constraints on childrenâĂŹs production of English third person singular–s. *Journal of Speech, Language, and Hearing Research*, 52(3):623–642, 2009.

J. N. Stanford. Language acquisition and language change. *The Routledge Handbook of Historical Linguistics*, page 466, 2015.

D. Steriade. The morphome vs. similarity-based syncretism. *The morphome debate*, 2016.

W. Streitberg. *Urgermanische Grammatik*, volume 1. na, 1896.

E. H. Sturtevant. An introduction to linguistic science. 1947.

J. Sverdrup. Der aorist im Germanischen verbalsystem und die bildung des starken präteritums. *Festschrift til Hjalmar Falk. Oslo: Aschehoug*, pages 269–330, 1927.

H. Sweet. *The practical study of languages*. London: Dent, 1899.

G. Szagun, C. Steinbrink, M. Franik, and B. Stumper. Development of vocabulary and grammar in young German-speaking children assessed with a German language development inventory. *First Language*, 26(3):259–280, 2006. doi: 10.1177/0142723706056475.

A. Taylor. Variation in past tense formation in the history of English. *University of Pennsylvania working papers in linguistics*, 1(1):10, 1994.

A. Taylor. The york—toronto—helsinki parsed corpus of old english prose. In *Creating and Digitizing Language Corpora*, pages 196–227. Springer, 2007.

H. Thráinsson. *The syntax of Icelandic*. Cambridge University Press, Cambridge, 2007.

P. M. Tiersma. *Frisian reference grammar*. Number 644. Foris Pubns USA, 1985.

M. Tomasello. *Constructing a language*. Harvard University Press, 2009.

C. Trips and A. Stein. Was old french-able borrowable? *English Historical Linguistics 2006: Lexical and semantic change*, 296:217, 2008.

A. Tyler and W. Nagy. The acquisition of English derivational morphology. *Journal of Memory and Language*, 28(6):649–667, 1989. doi: 10.1016/0749-596x(89)90002-8.

E. van Gelderen. *The linguistic cycle: Language change and the language faculty*. Oxford University Press, 2011.

T. J. Vance. " canadian raising" in some dialects of the northern united states. *American Speech*, 62(3):195–210, 1987.

F. T. Visser. *An historical syntax of the English language*. Brill Archive, 1963.

G. Walkden. The many faces of uniformitarianism in linguistics. *Glossa: A Journal of General Linguistics*, 4(1), 2019. doi: 10.5334/gjgl.888.

G. L. Walkden. The actuation problem, 2017.

J. C. Wallenberg, A. K. Ingason, E. F. Sigurðsson, and E. Rögnvaldsson. Icelandic Parsed Historical Corpus (IcePaHC) version 0.9, 2011. URL `http://linguist.is/icelandic\_treebank`.

U. Weinreich, W. Labov, and M. Herzog. Empirical foundations for a theory of language change. In W. Lehmann and Y. Malkiel, editors, *Directions for historical linguistics*, pages 95–195. University of Texas Press, 1968.

M. L. Weiss. *Outline of historical and comparative grammar of Latin*. Beech Stave Press, 2009.

J. F. Werker and R. C. Tees. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63, 1984.

E. Wonnacott. Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, 65(1):1–14, 2011.

F. Xu and S. Pinker. Weird past tense forms. *Journal of Child Language*, 22(3):531–556, 1995. doi: 10.1017/s0305000900009946.

C. Yang. *Knowledge and learning in natural language*. Oxford University Press, Oxford, 2002.

C. Yang. Who's afraid of George Kingsley Zipf? or: Do children and chimps have language? *Significance*, 10(6):29–34, 2013.

C. Yang. *The Price of Linguistic Productivity*. MIT Press, Cambridge, MA, 2016. doi: 10.7551/mitpress/9780262035323.001.0001.

G. K. Zipf. *Human behavior and the principle of least effort: An introduction to human ecology*. 1949.

K. Zuraw. Probability in language change. *Probabilistic linguistics*, pages 139–176, 2003.