LAISSEZ-FAIRE ANALOGICAL CHANGE[*]

JORDAN KODNER
STONY BROOK UNIVERSITY
DEPARTMENT OF LINGUISTICS AND INSTITUTE
FOR ADVANCED COMPUTATIONAL SCIENCE

ABSTRACT

A wide range of proposals have been put forth to account for the many tendencies of analogical change as well as the typological trends that they induce on morphological systems. These range from sensitivity to token frequency to the management of complexity, and many of them may play some role. However, their relative contributions are hard to differentiate, since they do all fit the data and are also correlated with one another. A well-defined baseline to compare against would help to evaluate the range of proposed accounts. The Poverty of the Stimulus itself suggests one such baseline. The evidence for parts of a morphological system may be so sparse in the input that the language faculty, no matter how well-endowed, might not steer all learners towards compatible grammars. This article shows that this input sparsity by itself can account for many observed correlations between frequency, paradigm size, and irregularity in morphological systems prior to the involvement of other factors. The often severe sparsity of morphological input is quantified in terms of saturation and measured in child-directed, adult, and historical corpora. Population-level simulations of linguistic transmission and change confirm the intuitions drawn from the corpora: sparsity in the input drives analogical change in consistent directions prior to the influence of any internal factors.

1 INTRODUCTION

The Poverty of the Stimulus, that is, the insufficiency of early linguistic input in uniquely specifying a grammar, is one of the most important arguments for an innate language faculty (Chomsky 1959, 1980). As it relates to

change, critically informative patterns in the input are sometimes so sparse or altogether absent that even a well-endowed human language faculty are not always enough to ensure that all learners land on the same grammar. These instances of "abject poverty" of the stimulus provide one mechanism for acquisition-driven language change. Language change is a population-level phenomenon, the sum total of innovations in the internal grammars of individuals. If abject poverty drives learners to different grammars in any way, the result is a change in the distribution of grammars in the population. Furthermore, if abject poverty biases learners in a particular direction, learners' innovations may accrue to yield more dramatic changes in that direction.

Morphology in particular presents an opportunity for the study of abject poverty and change because it often manifests finite pieces which are easily quantifiable and easily discoverable in corpora. The size of a given syntactic category's inflectional paradigm is generally fixed in a given language, and inflected forms, stems, and inflectional categories can be extracted from morphologically annotated corpora. *Analogical leveling*, the replacement of some forms with the effect of removing distinctions (see Hock (2003) for a summary of types of analogy and definitions), can often be explained straightforwardly in terms of the input, since it is well known that infrequent and irregular forms are more likely to be analogized away than high-frequency and regular forms. High frequency items are more likely to be but are not strictly more irregular than low frequency items (Bybee 1985, Michel, Shen, Aiden, Veres, Gray, Google Books Team, Pickett, Hoiberg, Clancy, Norvig et al. 2011, Fratini, Acha & Laka 2014). High frequency inflectional categories are more likely to have irregular forms and are more likely to be the basis for analogy rather than analogized away, and large paradigms tend to be more regular than smaller paradigms, for example, Turkish verb paradigm are larger than German verb paradigms, but they include many fewer irregular forms (Mańczak 1980, Bybee 1985, Hock 1991, 2003, Chapman & Skousen 2005, Ackerman & Malouf 2013).

The mechanisms by which frequency drives the tendencies of analogical change are still debated. Though previous work acknowledges a role for simple attestation, explanations lean in the direction of internal cognitive factors, appealing to token frequency directly or its *entrenchment* as memory imprints (Mańczak 1980, Bybee 1985, Tiersma 1982, Pinker 1995, Bybee & Thompson 1997, Diessel 2007), frequency as a proxy for reliability in productivity (Albright 2008), phonological neighborhood density or network effects interacting both directly and indirectly with token frequency (Bybee & Moder 1983, Hare & Elman 1995, Pinker 1995, Chapman & Skousen 2005, Blevins, Milin & Ramscar 2017, Frank, Smith & Cuskley 2020), and the maintenance of contrast

or balance of contrast and efficient reuse measured directly or in information theoretic terms (Kiparsky 1968, Vennemann 1972, Ackerman & Malouf 2013, Blevins et al. 2017), among others. While experimental evidence shows that humans are sensitive to frequency, phonological neighborhood density, and related measures (see Goldrick & Rapp (2007) and Vitevitch & Luce (2016) for surveys), this does not necessarily demonstrate that these are primary drivers of analogical change. Even if we were to grant that all that all the above factors play some role, there would still be an explanatory oversight: how do we know which of these factors, which all fit the data well and are highly-correlated with one another, is the *cause* of analogical leveling, and which proposals just appear to be driving factors because they are correlated with that cause? Unlike these cognitive proposals, which combine theoretical and often indirect experimental evidence, input sparsity is a directly measurable empirical fact. It exists prior to theories of morphology or psycholinguistic processing. How far does input sparsity go in accounting for the typological trends independent of additional internal factors? The real contribution of the language faculty and other proposed internal factors can only truly be assessed once such a baseline has been established. Given the consistently skewed nature of sparsity across language, that baseline is likely quite biased, and likely contributes substantially to the observed trends of analogy.

The outline of this article is as follows. Section 2 discusses abject poverty at different levels of linguistic representation. *Paradigm saturation* (PS) and *inflectional category saturation* (ICS) are introduced as two easily quantifiable measures of morphological sparsity in corpora, both of which can be seen as creating a biased baseline that may drive learners towards the observed typological trends of analogy. These measures are highly correlated with token frequency, adding a further confound into corpus work which argues for a link between token frequency and analogical change. Section 3 moves past description of biased baselines in the corpora and proposes a mechanism by which sparsity in the early linguistic input may induce the actuation and propagation of changes in the community. Following that, two proof-of-concept simulations are carried out. Modeling only simple learners, the simulations reproduce typologically observed patterns in frequency, paradigm size, and regularity. The learners are "*laissez-faire*" because they do not adopt an active strategy to maintain the paradigm or balance trade offs in complexity or even track token frequency. They only react consistently to their input.

These results have two primary implications which discussed are in Section 4. First, sparsity in early linguistic input is not innocuous or neutral with respect to change. It introduces directional biases which lead learners to modify their grammar in particular ways. Learners do not need to impose addi-

tional internal pressures to achieve these broad effects. Second, quantitative correlational evidence, while very important, is not sufficient by itself to distinguish competing causes. A good understanding of the baseline case and a mechanism are crucial for working out what competing causes are or are not necessarily responsible for. It is probably true that internal factors play important roles in analogical change – sparse paradigm and inflectional category saturation is shown here to account for the broadest trends – but these saturations are always present in the input and are unavoidable. They are "always on" factors in language acquisition and change. Thus, the impact of other factors has to be described *in addition* to this baseline.

2   Input Sparsity as a Baseline

Language acquisition presents a real challenge no matter how rich the innate faculty of language – Universal Grammar (UG) renders language learning tractable in the face of the Poverty of the Stimulus (Chomsky 1959, 1980), but it does not trivialize it. Input sparsity can be severe, and acquisition takes time. Some linguistic patterns are more challenging to acquire than others,[1] 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 might assume and could play an important role in language change.

It is important to emphasize that abject poverty cannot be reduced to a simple absence of direct evidence. The Poverty of the Stimulus is interesting because children acquire generally consistent grammars despite their under-specified input. One of the most impressive aspects of language acquisition is that children learn what *not* to say in the absence of actionable negative evidence (Brown & Hanlon 1970, Braine 1971, Bowerman 1988, Marcus 1993). English learners must learn island constraints without ever receiving

---

1 For example, English learners consistently acquire verbal *-s* and *-ing* before *-ed* (Brown 1973). Palestinian Arabic learners acquire suffixing 'sound' plurals around the same time as English past is acquired but take several more years to gain complete competence over stem-changing 'broken' plurals (Ravid & Farah 1999). English learners begin over-regularizing *-ed*, a sign that it has been productively acquired, by age 3 (Ervin & Miller 1963, Pinker & Prince 1988), but Spanish learners begin over-regularizing verbal stem alternations before age 2 (Clahsen, Aveledo & Roca 2002). On the other hand, acquirers of Turkish and Swahili acquire most aspects of their elaborate morphological systems by around age 3 as well (Aksu-Koç 1985, Deen 2005). Some aspects of syntax and semantics are among the latest to be acquired, for example, the complexities of English models (Papafragou 1998) and quantifier scope (van Koert, Koeneman, Weerman & Hulk 2015) (see Cournane (2017) for a review) which, like Arabic broken plurals, still exhibit divergences from adult-like performance into late childhood.

information that they are ungrammatical, for example. Nevertheless, English speakers *can* produce parasitic gaps, exceptions to these island constraints, even though they very rarely if ever appear in their input as positive examples (Pearl & Sprouse 2013). Thus parasitic gaps must fall out from the inner workings of the syntax. This is a case of conventional Poverty of the Stimulus, not abject poverty, because the language faculty is apparently sufficient despite the input sparsity.[2] Another related concept that does not necessarily fall under abject poverty is the imperfect learning proposed by Kiparsky (1968) *et seq.*, in which opaque phonological surface forms (perhaps the result of change among adults) drives learners to adopt more transparent surface realizations. In these cases, it is not that there is insufficient evidence *per se*, so much as that no evidence would be sufficient because the opacity, and an alternative preferred hypothesis grammar exist.

Cases of abject poverty are challenging to identify by definition. If the surface expression of the different grammars were clear and easily distinguishable by the linguist, they might be easily distinguishable for learners as well, so they would not be abject. Real cases must not differ in their extensions or differ so rarely that learners are statistically unlikely to receive distinguishing information in their input, yet they must differ in some measurable way or they would be indistinguishable even to researchers. As such, we need specific probes to identify individual cases of abject poverty, and there may be more instances of failed convergence than we can currently observe. Some known examples are worth discussing.

Contrasting with English parasitic gaps, Korean presents an example of abject poverty in morphosyntax (Han, Lidz & Musolino 2007). Because the language is head-final, it is not clear on the surface whether the language has verb-raising – both a raising and a non-raising grammar would produce surface-identical SOV strings. Unambiguous evidence for one grammar or the other can only come from differences in meaning. Scoping of negation and quantified object noun phrases was identified by Han et al. as a probe for distinguishing the two grammars. The authors used experimental data to demonstrate that there are underlyingly two distinct Korean grammars, one with and one without verb raising. Since the distinguishing constructions are quite rare in the input, and these rare instances would have to occur in a dialogue where the learner is set to notice and react to a mis-parse, something that is far from guaranteed (Labov 2011: ch. 2), the input is not enough to drive all learners to the same grammar. Korean verb raising has apparently

---

2 Even if one adopts an alternative analysis where the unacceptability of island violations is due to processing issues and not ungrammaticality (Liu, Winckel, Abeillé, Hemforth & Gibson 2022), this definition still holds. Our general cognitive faculties are apparently sufficient to account for them despite input sparsity.

"fallen through the cracks" between the language faculty and the Korean input.

In English morphology and phonetics, 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 suffix, but also employ some kind of stem change as in *tell ~ told* or *sleep ~ slept*. In principle, these could be decomposed into a root and the past suffix *tol-d*, or their forms might be represented like strong verbs (*sing ~ sang*) and suppletive verbs (*go ~ went*) that just happen to end in coronal obstruents. One way to distinguish between these possible grammars 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 the t/d-deletion rate for semi-weak verbs can provide insight into their (lack of) decomposition. However, Guy & Boyd (1990) find some variation in deletion rates even among adult speakers, consistent with some adults treating the final obstruent as the past suffix and some treating it as part of the stem. Guy & Boyd (1990) did not write in these terms, but this is abject poverty since the weak and indirect statistical signal uncovered by linguists is either not noticed or not usable by learners to ensure a consistent parse of these verbs. Learners would have to track deletion rates for this specific set of verbs in their past forms across contexts and across speakers, and then they would need to leverage this evidence in the reverse direction to set their parses. Rather than learning the community *t/d*-deletion rate for a morphological context, they would need to infer the morphological context from an already acquired span of context-dependent *t/d*-deletion rates.

Morphology actually presents many more systemic examples of abject poverty as well due to low-frequency irregular forms. A truly irregular form is irregular because it is not predictable according to some internalized rule or pattern available in the speaker's grammar. If a truly irregular form is not attested in the early linguistic input, then the learner will not be able to produce it and should usually let the grammar generate a regular form instead. The latter case will result in an over-regularization, one of the more common types of innovation in morphology acquisition (Clahsen & Rothweiler 1993, Xu & Pinker 1995, Maratsos 2000, Yang 2002, Maslen, Theakston, Lieven & Tomasello 2004, Mayol 2007). In some cases accurately quantifiable by the Tolerance Principle introduced in Section 3, there is no regular pattern for the child or adult speaker to fall back on, and they produce nothing. These are paradigmatic gaps (see Gorman & Yang (2019) for a review).

Over-regularization during individual-level acquisition has a direct par-

allel in *leveling*, the most common type of analogical change at the population-level is *leveling*, where infrequent and irregular patterns are replaced by more frequent and regular ones (Kuryłowicz 1945, Mańczak 1980, Hock 1991). The following sections quantify two measures of sparsity in morphological input which can influence the direction of over-regularization and analogy, *paradigm saturation* and *inflectional category saturation*. Following that, Section 3 discusses how over-regularization can sometimes result in analogical change.

2.1 Paradigm Saturation

Morphology lends itself to the quantification of sparsity more so than syntax because there is a finite number of inflected forms for a given root in a given language. However, these forms are not distributed equally in language use. Much has been said about the presence of Zipfian and other sparse long-tailed distributions of roots, inflectional categories, and inflected forms in everything from child-directed speech corpora to large natural language processing data sets (Zipf 1949, Miller 1957, Howes 1968, Baayen 1993, Jelinek 1997, Chan 2008, Yang 2013, Lignos & Yang 2018).

Following a Zipfian distribution, frequencies are proportional to the inverse of their frequency rank. That is, the second most frequent item should be about half as common as the most frequent, the third most frequent item should be proportionately about a third as frequent, and so on. Such distributions are dramatically skewed, with a few frequent items in the "neck" and very many infrequent items in the long thin tail. From the perspective of someone learning their language's morphology, this means that most roots will appear only occasionally in the input, maybe once or twice even in millions of tokens, and these will in turn only appear in one or two of their possible forms. The proportion of a root's licit inflected forms that are actually attested in a corpus is its *paradigm saturation* (PS), defined formally in (1). The "paradigm" in "paradigm saturation" is used descriptively to refer to the set of inflections available for a given root and not a commitment to a particular theoretical stance on the nature of paradigms.

(1) **Paradigm Saturation PS$_C$ (adapted from Chan 2008)**
     For a given corpus C, the proportion of a lemma's possible paradigm P that is attested p_C in a given corpus C:

$$\text{PS}_C = \frac{|p_C|}{|P|}$$

Chan (2008) shows that paradigm saturations also follow long-tailed distributions in corpora of various sizes and genres in thirteen languages. That

is, a few roots appear in many of their possible forms (they have high paradigm saturation), while the majority of them appear in only a tiny fraction of their possible forms (they have low paradigm saturation). Importantly, this is just an empirical description of the data. It is true regardless of prior conceptions of morphological theory or language processing.

The Uniformitarian Principle, as applied to linguistics (Labov 1972, Walkden 2019), predicts that these long-tailed distributions should recur across languages and genres and across time in the absence of time and place-specific factors that would alter them. This assumption holds for paradigm saturation. To illustrate this, Figure 1 and Table 1 show paradigm saturations for verbs from the English Brown child-directed speech (CDS) corpus (Brown 1973), Spanish FernAguado CDS (Fernald & Marchman 2012), and German Leo CDS (Behrens 2006), along with modern adult Finnish, modern adult Turkish, attested adult Gothic, and attested adult Latin from the Universal Dependencies Treebank (UD) (Nivre 2018).[3]

There are a few points of note. First, the distribution becomes sparser as paradigm size increases, which follows given that any given inflectional category accounts for a smaller proportion of productions as paradigm size increases. Modern English verbs only have five morphologically distinct inflectional categories, present, present 3sg, past, present progressive, and a past participle (though some of these are often suppletive), and it has by far the highest average paradigm saturation. Second, these patterns arise in CDS (Chan 2008: ch. 3), which is the "genre" that children primarily learn from, but also in modern adult and historical corpora. These patterns can be measured and used by historical researchers when CDS is unavailable (Kodner 2020a: ch. 3). Table 1 shows paradigm saturation distributions for each language. It is notable that similar patterns emerge for English, German, and Spanish CDS as in adult-directed texts from UD in the same languages. We can reasonable assume that Finnish, Turkish, Gothic, or Latin CDS would

---

3 The set of languages chosen here was subject to a number of constraints. Annotated running text was needed in order to extract both frequency information and morphological information. Such corpora do not yet exist for the overwhelming majority of the world's languages, and even most of the languages present in UD and CHILDES either lack morphological annotations or have very small corpora. UD and CHILDES morphological annotations are both relatively error-prone and had to be semi-manually normalized and corrected to be useable, so the set of available languages was further constrained by the author's familiarity with the annotated languages. Languages that met these constraints were chosen to cover a range of paradigm sizes. This paper focuses primarily on verbal paradigms because most of they are much larger that the nominal paradigms for the (Indo-European) languages used here. For example, English nouns only have two forms, singular and plural, and Spanish have two to four, depending on whether gender is overtly marked. Reported paradigm sizes are the empirical maximum number of unique tags in the cleaned data sets.

pattern similarly to the UD corpora as well if they were available. There is no meaningful difference between ancient Gothic and Latin on one hand and modern English, German, Spanish, Finnish, and Turkish on the other, a demonstration that the uniformitarian assumption is appropriate across time.
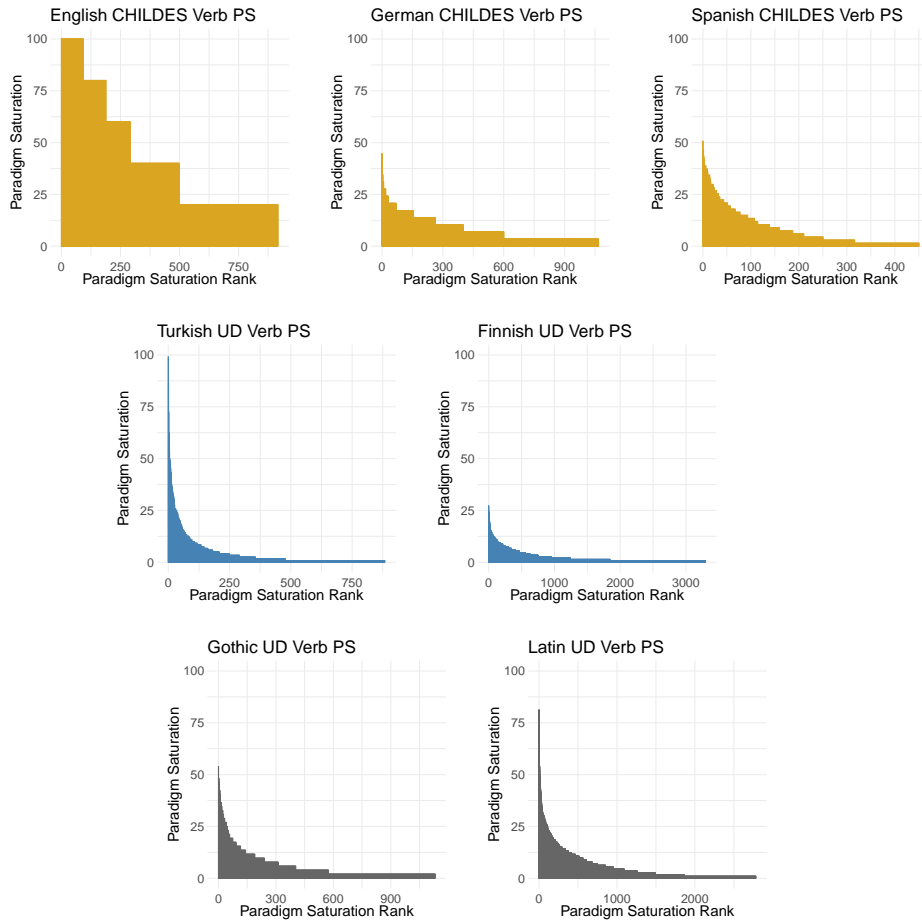
Figure 1 Verb paradigm saturation (PS) plots for English, German, and Spanish CDS, Finnish and Turkish from UD, and Gothic and Latin from UD for historical comparison.

Paradigm saturation has clear implications for learning (Chan 2008, Lignos & Yang 2018): The early linguistic input available to a given child constitutes a corpus. If a form is not attested in the input, it must be inferred by some kind of productive or regular pattern (the Paradigm Cell-Filling Problem Ackerman, Blevins & Malouf 2009), but if a form is truly irregular, this inference will produce an over-regularization. Truly irregular forms must be

Corpus	Language	|Paradigm|	Max PS	Mean PS	Med. PS
CDS	English	5	100%	43.40%	40.0%
CDS	German	29	44.82%	8.48%	6.90%
CDS	Spanish	67	50.75%	8.51%	4.48%
Modern	English	5	100%	48.88%	40.0%
Modern	German	29	79.31%	8.08%	3.45%
Modern	Spanish	67	41.79%	4.94%	2.99%
Modern	Finnish	150	27.33%	2.49%	1.33%
Modern	Turkish	120	99.17%	4.74%	1.67%
Historical	Gothic	52	53.85%	6.18%	3.85%
Historical	Latin	113	81.2%	5.91%	2.65%

Table 1   CDS, modern adult, and historical adult paradigm saturation (PS) statistics. Maximum, mean, and median PS are reported for each language/corpus.

memorized, but to be memorized, they must be attested in the input. Outside of English, the vast majority of stems in the investigated languages have very low paradigm saturation, with a median of only a few attested forms if that, so nearly all of their forms must be inferred by learners. These inferred forms cannot be truly unpredictable or irregular, so the vast majority of forms must be regular. Since learners have many fewer opportunities overall to be exposed to the irregular forms of stems with low paradigm saturation, over many individuals over time, stems with low paradigm saturation have a greater chance to be regularized, while those with high paradigm saturation have a better chance at remaining irregular. Items with high paradigm saturation are more likely to serve as the basis for analogy because they are more likely to be well-attested in the input and form the basis of a learners' evidence for their grammar.

It is important to recognize that paradigm saturation correlates strongly with frequency. This is not particularly surprising, since the more times a stem appears, the more opportunities it has to appear in all of its possible forms: a stem that appears fewer times than it has inflectional categories cannot possibly have full saturation. Table 2 provides Pearson and Spearman correlations for token frequency vs. paradigm saturation directly as well as between token *rank* vs. paradigm saturation *rank*.

A high Spearman's rank correlation coefficient ρ indicates strong monotonic relationships between frequency and paradigm saturation, while a high Pearson's r demonstrates a strong linear relationship between the two. Both

Corpus	Language	Verb			Noun		
		r rank	*r* freq	ρ	*r* rank	*r* freq	ρ
CDS	English	0.833	0.317	0.825	0.354	0.382	0.345
CDS	German	0.870	0.408	0.864	0.849	0.485	0.834
CDS	Spanish	0.900	0.588	0.895	0.647	0.484	0.640
Modern	English	0.921	0.343	0.916	0.671	0.344	0.661
Modern	German	0.909	0.733	0.902	0.928	0.514	0.912
Modern	Spanish	0.950	0.412	0.940	0.719	0.314	0.714
Modern	Finnish	0.960	0.617	0.954	0.904	0.596	0.922
Modern	Turkish	0.981	0.881	0.977	0.891	0.721	0.884
Historical	Gothic	0.953	0.586	0.947	0.910	0.572	0.903
Historical	Latin	0.974	0.515	0.973	0.932	0.395	0.930

Table 2   Pearson's *r* and Spearman's ρ for token rank vs paradigm saturation rank and token frequency vs paradigm saturation. Spearman correlations are identical for rank and frequency.

range from -1 (perfect negative correlation) to 0 (no correlation at all) to 1 (perfect positive correlation). For all languages except for English nouns, *r* for frequency is lower than *r* for rank or ρ. This is expected because token frequencies themselves are Zipfian, which is extremely non-linear.

Overall, all *r* rank and ρ correlations are quite high with the sole exception of English nouns, which only have a paradigm size of two (singular and plural) and do not provide a good measurement. This means that the mere presence of a correlation between frequency or paradigm saturation and regularization is not enough to exhibit a causal role to either. One would have to control for the influence of one before drawing conclusions about the other. Importantly though, explanations based on paradigm saturation and frequency do not have the same burden of evidence. At a minimum, the influence of paradigm saturation is just a logical argument: if a form is not present in the input, it cannot be memorized and thus must be inferred by the learner. If a form is irregular, it cannot be learned unless it can be memorized. Many forms are just not present, so they have to be inferred. In this way, paradigm saturation exists *prior* to the token frequency of inflected forms. The former is a measure of which inflected types are attested at all. The latter can can only come into play once a form is attested. Moreover, the influence of frequency requires additional theoretical or psycholinguistic proposals for how the frequency or measures derived from frequency compels the learner when constructing the morphological system.

Since paradigm saturation exists prior to the cognitive effects of token frequency, it is especially critical to take the former into account before considering the latter. This is why it makes for a strong baseline in the study of the causes of analogical change. A strong effect for paradigm saturation cannot disprove roles for token frequency in cognition, but it does suggest a weaker role than otherwise assumed. That is, the "landscape" of change is not level, because input sparsity itself exerts a substantial directional bias. Paradigm saturation creates an always present baseline bias towards the greater regularization of low-frequency items that exists prior to and independently of other factors. Even "*laissez-faire*" learners not proactively counteracting the effects of sparsity or optimizing their paradigms could introduce directional analogical change into a language.

2.2 Inflectional Category Saturation

The distribution of attested inflectional categories or paradigm slots, which I will refer to here as *inflectional category saturation* (ICS), also follows a highly skewed distribution (Chan 2008: ch. 3), but this is not as extreme as the long-tailed distribution of paradigm saturation. Like paradigm saturation, this pattern is present in both child-directed speech and other historical and modern corpora (Figure 2). Despite genre differences, there are major trends across CDS and other corpora. Compare the most frequent categories in German from CHILDES and UD, along with Gothic for historical comparison in Table 3. There is broad agreement in which categories are most common. For example, the most frequent categories include participles and indicative finite forms rather than subjunctives, though CDS has more second and first person forms reflecting the fact that it is largely dialogue. (2) provides a formula for inflectional category saturation. One key difference from paradigm saturation is that the denominator, the number of lemmas attested in the corpus, is dependent on the corpus sample rather than an inherent property of the language. This is because lemmas are an open class, unlike inflectional categories. To account for this, Figure 2 is scaled by maximum ICS for visualization, while the PS plots are not scaled by maximum PS.

(2) **Inflectional category saturation ICS$_C$**
Let R_C be the lemmas attested in a given corpus C and r_C be the lemmas attested in C in some inflectional category. ICS$_C$ is the proportion of lemmas attested in the inflectional category.

$$\text{ICS}_C = \frac{|r_C|}{|R_C|}$$

ICS Rank	German CHILDES	German UD	Gothic UD
1	**present 3sg**	**past participle**	pres. active participle
2	**past participle 1**	*infinitive*	*preterite act. indic. 3sg*
3	present 1/3pl	*past indic. 3sg*	*pres. act. infinitive*
4	present 2sg	**pres. indic. 3sg**	**pres. act. indic. 3sg**
5	present 1sg	past indic. 3pl	**past passive participle**

Table 3   Two categories (bolded) are shared across all three languages' top five. Two additional categories (italicized) are shared by two. Category labels are adapted from CHILDES and UD coding schemes. The German CHILDES Leo corpus specifies two past participles (1 and 2) and does not distinguish syncretic person/number forms.

Inflectional category saturation establishes a bias towards the leveling of irregulars among low-frequency inflectional categories as well as a trend in analogy from high-frequency to low-frequency categories. For example, it is more likely that a noun would have an irregular nominative singular than an irregular locative plural since the former will typically be much more common. More broadly, it is more likely that a singular would serve as the basis for analogy than a plural except in those cases where the plural happens to be more common, what Tiersma (1982) describes as "local markedness." The same applies to verbs. The third person singular indicative is more likely to take a unique fusional ending than a second person plural subjunctive because the latter will rarely appear in the input and will have to be inferred.

Inflectional category saturation is extremely colinear with the frequency rank of inflectional categories, as shown in Table 4, even more so than the equivalent for paradigm saturation. One again, the mere presence of a correlation between token frequency or rank and analogical processes is not evidence *per se* of a meaningful role for token frequency.
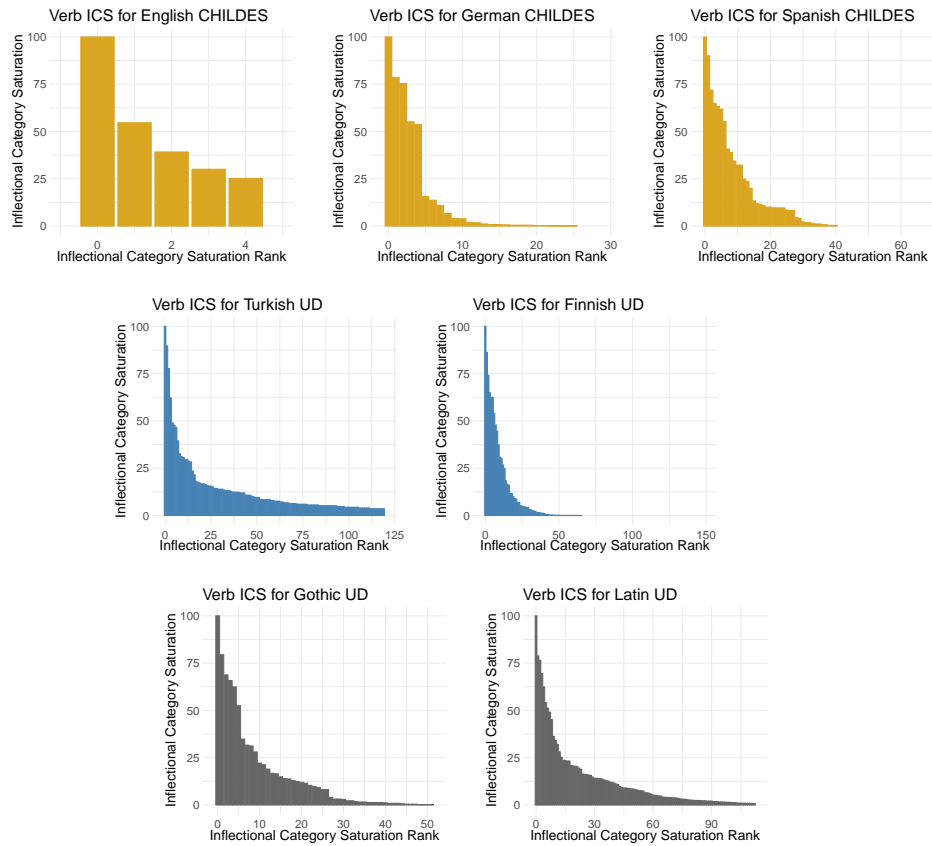
Figure 2 Inflectional category saturation (ICS) plots for English, German, and Spanish CDS, Finnish and Turkish from UD, and Gothic and Latin from UD for historical comparison. Scaled by maximum ICS.

Taken together, paradigm saturation and inflectional category saturation exert baseline biases towards analogical leveling of infrequent items. They cross-cut each other, affecting both lemmas and inflectional categories. Since saturations decrease when paradigm sizes increase, we should expect a trade-off between irregularity and paradigm sizes. Languages which have smaller paradigms can manage more irregularity than larger ones on average. This pattern does seem to hold on the aggregate (Ackerman & Malouf 2013).

Furthermore, observed long-tailed distributions of saturation are unavoidable, and not simply an artifact of small data sets. Further input hardly alleviates the sparseness or skew of the distributions past lengthening the tail for paradigm saturation. This is illustrated by Figure 3, which shows paradigm

Corpus	Language	Verb			Noun		
		r rank	*r* freq	ρ	*r* rank	*r* freq	ρ
CDS	English	1.00	0.960	1.00	0.909	0.999	0.901
CDS	German	0.839	0.872	0.835	0.857	0.997	0.857
CDS	Spanish	0.862	0.805	0.863	0.917	0.961	0.928
Modern	English	0.700	0.722	0.700	1.00	1.00	1.00
Modern	German	0.975	0.947	0.970	0.905	0.991	0.905
Modern	Spanish	0.994	0.916	0.994	1.00	1.00	1.00
Modern	Finnish	0.865	0.908	0.828	0.982	0.990	0.982
Modern	Turkish	0.966	0.930	0.962	0.976	0.991	0.976
Historical	Gothic	0.994	0.891	0.994	0.964	0.972	0.964
Historical	Latin	0.989	0.774	0.989	0.984	0.954	0.984

Table 4 Pearson's *r* and Spearman's ρ for token rank vs ICS rank and token frequency vs ICS. Spearman correlations are identical for rank and frequency.

saturation and inflectional category saturation plots extracted from the first half million tokens of the UD Latin data set in 100,000-token increments. The five-fold increase in corpus data does little to alter the sparsity or skew of the distributions. The insensitivity of these distributions to data size combined with the similarity of distributions across genres including CDS means that we can extrapolate quantitative observations drawn from finite text corpora to the relevant data at hand, the input to children during language acquisition.

3 MECHANISMS

Sparse saturations provide a baseline cause for biased over-regularizations during individual language acquisition. This is a useful observation, but it is not enough to explain analogical change. That requires a further link to population-level change. This section proposes a mechanism through which learners who innovate over-regularizations may actuate analogical change. It will then demonstrate through two proof-of-concept simulations that "*laissez-faire*" learners who receive sparse input and interact in a population can cause biased analogical change.
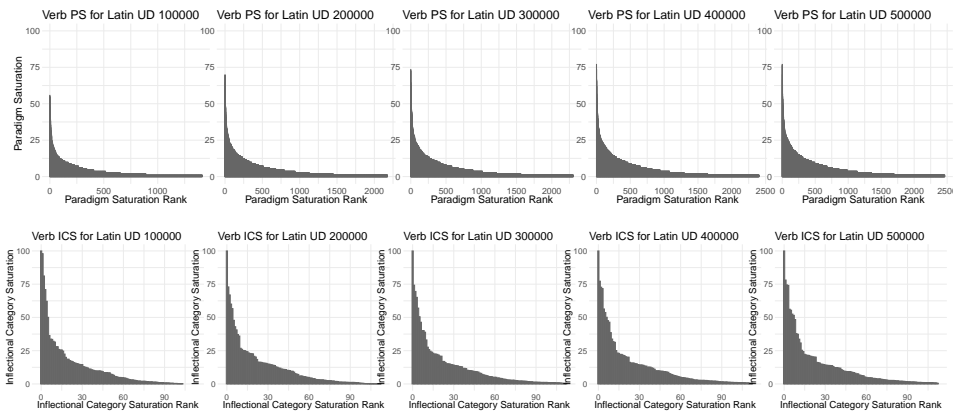
Figure 3 Sparsity and skew of PS (top row) and ICS (bottom row) remain as the corpus size increases from 100,000 to 500,000 for Latin UD.

3.1 "Sibling-Induced" Change

The actuation of a change requires both its innovation and its entry into a speech community (Labov, Yaeger & Steiner 1972: p. 7). If one is to build a theory of input sparsity and actuation, one must propose a mechanism by which individuals' transient sparsity-induced innovations might gain a foothold in the speech community. After all, children are famously accurate language learners. They typically grow out of any innovations that they make during development. How and when would they propagate these innovations instead of growing out of them?

Sparsity and variation together present an explanation. First, if some language-specific pattern is subject to abject poverty, it may not be evidenced for some children at all, who are then forced to infer it with their nascent grammars. The analysis in Section 2 of morphological sparsity shows that input is indeed seriously impoverished. Items in a paradigm of even moderate size will probably not be attested in the vast majority of their forms even after millions of tokens of input. Any item that is only attested once at all can only be attested with just one variant. Children should learn the evidenced variant even if it happens to be the innovative one. Second, children do not actually mature in ideal single input source environments. We know that children receive input from multiple people who themselves may exhibit internal variation in the form of competing grammars (Kroch 1994, Yang 2002) or otherwise.

This is enough to explain change among very rare things, such as the inflectional categories of a large paradigm with the lowest saturation or very rare stems with low saturation, but many other items or phenomena will be attested multiple times and may still change. What might prevent a child who learned an innovative form from growing out of it even if they receive the conservative form? The most important thing to recognize here is that language transmission is not strictly generational (Manly 1930, Weinreich, Labov & Herzog 1968). Children interact verbally with other children across societies, though the interaction setting may vary (Loukatou, Scaff, Demuth, Cristia & Havron 2021). Children are sensitive to their peers' language, especially that of slightly older peers, and they adopt features from them (Labov 1989, Roberts & Labov 1995, Labov 2001, Nardy, Chevrot & Barbu 2014).

The importance 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, one that consistently levels the past participle to the simple past, for example, and she sometimes produces innovative utterances. How might little Bob react to Alice? There are a few of cases to consider:

First, as discussed above, if Bob only hears Alice's innovation and not his parents' conservative production, then he has no way of identifying it as an innovation. He has no reason to doubt Alice in this case, since her language is mostly consistent with the adults' after all, and she communicates with them regularly. If Bob happens to pick up one of Alice's leveled forms, then that is the first step towards actuation of an analogical change. Such changes among low-frequency lexically-specific phenomena should often be explainable by this scenario because they are very sparse and must be directly attested to be learned.

Second, if Bob hears both Alice's innovation and also conservative productions from others, he may recognize the innovation. If he recognizes the innovation, he has three options. First, he could choose to reject it on the basis of Alice's unreliability or even the low frequency of the innovative form relative to the conservative one in his input. Second, he could accept the innovation and reject the conservative variant if he prefers Alice sociolinguistically or the innovation is meaningfully more frequent in his input. Third, he could learn both variants subject to some kind of conditioning by social context, subtle semantic distinction, or any other factor. In this case, the innovative and conservative variants would then be free to pattern like any other sociolinguistic variable. When Bob matures, 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 com-

munity as well as each other granting the innovative forms a foothold in the community. This is actuation.

All three of these scenarios, in which Bob reacts by rejecting Alice's innovation, leveling in favor of the innovation, or learning both the innovative and conservative variant, have been demonstrated to occur experimentally under certain conditions. The conditioning factors are complex and interacting, but generalizations can be made (see Austin, Schuler, Furlong & Newport (2022) for an overview). Typically, children are more likely to learn both variants rather than just one if a clear conditioning factor is present (Hudson Kam & Newport 2005). They are also more likely to level the input if the variant is idiosyncratic rather than clearly associated with community-level variation. (Singleton & Newport 2004, Samara, Smith, Brown & Wonnacott 2017). Sparsity in morphological systems, especially among low-frequency low-saturation items is likely to obscure language-internal or sociolinguistic conditioning factors and so may push learners towards leveling, but this is not always guaranteed by the setting.

This thought experiment is formalized in a framework called "Sibling-Induced" Change (Kodner 2020a), which only requires that learners receive input from multiple individuals, and that the input is sparse. These are both totally normal and empirically observable aspects of native language acquisition. Alice and Bob are siblings in this thought experiment, but the reasoning is meant to apply to any young peer-to-peer interaction. That said, literal sibling-to-sibling transmission of sociolinguistic variants is known to occur. Sankoff & Blondeau (2007) argue in their study of Montreal French /r/ (§7.2) that the first cohorts of speakers to acquire categorical [ʀ] (the innovative variant) acquired it from their older siblings who could produce either the innovative or conservative variant, even though they must have also heard many tokens of their parents' conservative [r].

Sibling-Induced Change can be seen as a clarification or further specification of the Andersen (1973) Z-model of change. Andersen's model depicts change as a cycle where the grammar of one generation generates some set of outputs, and these become the inputs over which the grammar of the next generation is abduced, and so on. To extend this, Sibling-Induced Change first emphasizes that learners are embedded in speech communities that contain many individuals spanning many ages, so transmission is from multiple people to one and repeats over periods much shorter than generations. Second, individuals vary in their productions, both across their lifetimes and across social settings, both for reasons of competence (in the case of competing grammars) and for reasons of performance, so learners receive variable input even from single individuals. Third, acquisition, while impressive in

the face of the Poverty of the Stimulus, is not instantaneous, and in the time it takes to acquire a language, innovating learners have the opportunity to influence others.

The Z of Sibling-Induced Change is visualized in Figure 4. Like the traditional Z-model, the cycle continues on indefinitely. Rather than a single grammar and a single speaker at time 1, a single grammar and a single speaker at time 2, and so on, there are some k speakers at time 1, $1_1, 1_2 \dots 1_k$, each of whom have at least one grammar, so there are $j > k$ grammars $1_1, 1_2 \dots 1_j$ with potentially complex relationships between them at time 1, and so on. The loop indicating transmission that occurs among learners of similar age captures the sociolinguistic findings regarding orientation towards peers. It is this loop that allows for children to influence their peers. The Z with back loops continues *ad infinitum*.
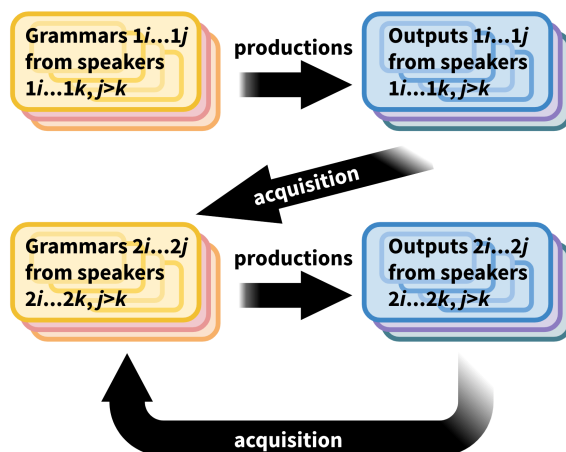
Figure 4   The Z-model extended for "Sibling-Induced" Change. As with the classic Z-model, it continues indefinitely in a chain. Thick arrows indicate bundles of individual arrows, and these may also skip "generations."

3.2   *Simulations*

This section presents a proof-of-concept for *laissez-faire* analogical change, that is, morphological leveling driven by the baseline effect of biased input sparsity before additional factors are taken into account. Two experiments are conducted using the same simulation model.[4] The first (Section 3.2.1) investigates the relationship between attestation and the retention of irregularity, a

4 Code available at https://github.com/jkodner05/jhs-sic

consequence of paradigm saturation. The second (Section 3.2.2) investigates the relationship between paradigm size and the retention of irregularity, a consequence of inflectional category saturation.

As a tool for investigation, simulation has some advantages which complement historical corpus work, sociolinguistic fieldwork, and experimental developmental work. It provides a white box model of both the speaker and the speech community. That is, we can build a "learner" that we know does not track frequency directly in any way, so any observed patterns of change can be blamed squarely on type attestation. This differs from a black box statistical approach where saturation and frequency are highly correlated and must be disentangled. Additionally, the grammars present in the simulated community can be seen to develop in real time without other confounding factors. However, like any other methodology, simulation has some drawbacks. Perhaps the most serious among these is that extra care has to be taken to connect the simulation to the real world systems being modeled. It is technically possible to simulate almost anything one desires, but that anything may not be something in reality. Thus it is important to motivate one's simulation assumptions to the greatest extent possible. One of the goals of this section is to motivate those assumptions.

The intuition for both simulations is as follows: say irregular items have been introduced into a morphological paradigm, perhaps by a sound change. To a first approximation, sound changes proceed in a regular fashion without regards to frequency (Paul 1880, Hoenigswald 1978, Labov 2020), so the new irregulars should be distributed throughout the frequency range rather than concentrated among the high-frequency items. Over time though, low-frequency irregulars will tend to be regularized simply because they are less likely to be attested in individuals' early linguistic input as a consequence of skewed paradigm and inflectional category saturation, and this will cause a strong correlation between frequency and irregularity to emerge. The higher the frequency (lower the frequency rank), the higher the chance of irregularity. Additionally, since larger paradigms are necessarily sparser, this effect should be stronger when the paradigm is larger. Furthermore, since this is no cognitive or learning *requirement* for the correlations to hold, it is possible for cases of irregularity to (temporarily) emerge which do not show strong frequency correlations, as are sometimes attested (Fratini et al. 2014).

At the start of each simulation, a fraction of the forms are irregular, selected at random uniformly from among the lexicon to represent the state of after a recent sound change. The items follow an empirically Zipfian frequency distribution. How this distribution is interpreted depends on the specifics of each simulation. For the simulations in this paper, the lexicon

is taken to be very simple, without declensions or particular semantic relationships.

Decades of literature, as early as the Monte Carlo simulations of Klein (1966), presents many options for representing the population of agents in the speech community. Classic iterated learning (Kirby & Hurford 2002, Zuidema 2002, Kirby, Dowman & Griffiths 2007, Ackerman & Malouf 2015), which focuses on linear transmission, is sufficient for representing Andersen's Z-Model. For example, Ackerman & Malouf (2015), argued against a cognitively-explicit No Blur Principle by showing that it could emerge as a consequence of transmission relied on iterated learning of artificial non-skewed paradigms. However, iterated learning is insufficient for Sibling-Induced Change because compressing transmission into a linear generational chain eliminates population-level and age cohort effects.

Contrasting with iterated learning, a large literature exists which models social networks and language change (e.g., Nettle 1999, Baxter, Blythe, Croft & McKane 2009, Fagyal, Swarup, Escobar, Gasser & Lakkaraju 2010, Blythe & Croft 2012, Kodner 2020b). The models that have been proposed are quite diverse in both the level of social detail and the mathematical interpretations. The fine details of social networks are certainly important for the study of language change, and have been one of the major focuses in sociolinguistics over the past several decades (cf. Milroy & Milroy 1985, Milroy & Llamas 2013), and this has included agent-based modeling which explicitly adopts or tests sociolinguistic proposals (Fagyal et al. 2010, Kodner 2020b). However, in the spirit of testing the baseline of Sibling-Induced Change, which does not rely on network effects, only a simple population model is needed here. The simulation encodes three concepts to model Sibling-Induced Change (3):

(3) **Desiderata for a simulation of Sibling-Induced Change**

   i. Language acquisition should not be instantaneous. It should extend for multiple iterations but taper off or stop eventually.

   ii. Interactions should occur in a population in which 'child' agents receive many inputs from many other agents.

   iii. Learning should be based partially on input from other child agents who themselves have not matured.

These desiderata were taken into account when developing the simulation framework. For tractability, the simulation represents a local speech community and contains 100 individuals. This number was chosen to be the same order of magnitude as Dunbar's Number for personal networks (Dunbar 1992), which is the approximate number of reasonably robust social connections that an individual may hold.

Every simulated agent in the community has an associated age, where the youngest individuals are "learners" who are still acquiring their language, and the rest are "linguistically mature" and are no longer updating their grammars. At each iteration, the oldest member of the community "dies" and is removed from the simulation, and a new child is "born" and added to the simulation. This both roughly reflects the changing composition of populations over time and keeps the population size steady, which facilitates interpetability.[5] Every surviving agent is incremented in age so that the oldest learner becomes the youngest linguistically mature speaker and stops updating their grammar. This results in a stable population size of both learners and linguistically mature members, with individuals constantly cycling through as they age.

At each iteration, each learner has 1,000 interactions with other members of the community. Each interaction is with a randomly chosen individual, and involves the uttering of one item by that individual to the learner. The items are sampled following a Zipfian distribution and may either be regular or irregular according to that individual's grammar. This is an implementation of the Sibling-Induced Change Z-model because each learner receives input from across the population including from older learners.

Every agent at the start of the simulation shares the same set of regular or irregular variants. Learners each record a regular or irregular variant for each item in the lexicon as they receive them through interactions. As such, each item is potentially subject to variation in the learner's input or may be unattested altogether. learners can then choose to regularize irregular forms: following the treatment of unordered variation in Sneller, Fruehwald & Yang (2019) and Kodner & Richter (2020), they each adopt the majority variant for their cumulative input up to the end of that iteration. This approach is motivated by the experimental literature on young learners and the leveling of variation.

Because of Zipfian input sparsity, some items may simply not be attested to a given learner, as is often the case empirically. In these cases, the learner needs to infer the missing forms. Contrasting with prior work on network modeling, simulated learners apply a cognitively motivated model of learning model to infer missing forms. Learners applies the Tolerance Principle (Yang 2016), a model of productivity learning which has been applied successfully in a range of experimental settings (Schuler 2017, Koulaguina & Shi 2019, Emond & Shi 2020), as well as questions of variation and change

---

5 Fixed population size as a modeling assumption has a long history in computational studies of population genetics (Moran 1958), where it facilitates much cleaner and interpretable mathematical results.

in phonology, morphology, and syntax (Yang 2016, Sneller et al. 2019, Kodner 2020a, Nowenstein, Sigurjónsdóttir, Yang, Ingason & Wallenberg 2020, Björnsdóttir 2021, Richter 2021, Ringe & Yang 2022, Trips & Rainsford 2022, Kodner 2022). If the regular pattern is productive for that learner, they will fill in the missing forms. If the regulars are not productive, then those forms will be left as paradigmatic gaps following (Gorman & Yang 2019).[6] If an individual extends a productive pattern and overwrites an irregular, then it has been over-regularized. Over-regularization of this type is extremely common in language acquisition, while the opposite, over-irregularization is quite rare (Clahsen & Rothweiler 1993, Xu & Pinker 1995, Maratsos 2000, Yang 2002, Maslen et al. 2004, Mayol 2007). Over-regularization is also present in languages with large highly regular paradigms such as Turkish, though at a lower rate commiserate with the rarity of irregular items (Aksu-Koç 1985).

The Tolerance Principle is a decision procedure which makes local choices based on type count. It does not track token frequency or paradigm complexity and is thus *laissez-faire* according to the definition used in this paper. This contrasts with learning models that optimize for some global properties of the lexicon or paradigm (see Yang (2017) for discussion). More formally, the Tolerance Principle determines whether an apparent pattern with exceptions should be entered into the grammar as such (i.e., it should be productive), or it should not be. If it is not, narrower regularities may be discovered, or all items may be memorized. The formal definition from (Yang 2016) is presented in (4):

(4) 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:

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

Importantly, N and e are measures of an individual's lexicon, which is itself dependent on the individual's input, which we can estimate with a corpus. These are not meant to be a data-independent external measure of the language or of a given corpus *per se*. Corpora are important inasmuch as they provide us with estimates for the lexicon of a typical individual. As part of a simulation, Tolerance Principle or any other *laissez-faire* learning model that can be motivated, connects corpus analysis to predictions about the grammars of individuals.

---

6 In practice, paradigmatic gaps do not emerge in these simulations, so they receive no further discussion. However, the simulation as implemented allows gaps to emerge in principle, and can be used to study them if different initialization parameters are used.

The Tolerance Principle view of productivity provides an explanation for the discrepancy between over-regularization and over-irregularization: it is very unlikely (but not necessarily impossible) for a rare minority pattern to achieve productivity and be over-extended. Individuals in the simulation can grow out of over-regularization, and they usually do, once they receive new inputs in subsequent iterations. However, if they fail to receive sufficient evidence that an item should be irregular, or they receive sufficient confirmatory evidence from older individuals who have also over-regularized, then they will carry the innovation into adulthood. High-frequency irregulars should almost always be learned as such because they will be present in the input, but low-frequency items have a real chance of being permanently regularized.

At the end of 100 iterations, when none of the original community members remain, the forms (regular or irregular) of each item in the youngest mature speaker's grammar are reported. Simulations were repeated for 500 trials and the final outcomes were averaged to calculate probabilities of regularization by item frequency and frequency rank.

3.2.1 Simulation 1: Varying the interaction model

This simulation tests whether Sibling-Induced Change and sparse attestation are sufficient to yield the relationship between frequency and regularization with *laissez-faire* learners who do not track token frequency directly or attempt to globally optimize their lexicons. Following the intuition of that model, the more a learner interacts with slightly older peers, the faster regularization should take place. These simulations are carried out with lexicons of 100 items with two initial conditions: one in which 10 out of 100 items are initially irregular and one in which 20 of 100 items are initially irregular. The 10 or 20 irregulars are re-assigned uniformly at random at the start of every trial. These ratios were chosen so that the irregulars lay below the TP tolerance threshold for 100 ($\theta = 21.7$), which prevents the irregulars from becoming productive. The items are expressed according to a Zipfian frequency distribution and can be more concretely interpreted as representing either 100 roots, which follow Zipfian attestation, or a particular inflectional category across 100 roots, which follow a similarly long-tailed sparse attestation due to paradigm saturation. While it is clear that items in the lexicon should follow a Zipfian distribution, it is less clear what distribution the interactions between individuals should follow. Three interaction models were tested: one in which learners interact with others at a uniform rate irrespective of their age, one in which interaction rate decreases linearly with distance in age so that learners are more likely to learn from their older peers than much older adults, and one in which interaction rate follows a Zipfian distribution

decreasing with age difference so that learners are much more likely to learn from their older peers than older adults. In all the following simulations, the interaction model clearly affects the rate of change but not the general trend of frequency-sensitive regularization.

Figure 5 shows the regularizing outcome for each initially irregular item for both the 10-irregular and 20-irregular simulations, with *x*-axes indicating sampled token frequency rank and *y*-axes indicating the rate at which the items which were initially irregular items retained irregularity. In all three cases, strong, clearly visible, but not perfect, negative correlations between token frequency rank and irregularity emerge (Zipfian interaction rate $\rho = -0.828$; linear interaction rate $\rho = -0.772$; uniform interaction rate $\rho = -0.617$). That is, high-frequency items are much more likely to retain their irregularity than low-frequency items are, but as is known empirically, they are not required to. As predicted by the corpus analyses in Section 2, Zipfian and other long-tailed distributions in stem frequency, paradigm saturation, and inflectional category saturation establish a biased baseline towards regularization of low-frequency inflectional categories and low-frequency roots relative to high-frequency ones. From a synchronic typological perspective, irregulars are more likely to be frequent. This is the "conserving effect" of frequency (Bybee 1985, Bybee & Thompson 1997), but here driven by learners who do not track frequency.

Quasibinomial models were fit for each interaction model predicting the rate of retention of irregularity by token frequency and the number of initial irregulars and by token frequency rank and the number of irregulars. Quasibinomial models were chosen because the data lies between zero (consistent regularization) and one (consistent retention of irregularity). For each interaction model, both token frequency and token frequency rank were significant predictors, while the number of initial irregulars was not. Estimates were strongest for the Zipfian interaction model and weakest for the uniform interaction model (Zipfian $\beta_{freq} = 0.115$, $\beta_{rank} = -0.175$; Linear $\beta_{freq} = 0.077$, $\beta_{rank} = -0.068$; Uniform $\beta_{freq} = 0.058$, $\beta_{rank} = -0.049$), consistent with visual inspection of Figure 5. It is also consistent with the prediction of Sibling-Induced Change that more immature peer-to-peer interaction should drive change more quickly. Details for each model can be found in the Appendix. Note that the model predicting irregularity by frequency rather than rank is likely more appropriate, since it is frequency in the input rather than rank *per se* that is driving the effect of retaining irregularity.

Retention of Irregularity by Frequency Rank:
Zipfian Interaction Probs by Age Difference
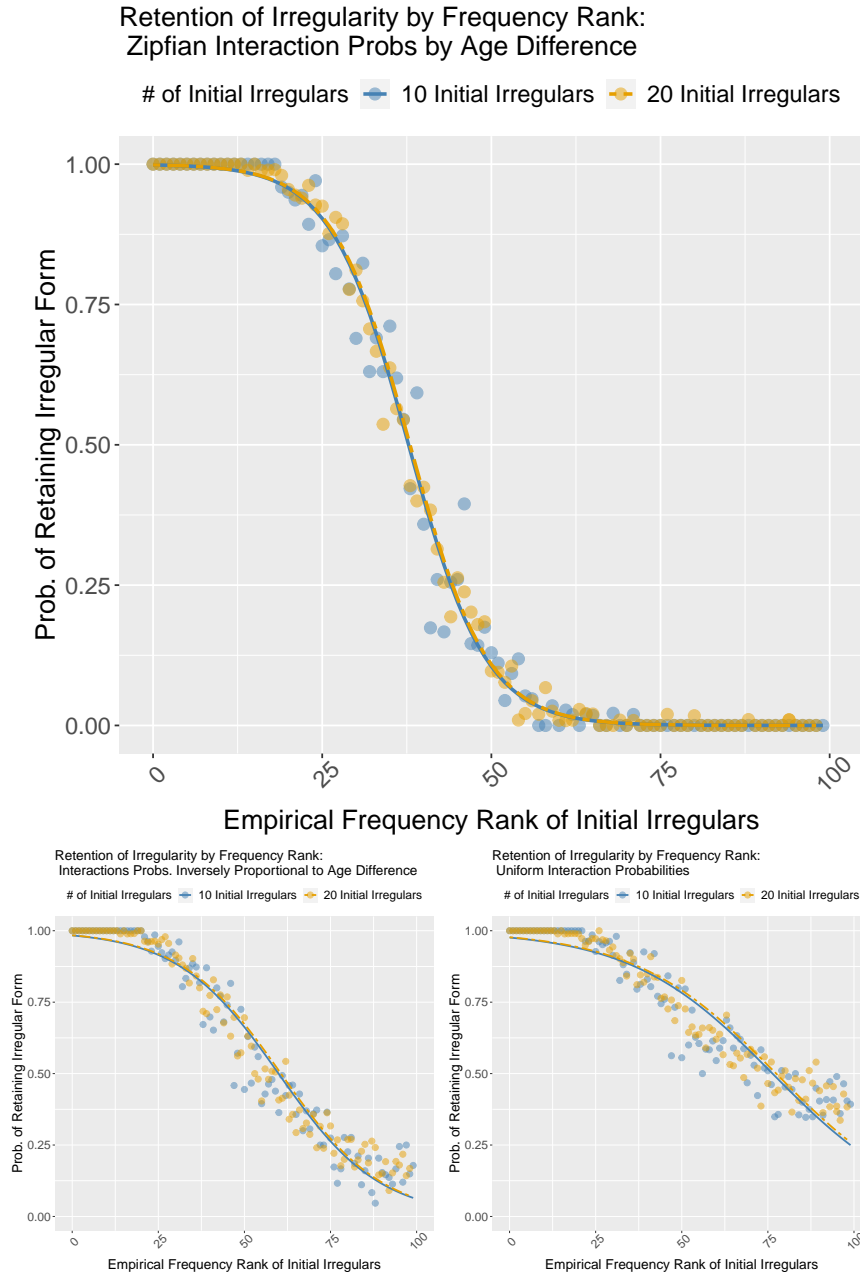
Figure 5 Relationship between token frequency rank and irregularity retention in a simulation of Sibling-Induced Change with *laissez-faire* Tolerance Principle learners and Zipfian input sparsity. Solid blue and dashed gold lines indicates the model mean estimates for overall irregularity retention by for 10 and 20 initial irregulars respectively.

3.2.2   *Simulation 2: Varying paradigm size*

This simulation is similar to the first, except that it investigates the effect of paradigm size on regularization. Rather than a lexicon of 100 items, this simulation models the inflectional paradigm of a single root or stem. Attestation of inflectional categories is skewed and sparse due to inflectional category saturation. Different paradigm sizes ranging from 10 to 100 are tested, each with an initial distribution of $\lfloor 0.9 * \theta_N \rfloor$ uniformly distributed initial irregulars. This number was chosen to lie just under the tolerance threshold. Inflectional categories in the paradigm are produced according to a Zipfian distribution, so the expectation is that high frequency categories should retain irregularity more readily than infrequent categories. Furthermore, paradigms as a whole should remain more irregular if they are smaller, since individual categories will be produced in the input more often on average.

Figure 6 shows the results for 500 trials. Each interaction model shows a similar significant trend, though it is once again strongest in models with the most child-to-child interaction (Zipfian $\beta = -0.046, p < 2e - 16$; Linear $\beta = -0.045, p < 2e-16$; Uniform $\beta = -0.041, p < 2e-16$). Each column in the plot represents a paradigm of given size, and the y-axis indicates the probability of retaining an initially irregular form. Small dots colored by frequency rank indicate individual categories or cells in the paradigm. Large black dots show the descriptive mean for each paradigm size, and the black curve shows the model's mean estimate. It is clear that high-frequency inflectional categories are more likely to remain irregular and that more leveling occurs in larger paradigms. The overall mean is pulled down by the lowest frequency items in the largest paradigms, since those were regularized in most trials.

In addition, there is still a significant (though subtle) effect even when only categories of equivalent frequency rank are compared across paradigms. That is, sparsity imposes a stronger biases towards regularization across all categories in a larger paradigm, not just the most infrequent ones. For visualization, an individual quasibinomial curve is plotted for each token rank independently. This shows generally downward trends as paradigm size increases, as expected. However, the resulting model has more parameters than data points, so the results are not quantitatively interpretable. Furthermore, it is frequency, not rank *per se* that facilitates the retention of irregularity. Therefore, a model which predicts the interaction of paradigm size and token frequency was employed for the quantitative analysis. It predicts a subtle but significant positive trend: items of the same frequency are more likely to be regularized if they are part of a larger paradigm. (Zipfian $\beta = 0.001, p = 0.001$; Linear $\beta = 3.916e - 04, p = 6.76e - 12$; Uniform $\beta = 1.542e - 04, p = 0.001$). Full analyses are provided in the Appendix.
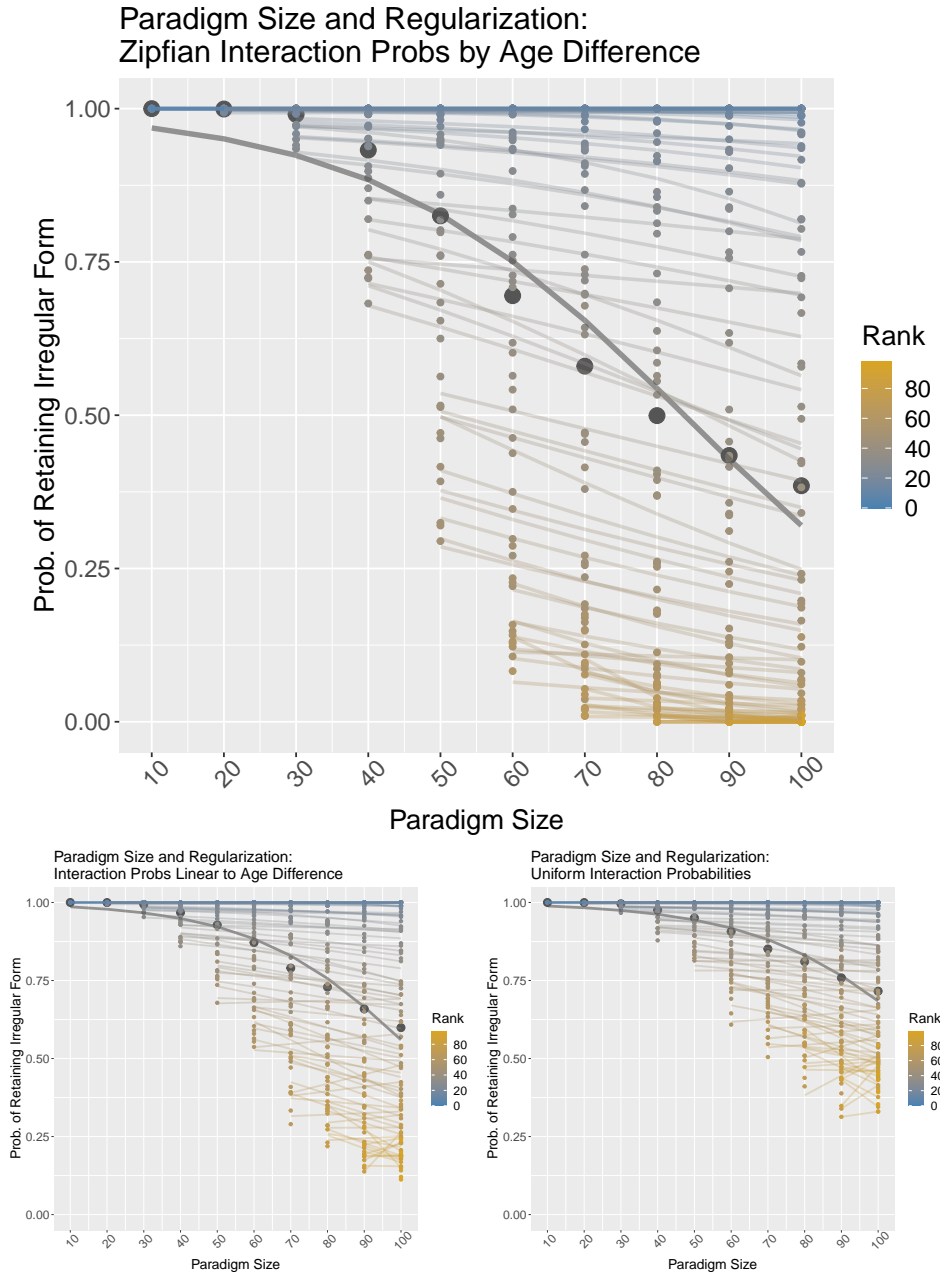
Figure 6 Relationship between paradigm size and irregularity retention. Black dots indicate the empirical mean retention of irregularity across inflectional categories for each paradigm size. The black curve indicates the model mean estimates for overall irregularity retention, while colored curves are estimates for each token rank.

This simulation produces the trade-off that Ackerman & Malouf (2013) observe between paradigm size and irregularity: smaller paradigms support more irregularity because their forms are less likely to regularize over of time, while larger paradigms tend towards regularity because their forms are more likely to regularize over time. This can be accounted for by the simple fact that any given inflectional category will be less likely to be reliably attested in a larger paradigm than a smaller one, which is supported by the additional observation that paradigm size yields more regularization even for items with similar token frequency. Taken together, the two simulations presented in this section complement the empirical corpus investigation in Section 2 to provide support for the framework Sibling-Induced Change. They are a proof-of-concept, showing that, together with the Tolerance Principle and sparse type attestation, it is capable of reproducing the conserving effect of token frequency against analogical leveling.

4 CONCLUSIONS

A wide range of proposals have been put forth to account for the many tendencies of analogical change as well as the typological trends that they induce on morphological systems. Many of them, from sensitivity to token frequency to management of complexity, may play some role. However, their relative contributions are hard to differentiate, since they all fit the data, are correlated with one another, and are often conceptually related. A proper baseline is necessary to compare them, that is, we need to know what biases in analogical change would exist, if any, prior to any of the cognitive proposals being taken into account?

This builds such a around the idea of "*laissez-faire*" learner-speakers who do not take a proactive role in reorganizing or optimizing their paradigms. Two trends are investigated: First, the tendency for analogy to level lower frequency items (whether the paradigms of lower-frequency stems or the lower frequency inflectional categories within a paradigm) on the basis of higher frequency items rather than vice-versa. This manifests synchronically as a positive correlation between irregularity and frequency. Second, the tendency for larger paradigms to exhibit less irregularity than smaller paradigms. In both cases, the *laissez-faire* baseline predicts these trends as a consequence of simple attestation. An item that is unattested in the input needs to be inferred by a learner on the basis of items that are attested, and lower-frequency items are less likely to be attested than higher-frequency items.

Supporting evidence for this baseline was drawn from morphologically annotated corpora as well as computational simulations. Section 2 investigates morphological sparsity in child-directed, modern adult-directed, and

historical adult-directed corpora, and identifies two measures of type attestation of inflected forms and inflectional categories,paradigm saturation and inflectional category saturation, which correlate strongly with raw token frequency. Both of these point to highly skewed distributions of attestation that drive the *laissez-faire* baseline. Furthermore, these sparse distributions recur regardless of text genre, including in child-directed speech, and are only alleviated by small paradigm sizes.

Section 3 presents a framework called "Sibling-Induced" Change for reasoning about how sparse input during language acquisition snowballs into the actuation of population-level change. Combining this framework with the Tolerance Principle, a concrete *laissez-faire* model for the acquisition of linguistic generalizations, provides sufficient formalization to implement baseline simulations. Two population-level simulations were set up, one for both tendencies of analogy. Both show that simple attestation, in the context of *laissez-faire* learners, creates a strong bias towards thet typological tendencies of analogical change. Furthermore, the predictions of "Sibling-Induced" Change are borne out, namely, that peer-to-peer interaction among young speakers is sufficient for actuating changes derived from learner innovations.

The simulations carried out in this study are simple. This was a deliberate choice in order to pinpoint a set of basic conditions for the observed trends in analogy. However, the downside of this approach is that they have limited resolution. They cannot evaluate more specific patterns of analogy than they can conceivably simulate. For example, we know that paradigms express internal semantic and phonological patterns that may influence the course of analogical change. At a broader level, items often group into declensions or conjugational classes according to the expression of their paradigms. More complex simulations, perhaps built on the ones here, would be needed to determine what effect such patterns would have on a baseline.

There are also clear limitations to simple attestation (and token frequency) as an explanation for analogical change. Most notable is the fact that analogy sometimes affects very high frequency items that must be thoroughly attested to all learners. A clear example of this lies in levelling of the copula independently in varieties of Germanic. While English retains more person/number distinctions in the the copula than in other verbs (e.g., *am, are, is* in the present), some other languages have leveled it completely to the third person singular (e.g., Afrikaans *is*, continental North Germanic textiter/*är*). But it is important to recognize simple attestation is distinct from the concept of a *laissez-faire* learner, so one may ask if such learners still play some role in these changes.

This study assumes that analogical change is driven primarily by children,

this is not a new viewpoint (Paul 1880, Halle 1962, Andersen 1973, Baron 1977, Lightfoot 1979, Niyogi & Berwick 1997, Yang 2002, Kroch 2005, van Gelderen 2011, Yang 2016, Cournane 2017, Kodner 2020a: *inter alia*). While it is likely that factors other that simple attestation play some part in analogical change, and it is certain that acquisition is not the only driver of change (Labov 1994, 2001), the corpus study and simulations provided here provide support for the feasibility of language acquisition as a driver of morphological change and strong evidence that linguistic transmission exerts biases which drive typological trends even in the absence of more complex theoretical and psycholinguistic models. Future work investigating mechanisms of analogical change must analyze their impact *in addition* to the biased baseline. Since it is sufficient to achieve the generally observed typological trends, additional proposals should focus on finding the necessary mechanisms for yielded the specifics of analogical change passed the basic trends. The same applies to the study of language change in general. What other biased baselines have been out there all along?

REFERENCES

Ackerman, Farrell, James P Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P Blevins & J Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–81. Oxford, UK: Oxford University Press.

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89(3). 429–464.

Ackerman, Farrell & Robert Malouf. 2015. The No Blur Principle effects as an emergent property of language systems. In *Proceedings of the annual meeting of the Berkeley Linguistics Society*, vol. 41 41, 1–14.

Aksu-Koç, Ayhan A. 1985. The acquisition of Turkish. In D Slobin (ed.), *The cross-linguistic studies of language acquisition. vol. 1: The data*, 839–876. Hillsdale, NJ: Lawrence Erlbaum Associates.

Albright, Adam. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.), *Linguistic universals and language change*, 144–181. Oxford, UK: Oxford University Press.

Andersen, Henning. 1973. Abductive and deductive change. *Language* 49(4). 765–793.

Austin, Alison C, Kathryn D Schuler, Sarah Furlong & Elissa L Newport. 2022. Learning a language from inconsistent input: Regularization in child and adult learners. *Language Learning and Development* 18(3). 249–277.

Baayen, Harald. 1993. On frequency, transparency and productivity. In

Geert E. Booij & Jaap Van Marl (eds.), *Yearbook of morphology 1992*, 181–208. Dordrecht: Kluwer.

Baron, Naomi S. 1977. *Language acquisition and historical change*. Amsterdam: North-Holland Publishing Company.

Baxter, Gareth J, Richard A Blythe, William Croft & Alan J McKane. 2009. Modeling language change: an evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21(02). 257–296.

Behrens, Heike. 2006. The input–output relationship in first language acquisition. *Language and Cognitive Processes* 21(1-3). 2–24.

Björnsdóttir, Sigríður Mjöll. 2021. Productivity and the acquisition of gender. *Journal of Child Language* 48(6). 1209–1234.

Blevins, James P, Petar Milin & Michael Ramscar. 2017. The Zipfian paradigm cell filling problem. In *Perspectives on morphological organization*, 139–158. Brill.

Blythe, Richard A & William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88(2). 269–304.

Bowerman, Melissa. 1988. The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In J A Hawkins (ed.), *Explaining language universals*, 73–101. Oxford: Basil Blackwell.

Braine, Martin D S. 1971. On two types of models of the internalization of grammars. In Dan Slobin (ed.), *The Ontogenesis of Grammar: A Theoretical Symposium*, vol. 1971, 153–186. New York: Academic Press.

Brown, Roger. 1973. *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Brown, Roger & C Hanlon. 1970. Derivational complexity and order of acquisition in child speech. In J.R. Hayes (ed.), *Cognition and the development of language*, New York: Wiley.

Bybee, Joan & Sandra Thompson. 1997. Three frequency effects in syntax. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, vol. 23 1, 378–388.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, Joan L & Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language* 259(2). 251–270.

Chan, Erwin. 2008. *Structures and distributions in morphological learning*. Philadelphia, PA: University of Pennsylvania dissertation.

Chapman, Don & Royal Skousen. 2005. Analogical modeling and morphological change: the case of the adjectival negative prefix in english. *English Language & Linguistics* 9(2). 333–357.

Chomsky, Noam. 1959. A review of BF Skinner's Verbal Behavior. *Language* 35(1). 26–58.

Chomsky, Noam. 1980. On cognitive structures and their development: A reply to Piaget. In In M. Piatelli-Palmarini (ed.), *Language and Learning*, 35–52. MIT Press Cambridge, MA.

Clahsen, Harald, Fraibet Aveledo & Iggy Roca. 2002. The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language* 29. 591–622.

Clahsen, Harald & Monika Rothweiler. 1993. Inflectional rules in children's grammars: Evidence from German participles. In *Yearbook of morphology 1992*, 1–34. Springer.

Cournane, Ailís. 2017. In defense of the child innovator. In Eric Mathieu & Robert Truswell (eds.), *Micro change and macro change in diachronic syntax*, 10–24. Oxford: Oxford University Press. doi:10.1093/oso/9780198747840.003.0002.

Deen, Kamil Ud. 2005. *The acquisition of Swahili*. John Benjamins Publishing.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25(2). 108–127.

Dunbar, Robin IM. 1992. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 22(6). 469–493.

Emond, Emeryse & Rushen Shi. 2020. Infants' rule generalization is governed by the Tolerance Principle. In Danielle Dionne & Lee-Ann Vidal Covas (eds.), *45th Annual Boston University Conference on Language Development*, 191–204.

Ervin, Susan M & Wick R Miller. 1963. Language development. *Child Psychology* 108–143.

Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser & Kiran Lakkaraju. 2010. Centers and peripheries: Network roles in language change. *Lingua* 120(8). 2061–2079.

Fernald, Anne & Virginia A Marchman. 2012. Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development* 83(1). 203–222.

Frank, Stella, Kenny Smith & Christine F Cuskley. 2020. Learner dynamics in a model of wug inflection: Integrating frequency and phonology. In *Proceedings of the 42nd annual conference of the cognitive science society*, 709–715.

Fratini, Viviana, Joana Acha & Itziar Laka. 2014. Frequency and morphological irregularity are independent variables. evidence from a corpus study of spanish verbs. *Corpus Linguistics and Linguistic Theory* 10(2). 289–314.

van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language*

faculty. Oxford University Press.

Goldrick, Matthew & Brenda Rapp. 2007. Lexical and post-lexical phonological representations in spoken production. *Cognition* 102(2). 219–260.

Gorman, Kyle & Charles Yang. 2019. When nobody wins. In F. Rainer, F. Gardani, W. U. Dressler & H. C. Luschützky (eds.), *Competition in Inflection and Word-Formation*, 169–193. Springer.

Guy, Gregory R & Sally Boyd. 1990. The development of a morphological class. *Language Variation and Change* 2(1). 1–18.

Halle, Morris. 1962. Phonology in generative grammar. *Word* 18(1-3). 54–72. doi:10.1080/00437956.1962.11659765.

Han, Chung-hye, Jeffrey Lidz & Julien Musolino. 2007. V-raising and grammar competition in Korean: Evidence from negation and quantifier scope. *Linguistic Inquiry* 38(1). 1–47.

Hare, Mary & Jeffrey L Elman. 1995. Learning and morphological change. *Cognition* 56(1). 61–98.

Hock, Hans Henrich. 1991. *Principles of historical linguistics*. Berlin: Walter de Gruyter.

Hock, Hans Henrich. 2003. Analogical change, 441–460. Malden: Blackwell.

Hoenigswald, Henry M. 1978. The annus mirabilis 1876 and posterity. *Transactions of the Philological Society* 76(1). 17–35.

Howes, Davis. 1968. Zipf's Law and Miller's random-monkey model. *The American Journal of Psychology* 81(2). 269–272.

Hudson Kam, Carla L & Elissa L Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1(2). 151–195.

Jelinek, Frederick. 1997. *Statistical methods for speech recognition*. MIT press.

Kiparsky, Paul. 1968. Linguistic universals and linguistic change. In *Universals in linguistic theory*, 171–202. New York: Holt, Rinehart, and Winston.

Kirby, Simon, Mike Dowman & Thomas L Griffiths. 2007. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences* 104(12). 5241–5245.

Kirby, Simon & James R Hurford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language*, 121–147. Springer.

Klein, Sheldon. 1966. Historical change in language using Monte Carlo techniques. *Mechanical Translation and Computational Linguistics* 9(3). 67–81.

Kodner, Jordan. 2020a. *Language Acquisition in the Past*: University of Pennsylvania dissertation.

Kodner, Jordan. 2020b. Modeling language change in the st. louis corridor. *Language Variation and Change* 32(1). 77–106.

Kodner, Jordan. 2022. What Learning Latin Verbal Morphology Tells us about Morphological Theory. *Natural Language & Linguistic Theory* 41. 733–792.

Kodner, Jordan & Caitlin Richter. 2020. Transparent /ai/-raising as a contact phenomenon. In Ruaridh Purse & Yosiane White (eds.), *Penn Working Papers in Linguistics: Selected Papers from NWAV*, vol. 25 2, 61–70.

van Koert, Margreet, Olaf Koeneman, Fred Weerman & Aafke Hulk. 2015. The quantificational asymmetry: A comparative look. *Lingua* 155. 140–153.

Koulaguina, Elena & Rushen Shi. 2019. Rule generalization from inconsistent input in early infancy. *Language Acquisition* 26(4). 416–435.

Kroch, Anthony. 1994. Morphosyntactic variation. In *Proceedings of the thirtieth annual meeting of the chicago linguistic society*, vol. 2, 180–201. Chicago, IL: Chicago Linguistics Society.

Kroch, Anthony. 2005. Modeling language change and language acquisition. In *Expansion of an lsa institute forum lecture*, .

Kuryłowicz, Jerzy. 1945. La nature des procès dits «analogiques». *Acta linguistica* 5(1). 15–37.

Labov, William. 1972. Some principles of linguistic methodology. *Language in society* 1(1). 97–120.

Labov, William. 1989. The child as linguistic historian. *Language Variation and Change* 1(1). 85–97.

Labov, William. 1994. *Principles of Linguistic Change, Volume 1: Internal Factors*. John Wiley & Sons.

Labov, William. 2001. *Principles of Linguistic Change, Volume 2: Social Factors*. John Wiley & Sons.

Labov, William. 2011. *Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors*. John Wiley & Sons.

Labov, William. 2020. The regularity of regular sound change. *Language* 96(1). 42–59.

Labov, William, Malcah Yaeger & Richard Steiner. 1972. *A quantitative study of sound change in progress*, vol. 1. US Regional Survey.

Lightfoot, David W. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.

Lignos, Constantine & Charles Yang. 2018. Morphology and language acquisition, 765–791. Cambridge, UK: Cambridge University Press.

Liu, Yingtong, Elodie Winckel, Anne Abeillé, Barbara Hemforth & Edward Gibson. 2022. Structural, functional, and processing perspectives on linguistic island effects. *Annual Review of Linguistics* 8. 495–525.

Loukatou, Georgia, Camila Scaff, Katherine Demuth, Alejandrina Cristia & Naomi Havron. 2021. Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language* 40. 1173–1192.

Mańczak, Witold. 1980. Laws of analogy. In J Fisiak (ed.), *Historical Morphology*, 283–288. De Gruyter Mouton.

Manly, John Matthews. 1930. From generation to generation. In N Bøgholm, A Brusendorff & C Bodelson (eds.), *A Grammatical Miscellany Offered to otto Jespersen on His 70th Birthday*, 287–289. Copenhagen: Levin & Munksgaard.

Maratsos, Michael. 2000. More overregularizations after all: New data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen & Xu. *Journal of Child Language* 27(1). 183–212.

Marcus, Gary F. 1993. Negative evidence in language acquisition. *Cognition* 46(1). 53–85.

Maslen, Robert JC, Anna L Theakston, Elena VM Lieven & Michael Tomasello. 2004. A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and Hearing Research* 47. 1319–1333.

Mayol, Laia. 2007. Acquisition of irregular patterns in Spanish verbal morphology. In Ville Nurmi & Dmitry Sustretov (eds.), *Proceedings of the 12th ESSLLI Student Session*, 1–11. Dublin.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Google Books Team, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182.

Miller, George A. 1957. Some effects of intermittent silence. *The American Journal of Psychology* 70(2). 311–314.

Milroy, James & Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21(2). 339–384.

Milroy, Lesley & Carmen Llamas. 2013. Social networks 407–427.

Moran, Patrick Alfred Pierce. 1958. Random processes in genetics 54(1). 60–71.

Nardy, Aurélie, Jean-Pierre Chevrot & Stéphanie Barbu. 2014. Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study. *Language Variation and Change* 26(3). 273–301.

Nettle, Daniel. 1999. Using social impact theory to simulate language change. *Lingua* 108(2-3). 95–117.

Nivre, Joakim et al. 2018. Universal Dependencies 2.3. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. http://hdl.handle.net/11234/1-2895.

Niyogi, Partha & Robert C Berwick. 1997. A dynamical systems model for language change. *Complex Systems* 11(3). 161–204.

Nowenstein, Iris, Sigríður Sigurjónsdóttir, Charles Yang, Anton Ingason &

Joel Wallenberg. 2020. The meaning of case: Morphosyntactic bootstrapping and icelandic datives. In Megan M. Brown & Alexandra Kohut (eds.), *45th Annual Boston University Conference on Language Development*, 402–415. Sommerville, MA: Cascadilla Press.

Papafragou, Anna. 1998. The acquisition of modality: Implications for theories of semantic representation. *Mind & language* 13(3). 370–399.

Paul, Hermann. 1880. *Prinzipien der sprachgeschichte*. Tubingen, Germany: Niemeyer.

Pearl, Lisa & Jon Sprouse. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition* 20(1). 23–68.

Pinker, Steven. 1995. Why the child holded the baby rabbits: A case study in language acquisition. In Lila R. Gleitman & Mark Liberman (eds.), *An invitation to cognitive science*, vol. 1, 107–133. Cambridge, MA: MIT Press.

Pinker, Steven & Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28(1-2). 73–193.

Ravid, Dorit & Rola Farah. 1999. Learning about noun plurals in early Palestinian Arabic. *First Language* 19(56). 187–206.

Richter, Caitlin. 2021. *Alternation-sensitive phoneme learning: Implications for children's development and language change*: University of Pennsylvania dissertation.

Ringe, Don & Charles Yang. 2022. The threshold of productivity and the 'irregularization'of verbs in early modern english, 91.

Roberts, Julie & William Labov. 1995. Learning to talk Philadelphian: Acquisition of short a by preschool children. *Language variation and change* 7(1). 101–112.

Samara, Anna, Kenny Smith, Helen Brown & Elizabeth Wonnacott. 2017. Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology* 94. 85–114.

Sankoff, Gillian & Hélèn Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83(3). 560–588.

Schuler, Kathryn D. 2017. *The acquisition of productive rules in child and adult language learners*: Georgetown University dissertation.

Singleton, Jenny L & Elissa L Newport. 2004. When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology* 49(4). 370–407.

Sneller, Betsy, Josef Fruehwald & Charles Yang. 2019. Using the Tolerance Principle to predict phonological change. *Language Variation and Change*

31(1). 1–20.

Tiersma, Peter Meijes. 1982. Local and general markedness. *Language* 58(4). 832–849.

Trips, Carola & Thomas Rainsford. 2022. Tolerating subject-experiencers? Yang's Tolerance Principle applied to psych verbs under contact in Middle English. *Journal of Historical Syntax* 6(12). 1–43.

Vennemann, Theo. 1972. Phonetic analogy and conceptual analogy. In *Schuchardt, the neogrammarians, and the transformational theory of phonological change*, 181–204. Frankfurt: Athenäum.

Vitevitch, Michael S & Paul A Luce. 2016. Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics* 2. 75–94.

Walkden, George. 2019. The many faces of uniformitarianism in linguistics. *Glossa: A Journal of General Linguistics* 4(1).

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (eds.), *Directions for historical linguistics*, 95–195. University of Texas Press.

Xu, Fei & Steven Pinker. 1995. Weird past tense forms. *Journal of Child Language* 22(3). 531–556.

Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yang, Charles. 2013. Who's afraid of George Kingsley Zipf? or: Do children and chimps have language? *Significance* 10(6). 29–34.

Yang, Charles. 2016. *The Price of Linguistic Productivity*. Cambridge, MA: MIT Press.

Yang, Charles. 2017. Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition* 24(2). 100–125.

Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley.

Zuidema, Willem. 2002. How the poverty of the stimulus solves the poverty of the stimulus. *Advances in neural information processing systems* 15.

Jordan Kodner
Social & Behavioral Sciences Bldg, S201
Stony Brook, NY, USA 11790
....
jordan.kodner@stonybrook.edu
jkodner05.github.io

A APPENDIX

A.1 Models for Figure 5

Variables:

```
init_variant_rate: proportion of trials in which item has been regularized
tokenfreq:         item's token frequency across the simulations
tokenrank:         item's empirical token frequency rank across simulations
ratio:             whether the simulation began with 10 or 20 irregulars
```

A.1.1 Zipfian interaction rate

```
Call:
glm(formula = init_variant_rate ~ tokenfreq + ratio, family = quasibinomial(link = "logit"),
    data = data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.41596  -0.18163  -0.10388   0.00532   0.60174

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.477121   0.365132 -20.478   <2e-16 ***
tokenfreq    0.115166   0.004813  23.927   <2e-16 ***
ratio        0.003586   0.015117   0.237    0.813
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.06560832)

    Null deviance: 194.3763  on 199  degrees of freedom
Residual deviance:   6.5916  on 197  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 10


Call:
glm(formula = init_variant_rate ~ tokenrank + ratio, family = quasibinomial(link = "logit"),
    data = data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.41502  -0.04768  -0.00815   0.08821   0.46977

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.571717   0.276448  23.772   <2e-16 ***
```

```
tokenrank   -0.174965   0.005567 -31.429   <2e-16 ***
ratio        0.003611   0.011415   0.316    0.752
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for quasibinomial family taken to be 0.03719037)

    Null deviance: 194.3763  on 199  degrees of freedom
Residual deviance:   2.9902  on 197  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 8
```

A.1.2   Inverse linear interaction rate by age

```
Call:
glm(formula = init_variant_rate ~ tokenfreq + ratio, family = quasibinomial(link = "logit"),
    data = data)


Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.45959  -0.10888   0.00000   0.06445   0.41846


Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.442963   0.129749 -26.536   <2e-16 ***
tokenfreq    0.077107   0.002170  35.535   <2e-16 ***
ratio        0.007440   0.005986   1.243    0.215
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for quasibinomial family taken to be 0.02484054)

    Null deviance: 107.4351  on 199  degrees of freedom
Residual deviance:   4.0259  on 197  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 9


Call:
glm(formula = init_variant_rate ~ tokenrank + ratio, family = quasibinomial(link = "logit"),
    data = data)


Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.51786  -0.09078   0.06504   0.20947   0.54195


Coefficients:
```

Laissez-Faire Analogical Change

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.039093   0.158057  25.555   <2e-16 ***
tokenrank   -0.068214   0.001814 -37.611   <2e-16 ***
ratio        0.004744   0.007544   0.629     0.53
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.03863955)

    Null deviance: 107.4351  on 199  degrees of freedom
Residual deviance:   8.3206  on 197  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

A.1.3   Uniform interaction rate

```
Call:
glm(formula = init_variant_rate ~ tokenfreq + ratio, family = quasibinomial(link = "logit"),
    data = data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.36040  -0.06945   0.00178   0.06276   0.36543

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.878068   0.090548  -20.74   <2e-16 ***
tokenfreq    0.058006   0.001555   37.30   <2e-16 ***
ratio        0.005250   0.004340    1.21    0.228
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.01431069)

    Null deviance: 63.2009  on 199  degrees of freedom
Residual deviance:  2.7717  on 197  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 9


Call:
glm(formula = init_variant_rate ~ tokenrank + ratio, family = quasibinomial(link = "logit"),
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5546  -0.1266   0.0543   0.2305   0.4250
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.656992   0.150925  24.231   <2e-16 ***
tokenrank   -0.048586   0.001565 -31.046   <2e-16 ***
ratio        0.005229   0.007212   0.725    0.469
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.03894411)

    Null deviance: 63.2009  on 199  degrees of freedom
Residual deviance: 8.9375  on 197  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

A.2   Models for Figure 6

Variables:

```
  init_variant_rate: proportion of trials in which item has been regularized
  paradigm_size:     size of paradigm
  tokenfreq:         item's token frequency across the simulations
```

A.2.1   Zipfian interaction rate

```
Call:
glm(formula = init_variant_rate ~ paradigm_size, family = quasibinomial(link = "logit"),
    data = variantdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4239  -0.8200   0.2411   0.6860   1.5093

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.888804   0.305220   12.74   <2e-16 ***
paradigm_size -0.046419   0.003832  -12.11   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.5775065)

    Null deviance: 494.12  on 549  degrees of freedom
Residual deviance: 378.54  on 548  degrees of freedom
AIC: NA
```

Laissez-Faire Analogical Change

```
Number of Fisher Scoring iterations: 4


Call:
glm(formula = init_variant_rate ~ paradigm_size * tokenfreq,
    family = quasibinomial(link = "logit"), data = variantdata)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.51409  -0.13900   0.00000   0.01914   0.40384

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -2.5937462  1.1331415  -2.289 0.022461 *
paradigm_size           -0.0476415  0.0146119  -3.260 0.001182 **
tokenfreq                0.0429921  0.0154706   2.779 0.005641 **
paradigm_size:tokenfreq  0.0007009  0.0002091   3.352 0.000859 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.212565)

    Null deviance: 494.123  on 549  degrees of freedom
Residual deviance:  11.492  on 546  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 12
```

A.2.2   Inverse linear interaction rate

```
Call:
glm(formula = init_variant_rate ~ paradigm_size, family = quasibinomial(link = "logit"),
    data = variantdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1903  -0.3870   0.1989   0.4972   1.0771

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.710087   0.279028   16.88   <2e-16 ***
paradigm_size -0.044695   0.003309  -13.51   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.2818262)

    Null deviance: 246.38  on 549  degrees of freedom
Residual deviance: 174.91  on 548  degrees of freedom
```

```
AIC: NA

Number of Fisher Scoring iterations: 5


Call:
glm(formula = init_variant_rate ~ paradigm_size * tokenfreq,
    family = quasibinomial(link = "logit"), data = variantdata)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.37487  -0.05934   0.00012   0.05168   0.34667

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -8.970e-01  2.918e-01  -3.074  0.00222 **
paradigm_size           -2.368e-02  3.391e-03  -6.982 8.51e-12 ***
tokenfreq                3.549e-02  4.433e-03   8.007 7.10e-15 ***
paradigm_size:tokenfreq  3.916e-04  5.582e-05   7.017 6.76e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.0223863)

    Null deviance: 246.3822  on 549  degrees of freedom
Residual deviance:   5.7955  on 546  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 12
```

A.2.3   *Uniform interaction rate*

```
all:
glm(formula = init_variant_rate ~ paradigm_size, family = quasibinomial(link = "logit"),
    data = variantdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9575  -0.2970   0.1782   0.3944   0.8711

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.879283   0.252785   19.30   <2e-16 ***
paradigm_size -0.041059   0.002965  -13.85   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.1786038)
```

```
    Null deviance: 156.84  on 549  degrees of freedom
Residual deviance: 110.16  on 548  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6




Call:
glm(formula = init_variant_rate ~ paradigm_size * tokenfreq,
    family = quasibinomial(link = "logit"), data = variantdata)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.36983  -0.03610   0.00174   0.05420   0.29102

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)           -9.009e-01  2.375e-01  -3.793 0.000166 ***
paradigm_size         -9.294e-03  2.699e-03  -3.444 0.000618 ***
tokenfreq              4.264e-02  3.712e-03  11.489  < 2e-16 ***
paradigm_size:tokenfreq 1.542e-04  4.551e-05   3.389 0.000753 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.01302098)

    Null deviance: 156.8401  on 549  degrees of freedom
Residual deviance:   4.6443  on 546  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 12
```