

UNIVERSITY OF AMSTERDAM

MSC ECONOMICS

TRACK: BEHAVIOURAL ECONOMICS & GAME THEORY

MASTER THESIS

**Inflectional Morphology as Shaped by the
Brain: An Evolutionary Model of
Language Change**

by

YUVAL FERTIG

11085525

August 18, 2016

Supervisor:

PROF. VAN VEELLEN

Statement of Originality

This document is written by Yuval Fertig who declares to take full responsibility for the contents of this document. I declare that the text and the work presented in this document is original and that no sources other than those mentioned in the text and its references have been used in creating it. The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.

Abstract

One of the most polarising problems in linguistics has been to explain how language – with its seemingly infinite set of syntactic, semantic, pragmatic and phonological conventions – can be learnt by children barely able to tie their shoelaces. Chomsky (1980) deduced from this that humans have evolved to conceive of language in very specific ways, allowing them to infer the same (universal) grammar given disparate and scarce data. A more recent solution inverts this logic: humans have not evolved to easily understand specific grammar, but rather specific grammar has evolved to be easily understood by humans (Christiansen & Chater (2008)). In this dissertation I present a usage-based model of language change which provides evidence in favour of the latter solution. I model diachronic patterns as the emergent properties of cognitive processes that evolved long before language, and find that simulations robustly reproduce six stylized facts that have marked the competition between Germanic inflectional paradigms. I therefore conclude that inflectional morphology is, to a great extent, shaped by the brain.

1 Introduction

The similarity between the development of linguistic structures and biological organisms has long teased scholars into entertaining the possibility of design without a designer. Writing a century before Darwin, Hume (1779) proposed the thought experiment that “there is a natural, universal, invariable language, common to every individual of human race, and that books are natural productions, which perpetuate themselves in the same manner with animals and vegetables, by descent and propagation.” Partly inspired by “Hume’s close encounter” (Dennett (1995)), Darwin wrote that “the survival of certain favoured words in the struggle for existence is natural selection” (Darwin (1888)). Like biological systems, languages consist of variants which compete for selection through a (anthropological) niche: languages compete for speakers; synonyms compete for meanings; homonyms compete for sounds etc. I have chosen to focus on the competition between inflections because morphological success is in many respects the easiest to explain. It is immediately apparent why “–thisisnotagoodsuffix” would not survive as an inflection: it is hopelessly inefficient, unaesthetic, and downright confusing. More importantly for my purposes, the morphological success of an inflection can often be explained in terms endogenous to the system itself. This is because people have a tendency to inflect words in the same way they hear other words inflected. Paradigms are therefore subject to cumulative advantage, whereby popular inflections become more popular and unpopular become less popular. As Fertig (2013) quips, “Morphology is the reality TV of language. It is the domain of dominant patterns that have nothing to recommend them other than their dominance”.¹

Cumulative advantage is the basis for many existing models of inflectional morphology. Without any countervailing forces, these models predict that systems undergo unconditional regularization, stabilizing only at the point at which there is a one-to-one correspondence between grammar and inflectional form. Lieberman et al. (2007), for example, assume that the question as to whether a word is regularized is not a matter of ‘if’ but ‘when’. But this unidirectionality is simply not borne out by the facts. Contrary to Lieberman’s assumption, many words have in fact been *irregularized*. Cumulative advantage can therefore only be part of the story. Cultural, social and biological factors may also play a role, but I am at present interested in the extent to which *cognitive* factors can explain inflectional morphology. I would therefore like to focus on phonology.

¹ Despite sharing the same name and a passion for inflectional morphology, I do not know David Fertig

In a recent corpus study of the English past tense, Cuskley et al. (2014) show that much of what can't be explained by cumulative advantage – namely irregularization - can be accounted for by phonology. The authors show that irregularized words generally adopt the inflectional form of irregular phonetic neighbours. Take the verb “wed”. Lieberman et al. (2007) explicitly predict that this will be the next word to regularize; that the conjugation “wedded” will soon be more widespread than “wed”. Yet if we look at the data the recent trend has actually been in favour of “wed”. This makes little sense in terms of cumulative advantage: “-ed” is the more popular inflection and therefore ought to be the more productive paradigm. But it chimes with Cuskley et al.'s (2014) observation when we consider the conjugations of phonetic neighbours like “bed”, “shed” and “wet”. Thus it appears that people also have a tendency to inflect words in the same way they hear similar words inflected. We therefore have two explanations of paradigm change: cumulative advantage and phonetic consistency.

Both explanations have cognitive causes. In fact, their causes are intimately related: cumulative advantage is caused by the tendency to inflect words according to the *global* popularity of inflections; phonetic consistency is caused by the tendency to inflect words according to the (phonetically) *local* popularity of inflections. In psycholinguistics the latter tendency is generally assumed to take precedence: if a person forgets how a word has been inflected in the past, they first look to how similar words are inflected, and if that is not helpful they resort to a global default. This threefold thought process - *memory*, *analogy* and then *default* - takes centre stage in my model of language change. The model successfully simulates six stylized facts in Germanic inflectional morphology, and hence I am led to the conclusion that diachronic change is caused by (deficient) memory and created by analogy and default.

The paper proceeds as follows. Section 2 formalizes the evolutionary approach to language change and surveys other evolutionary models of inflectional morphology. Section 3 sets out the diachronic changes I hope to explain. Section 4 expounds the model and the results are evaluated in Sections 5 and 6. Section 7 addresses concerns about the validity of the model and points to further applications. Section 8 concludes.

2 Language and Evolution

Evolutionary models of language change have in recent years been popularized by Croft's (2000) seminal work "Explaining Language Change: an evolutionary approach". The book argues that language shares Darwin's two fundamental principles of evolution – variation and selection - and that this motivates embedding language change within Hull's (1988) generalized evolutionary framework. The units of variation in language change are linguistic structures, which in my case take the form of inflectional paradigms. Variation occurs when someone inflects a word differently to how they have heard it inflected before. In true evolutionary spirit, mutant inflections may be unintended, if for example the speaker mistakenly infers a novel suffixation. But innovation does not have to be 'blind'; people may employ a particular inflection for clarity, prestige or comic effect. Mutant inflections can be seen as in competition with rival inflections, vying for imitation.² Croft (2000) interprets them as Dawkinian replicators whose fate lies in the hands – or mouths - of the interactors qua humans. On the strength of the analogy to genes, Croft (2000) even goes so far as to coin the term "lingueme", defined as "a unit of linguistic structure, as embodied in particular utterances, that can be inherited in replication." I will take up Croft's interpretation in what follows.

The biological metaphor is particularly apt for the case of inflections. When an uninflected word is heard for the first time it simply enriches a person's vocabulary; uninflected words, even synonyms to an extent, do not fight for places in our vocabulary like animals fight for an ecological niche.³ In inflectional morphology, however, there is a dogma that an inflection is either right or wrong; that the correctness of "octopi", for example, stands and falls with the correctness of "octopuses". Even in derivational morphology "generosity" and "generousness" are not viewed in this same way. The perception means that inflections are engaged in an especially fierce zero sum competition for words. They are subject to a kind of linguistic 'Competitive Exclusion Principle' (Landsbergen (2009)), in which the presence of one precludes the other. This perception manifests at the emergent level in the form of the 'Unique Entry Principle' (Pinker (2009)), whereby inflections inhabiting the same semantic and grammatical niche must differentiate or die. It is important to note that this principle applies to intra-word competition only. Different inflections can stably serve the same grammatical function (as implied from the persistence of irregularization), but coexistence

² Cousse von Mengden (2014) argue that the two criteria for an inflection to be considered a mutant are (i) understandability and (ii) originality

³ Landsbergen (2009) makes the same point for syntactic competition between compounds and phrases (e.g. widescreen vs. wide screen) and Petersen (2012) makes the same point for spelling (e.g. X-ray and X ray)

does not last long when competing for the same grammatical function *and* the same word stem.

It is thus clear that the biological metaphor extends well to linguistic structures. Just how far the metaphor ought to be extended is however a major point of controversy. In particular it is hotly debated whether linguemes, like genes, are passed from parent to child, or whether variants propagate “in vivo”. This debate is more or less derivative to the debate over whether humans have evolved for language or whether language has evolved for humans. If one believes that humans have evolved for language, and hence that we have a ‘grammar gene’, then grammatical selection collapses into natural and sexual selection. If, on the other hand, one believes that grammar has adapted to the pre-existing cognitive biases of humans, then the locus of change must be in language acquisition and usage. Of course, grammar may have also evolved in response to non-cognitive factors as well, but these cannot be the whole story, as they do not resolve Chomsky’s “poverty of stimulus” (Chomsky (1980)). I therefore only consider biological and cognitive causes of inflectional variation.

Let us first consider the generative position, that humans have evolved to understand the same grammars. On this view, nature has selected for humans able to rapidly learn and process specific grammatical structures. This could be because of inherent grammatical superiorities, or as Pinker and Bloom (1990) suggest, it may well be another case of cumulative advantage: nature selects for humans that communicate effectively with the most other humans. Inflectional morphology might seem to provide evidence for genetically-encoded language universals. Indeed, in the following section I present a list of diachronic patterns that are sufficiently ubiquitous as to warrant cognitive explanation. It therefore stands to reason that these changes reflect the unfolding of genetic biases, causing languages everywhere to favour the same linguistic variants. Regularization, for example, could reflect a universal “optimizing impulse”; a “preference for . . . uniform paradigms in particular” (Garrett (2008)). But for the “optimizing impulse” to be a *language* gene, it must have evolved in response to a language that was *already* regular. This line of reasoning is therefore hopelessly circular: it assumes the existence of the very changes I wish to explain. If regularization is the result of some type of “optimizing impulse”, this impulse could not be a language gene per se; it must have evolved for some other non-linguistic purpose.

I believe the domain-general trait that largely shapes morphology is inductive inference, whereby we reason from the assumption that “similar causes are always conjoined with similar effects” (Hume (1738)). When applied to inflectional morphology, this assumption of regularity entails that similar

words are conjoined with similar inflections. Viewed in terms of phonetic similarity, the assumption motivates analogy; viewed in terms of grammatical similarity, inductive inference motivates the process of default. The logic is sound inasmuch as inflectional morphology is regular. And in this special case, it may even be self-fulfilling. For if enough people make the same regularizing mistake, and if others begin to imitate, then what people *think* is the correct inflection comes to define what *is* the correct inflection. If, for example, people increasingly conjugate “bring” as “brang” (by analogy to the conjugation “rang”), then by virtue of its widespread usage “brang” will eventually supersede “brought” in the battle for morphological success (after all, what other vantage point aside from convention do we have to deny inflectional status?). People *assume* morphological regularity and in doing so *impose* regularity on morphology. Perhaps more literally than Hume intended, this is an example of the mind’s “great propensity to spread itself on external objects”. It is because people extrapolate from popular patterns – both local and global – that popular inflections become more popular.

I have just suggested the pre-linguistic foundation of analogy and default. I will later expand on the role of these analogy and default in shaping inflectional variation and selection. Despite the processes being well-known in psycholinguistics, I believe my model will be the first evolutionary model to explain inflectional morphology in such deeply cognitive terms. The closest model in this respect was designed by Hare & Elman (1995) to explain the evolution of the English past tense. Agents in their model learn inflections using associative networks, which works in much the same way as analogy. The locus of change is on intergenerational learning and the source of variation is intergenerational error. The mechanism by which learning takes place is however rather robotic. Each new agent engages in an iterative process in which he repeatedly compares his output to that of his teacher’s and adjusts for error. Regularities are then borne out by the fact that a reduction in error on one presentation of patterns reduces error in all like patterns in proportion to their similarity, hence propagating most frequent and regular patterns. Besides being cognitively unrealistic, intergenerational learning does not seem, as a matter of empirical investigation, to be the main impetus of language change: infantile error simply does not reflect diachronic change (Croft (2000), Baxter et al. (2009)). By focusing on language learning as opposed to language use, Hare & Elman’s model fails to fully shake the generative dogma of the day.

More recent models, especially since Croft (2000), emphasize the role of language use over acquisition. Pijpops et al. (2015) develop a particularly elegant model to show how functional

advantages can explain the rise of the weak inflection from relative obscurity. Following Baxter et al.’s (2006) Utterance Selection Model, the language evolves as a result of pairwise conversations. Each conversation takes the form of a one-word monologue, in which a speaker inflects a word for another agent to hear. Upon hearing the word inflected, the listener adds a ‘token’ to his memory. An agent’s memory therefore consists of a bank of tokens, one for every time he has heard a word inflected. When required to inflect a word, agents do so based on the inflection types of tokens in memory. If an agent does not possess any tokens for the word, he then randomizes across any inflection that applies. Propagation is biased by the fact that the weak inflection (requiring no stem alternation) is more widely applicable than its strong counterparts. My model borrows this basic structure and essentially sandwiches another phase into the agent’s thought process: after checking their memory but before resorting to a default, agents in my model see if they can infer an inflection by analogy to the word’s phonetic neighbours.

3 Evaluation Criteria

In order to establish a perspective from which to evaluate the explanatory success of my model, I now list a number of stylized facts I wish to explain. The facts consist of six diachronic trends that have been widely observed in Germanic inflectional morphology. The examples I draw upon to elucidate the facts are primarily taken from the English past tense, which has come to be seen as an exemplar for how language is cognitively structured (Pinker & Ullman (2002)), but the trends have been observed across various grammatical cases and in highly divergent populations. Indeed, two of the trends are lifted directly from Pijpops et al. (2015) who are concerned with explaining the evolution of *Dutch* inflectional morphology.⁴ I therefore deem the trends to be sufficiently universal as to suggest a cognitive explanation.

1. Morphologization

Morphologization is the process by which a phoneme or morpheme takes on grammatical significance (Joesph (2003)). Let us consider an example, taken from Bybee (2015). It appears that the sounds we now associate with “v” and “z” evolved from the Old English voiceless fricatives “f” and “s” somewhere around the 15th century (Bybee (2015)). These

⁴ (Pijpops et al.’s (2015) evaluation criteria are (i) rise of the weak inflection (ii) gradual rise (iii) conserving effect (iv) class resilience. The very notion of evaluation criteria has also been borrowed from their paper.

alternations were not applied at random. In many cases the hardened fricative was employed specifically as a plural marker e.g. “wife/“wives, “thief”/“thieves”, “house”/“houses” (hazz) etc. Since these once meaningless phonemes are now indicative of grammar this is an example of morphologization. My model is broader in scope than standard usage-based theory in that I wish to model not just *any* instance of morphologization but also the *first* instances of morphologization.⁵ This may seem ambitious, but it is a logical conclusion of Christian Chater (2008) research programme that I embrace herein. For if one believes that linguistic structures adapt and have *always* adapted to the brain, then “it is plausible that historical processes of language change provide a model of language evolution; indeed historical language change may be language evolution in microcosm”.⁶ Following Christian & Chater’s (2007) reasoning, we can infer that the origin of morphology may have also been morphophonological.

2. Rise of the weak inflection

The trend that has been at the centre of linguistic attention is the global process of regularization. That there is a single regular inflection for each grammatical case has in fact been the premise of one of the most fiercely fought debates in psycholinguistics (see Pinker (1998) etc.). It is certainly true that inflectional systems are a lot more regular than they once were. Johnson (2016) refers to Old English as having a “jungle of endings”, and proto-Germanic as being more chaotic still. Irregular paradigms are nevertheless still rife and rising, and in many cases there is still no clearly dominant paradigm. More striking than the emergence of some dominant inflections is the nature of these nascent paradigms. Germanic languages have been subject to an unprecedented surge in the pervasiveness of weak inflections, inflections that do not alter the word stem. Weak inflections are opposed to strong inflections, which usually function by the altering of a vowel or consonant. To give a sense of the scale of the invasion, 74 of the 293 basic strong verbs in Old English are now weak (Fertig (2016)).⁷ As pointed out by Knooihuizen & Strik (2014), regularization may therefore more aptly be labelled as “weakenization”. Whilst I do not expect the authors hold out much hope for the evolutionary success of the term, the point still stands: any satisfactory model of diachronic change must

⁵ My model therefore straddles language evolution (the origin of language) and language change (the ongoing development of language).

⁶ Bybee (2010) makes this same argument: “there is every reason to assume that the first grammatical constructions emerged in the same way as those observed in more recent history.”

⁷ Of the others, 65 are still strong, 13 are mixed, 15 have become irregular weak verbs, and 23 have merged with an originally weak verb (Fertig (2016)).

explain why it is weak inflections in particular that have prevailed.

3. The conserving effect of frequency

As well as explaining why some words have been regularized, linguists must also explain why others have not. One reason is remarkably clear: frequency retards regularization (Bybee Hooper (2001)). Bybee (1997) calls this the ‘conserving effect’ of frequency. Working on a database of the English past tense, Lieberman et al. (2007) find that a verb that is 100 times less frequent regularizes 10 times as fast; a relationship that has been found to more or less hold across languages and grammatical cases, at least qualitatively (Carroll et al. (2012), Glushko (2004)).⁸ Notice that the most frequent verbs in English do not even conform to sub-regularities; conjugations like “was” and “went” bare no recognisable inflection whatsoever. These ‘suppletive’ verbs stand in stark contrast to weakly inflected verbs, which can be mechanically derived from their present tense. When analysing the effect of frequency it will therefore be fruitful to view regularity in continuous rather than binary terms.

4. Phonetic ‘islands of resilience’

Frequency is not the only thing that modulates regularization. Another factor is inflectional form. It is not immediately obvious why this should matter because the weak inflection can, by definition, invade any word regardless of their inflectional form. Nevertheless, Cuskley et al. (2014) find that the likelihood that English verbs have been regularized is determined in large part by the paradigm to which they belong. The authors analyse four different paradigms and find that whilst members of the “sing”/“sang” paradigm have resisted regularization, the last two centuries have seen the “dwell”/“dwelt” paradigms in rapid decay.⁹ One clue as to why this might matter case comes from Mailhammer’s (2007) study of the German verbal system. Mailhammer concludes that “it is exactly the remnants of the old system that keep the German strong verbs together as a group”. He coins these groups “islands of resilience”. Crucially, the islands Mailhammer speaks of are *phonetic* islands. They take this form because the original system of strong inflections partitioned phonetic islands of words very efficiently according to the sound of their stem. His point is that regularization has

⁸ Carroll (2012) repeats the test for German and Glushko (2004) finds that a similar relationship holds for the past participle.

⁹ Another decaying island is that consisting of the Latin declensions “octopus”, “platypus”, “hippopotamus”, “abacus”, “doofus”, “circus”, “status”, “bonus”, “virus”, “focus”.

ensued where phonological change has eroded the size and consistency of these islands.

5. Irregularization

Some islands have not just resisted regularization but have also been reversing it. In the last few centuries the “sing”/“sang” paradigm has attracted three new members: “ring”, “fling” and “string”. Fertig (2016) counts 18 cases of irregularization: “strive”, “dive”, “wear”, “stick”, “dig”, “ring”, “show”, “prove”, “sneak”, “catch”, “kneel”, “make”, “fit”, “bet”, “quit”, “plead”, “hurt”, “cost”, “quake”, “fling”, “string”, “hide”, “chide”, “strew”. To this list, I append six more: “wed”, “wet”, “wake”, “broadcast”, “light” and “leap”. Irregularization has been prevalent in other grammatical cases as well (the declensions of “cactus” and “hoof”, for example). Challenging the conventional wisdom of the day, Lounsbury (1908) writes: “A movement in one direction which threatened to sweep everything before it was much more than arrested. It was actually reversed” (as seen in Fertig (2016)). This was not an exaggeration. In a recent study of the English past tense, Cuskley et al. (2014) find that irregularization has been counteracting (endogenous) regularization since the 18th century.¹⁰ Bidirectionality should therefore not be overlooked.

6. S-curves

Thus far I have identified three types of change: morphologization, regularization and irregularization. My final diachronic trend concerns the dynamics of these changes. When plotted against time, language change frequently exhibits a distinctive S-shaped curve; a trajectory indicative of evolutionary systems. According to Blythe (2016), the S-curve is a language universal, and in a recent study of the English past tense Ghanbarnejad et al. (2014) show that regularization and irregularization is no exception: of the 10 cases analysed, they find that 8 satisfy their mathematical definition of the S-curve.¹¹ Being a language universal, we would expect the S-curve to manifest at the emergent level as well. Fertig (2016) talks about the ‘Great English Verb Regularization’ of the 14th and 15th century. He describes the weak verb inflection as spreading like “wildfire” which, according to the above story, feeds off the low frequency and phonetically vulnerable words before eventually running its course. Blythe

¹⁰ Cuskley et al. (2014) reconcile this fact with the undeniable ascension of weak verbs by disentangling endogenous from exogenous regularization, the former being the conversion of old words and the latter being the classification of new ones.

¹¹ “Abide”, “burn”, “chide”, “light”, “smell”, “spell”, “spill” and “thrive” all exhibited S-curves; “cleave” and “wake” did not.

Croft (2012) also documents the S-curve in the Brazilian Portuguese future and the French negation.

The S-curve is perhaps the hardest criterion to demonstrate because it is not a binary phenomenon or simple relationship. Moreover, it may not be entirely clear what it rules out. I will be interpreting the criterion qualitatively as referring to any change that starts slow, accelerates, and then plateaus. On the level of individual word change, this corresponds to Rastorgueva's (1989) three stages of language change: the appearance of the new features; coexistence and competition with the old ones; and the acceptance of new features and (consistent with the Unique Entry Principle) the disappearance of the old ones. I therefore deem any model inadequate that predicts the trajectory of change as being linear, exponential, or a random walk. Interpreted in this way, I believe this criterion is a valuable indicator that I have identified the right kind of mechanism driving language change.

The diachronic trends provide me with six falsifiable and non-trivial hypotheses: that morphemes reflect grammar; that a weak inflection emerges as regular; that high frequency words resist regularization; that phonetically consistent paradigms resist regularization; that irregularization is possible; and that the dynamics follow an S-curve. I will later evaluate the success of my model based on whether the results meet these criteria.

4 The Model

I have implemented my model using an object-oriented (OO) programming style in Python. The model consists of a community of agents each with a private lexicon. The words in the lexicon do not correspond to any words in the real world. Rather, they are words in the abstract, characterized by a single integer, which represents their unique position in phonetic space.¹² The word with the phonetic characteristic 7 (w^7), for example, would be the phonetic neighbour of w^8 . In order to have an organic phonetic distribution, I generate the phonetic positions at random between a lower and upper bound. The words are all assumed to be of the same grammatical case, meaning that they are all subject to analogy and the same default.

Each word in an agent's lexicon has a token count for each inflection - one token for each time that the word has been heard in that inflection. A lexicon can therefore be thought of as a frequency

¹² Landsbergen (2009) abstracts words in a similar fashion, but on a lexical rather than phonetic continuum.

table with dimensions ‘Number of Words’ and ‘Number of Inflections’, with each element of the table being the count of how many times an agent has heard a word inflected in a particular way. In OO modelling terms, the model has a community of agents; each agent has a lexicon; each lexicon has a sequence of words; and each word has an array of word tokens. The community itself has an emergent language, which is calculated by element-wise addition across all agents’ lexicons, resulting in an emergent language in the same format as the lexicon of an agent. The interaction of objects in the model is illustrated in Figure 1.

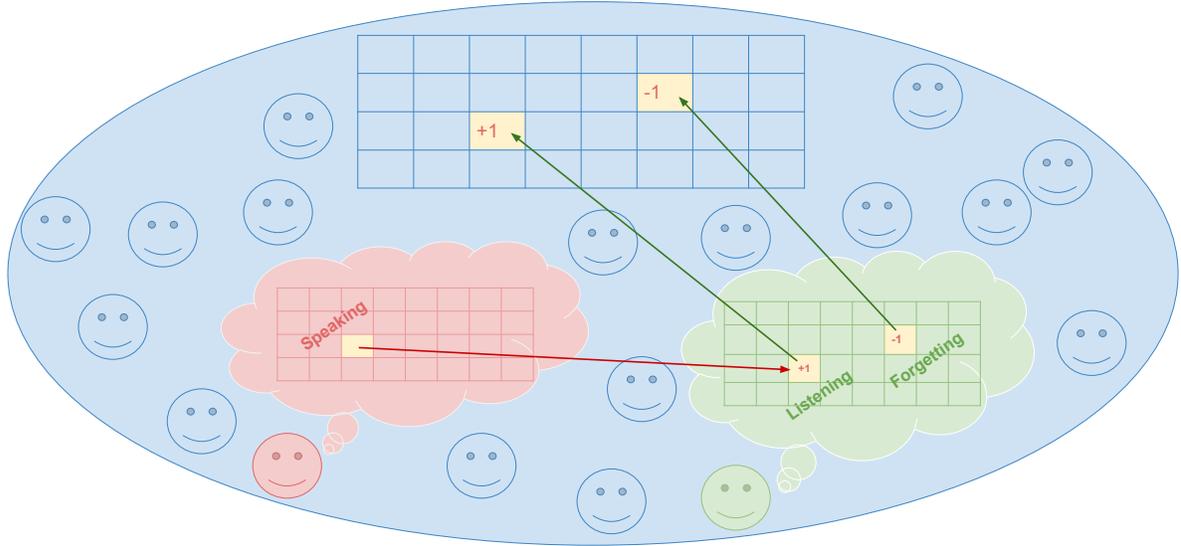
In order to generate a realistic distribution, tokens are initialized and allocated in a top-down process that works as follows. The initial frequency of each word in the language is determined by a Zipfian distribution, the trademark frequency distribution of words in a sociolinguistic language (Zipf (1932)). Within each word the tokens are allocated to a single inflection, the chosen inflection being different for each word, in effect modelling a pre-morphological proto-language. The weak inflection inhabits just one word of many. Each agent of the community is then allocated the tokens in proportion to their ‘vocabulary’, an integer determined on a bell-shaped curve. The initial lexicons therefore differ only in scale.

The structure of the game is adapted from Pijpops et al. (2015) and runs for a pre-determined number of iterations. Each iteration begins with a ‘speaker’ and ‘listener’ being chosen at random from the community of agents. The speaker is designated a word to inflect with a probability given by (a linear transformation of) the frequency of the word in the emergent language. The speaker then inflects the word. The listener ‘hears’ the speaker, and adds the inflected word to his lexicon in the form of an additional token. Finally, the listener ‘forgets’ a token (with each token equally likely to be forgotten), thus leaving the size of his vocabulary unchanged. In the rare instance that this leads to a word in the emergent language being forgotten altogether, a new word (with a new phonetic characteristic) is added to the language.¹³ The new word is initialized in exactly the same way as above (from a Zipfian distribution and apportioned according to vocabulary) except for that fact that it consists solely of weak (regular) tokens. This models Cuskley et al.’s (2014) observation that new verbs generally adopt the regular inflection.

Besides the entrance of new words, the main source of asymmetry in the model is in how speakers choose their inflection. They do so according to a K -Nearest Neighbour (KNN) algorithm, inspired

¹³ A word that is forgotten altogether becomes fixated at zero frequency because a positive number of tokens is necessary for a positive probability of the word being spoken again.

Figure 1: An iteration in the Agent-based Model



My model consists of a community of agents (the smiley faces); each agent has a lexicon (a table); each lexicon has a sequence of words (columns in the table); and each word has an array of tokens (counted by inflection in the elements of the table). The lexicons feed into the emergent language, which is an element-wise sum of all agents' lexicons. In each iteration, a listener (green) 'hears' a speaker (red) inflect a word, which is then added to his lexicon in the form of an additional token. The listener then forgets a token at random. The changes to the emergent language mirror the changes to the listener's lexicon because his is the only lexicon to be affected by the conversion.

by van Noord's (2015) model of memory-based learning. Firstly, the designated word activates all corresponding tokens in the speaker's lexicon. If the token count (across all inflections) for this word exceeds K , the speaker bases his inflection on how he has heard the word inflected in the past (as in Pijpops et al. (2015)). If the tokens do not exceed K , then the phonetically-next-nearest words are activated incrementally until the aggregate token count (of all words activated so far) exceeds K . If K tokens are successfully found within the speaker's 'phonetic search set' (an endogenously-determined range within phonetic space), the speaker inflects the word by means of analogy, based on the inflectional form of all activated tokens. If however the speaker fails to find K tokens within his phonetic search set, he resorts to the default (weak) inflection, invoking its general applicability.

Consider the situation in Figure 2, which shows lexicons in a language of 10 words and 3 inflections. For ease of analysis I have ordered the words according to their position in phonetic space. Suppose the speaker is required to inflect w^{37} (word with phonetic position 37) and suppose

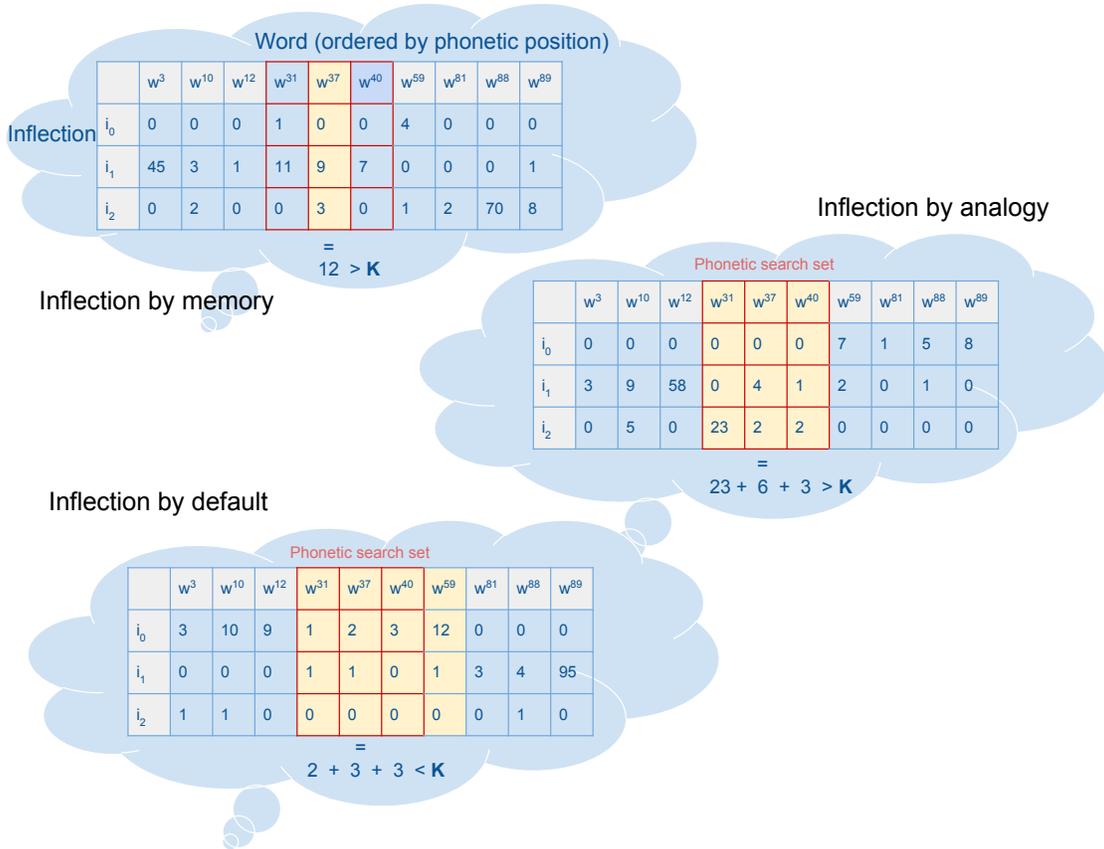
further that the token threshold is 10 ($K = 10$) and the phonetic search set is 20. In this first case, the number of times the speaker has heard w^{37} exceeds the threshold ($0 + 9 + 3 = 12 > K$), so he uses inflections i_0^{37} , i_1^{37} and i_2^{37} with probability $p_0^{37} = 0/12 = 0\%$, $p_1^{37} = 9/12 = 75\%$ and $p_2^{37} = 3/12 = 25\%$. The second speaker must call on tokens from w^{40} and w^{31} in order to pass the threshold. The probability of each inflection is determined *as if* the last word took the count to exactly K tokens, so the w^{31} tokens are down-weighted by $1/23$ because only one more of the word's 23 tokens was needed. The probability distribution is therefore $p_0^{37} = 0/K + 0/K + (1/23) * (0/K) = 0\%$; $p_1^{37} = 4/K + 1/K + (1/23) * (0/K) = 50\%$; and $p_2^{37} = 2/K + 2/K + (1/23) * (23/K) = 50\%$. The final speaker cannot find enough tokens within his phonetic search set to meet the threshold. He therefore resorts to the default and applies the weak inflection.

This threefold thought process may seem mechanical, but it embodies two key insights from psycholinguistics. The first is the process of analogy, and in particular the notion of proportional equations. As argued earlier, analogy is rooted in inductive inference. It proceeds from the assumption that similar words are conjoined with similar inflections. But this in itself is not enough to give a decisive call to action. If we are trying to conjugate the verb “dive”, for example, and we have as exemplars both “drive” (“drove”) and “thrive” (“thrived”), the assumption does not help us decide between “dove” and “dived”. What we need is a *measure* of similarity in order that we can calculate *which* example is the most relevantly similar. This is where proportional equations come in. A proportional equation is a formula which uses known words in order to solve unknown ones. As Fertig (2013) explains: “a complete proportional equation is typically a vast multi-dimensional array of forms. In a language like English, with a very simple inflectional system, we can largely make do with a two dimensional table. Imagine a very wide spreadsheet with a column for each of the thousands of verbs in English and five rows: [one for each grammatical case].” The above tables can therefore be interpreted in this light, with the threefold thought process now a general method for solving proportional equations.¹⁴ Crucially for my argument, Fertig also explains that “the proportional model was developed as an integral part of a coherent theory of the acquisition, mental representation and productive use of a grammatical system”. In other words, the proportional model is to be understood as a descriptive theory of what actually goes on in the brain.¹⁵

¹⁴ My table has also has two dimensions, but I only consider one grammatical case and instead add another dimension to allow for the coexistence of different inflectional forms that is necessitated by the token-based approach.

¹⁵ My model may in fact help corroborate this claim. This is because it side-steps the accusation of post hoc analysis that is frequently levelled against the proportional model. Here is Fertig (2013) summarizing the

Figure 2: Example Lexicons



Words in the lexicon are ordered according to their position in phonetic space. The first speaker inflects w^{37} by memory because he has enough w^{37} tokens to satisfy the threshold ($12 > K$). The second speaker inflects w^{37} by analogy because he requires tokens from the two nearest neighbours to satisfy the threshold. The third speaker defaults because he does not have sufficient tokens within his phonetic search set; his phonetic search set 20 and so the agent only searches words with phonetic properties between 27 and 47.

objection: “linguists start from the fact that a change has occurred and are satisfied as soon as they have identified a plausible proportional model to account for it”. To reuse the above example, the irregularization of “dive” (from “dived” to “dove”) could be explained by the analogy to “drive” (conjugated as “drove”) . The problem is that one can always find a proportional explanation in retrospect: if dive hadn’t irregularized, this could have been rationalized with the conjugations of “thrive”, “skive” and “jive”. The emergentist approach can help in this respect because if it is found that certain diachronic phenomenon, such as phonetic islands of resilience, could have only been generated by analogy, then this corroborates the existence of, and perhaps even sheds light on, the cognitive process of analogy. If inputted with real data the model would even predict precise historical instances of language change such as the irregularization of “dive”; a more rigorous but possibly futile test (Blythe (2011)). Thus paraphrasing Pinker, computational linguistics opens up a novel “window into human

Thus proportional equations, according to Fertig, are realistic picture of mental grammar, and they are to be solved with the use of a table much like mine.

The second insight concerns the theoretical underpinnings of the final stage thought process, the process of default. Being much simpler than the process of analogy, the realisticness of default will perhaps require less justification. The psycholinguistic literature nevertheless brings to light some non-obvious details regarding the scope and strength of the process. Intriguingly, Pinker (1998) finds evidence that an inflection can be the default even if it is not the most numerous. For example, the German plural marker “-s” is used in just 7% of nouns, and is nevertheless applied to nonce words, unusual-sounding nouns, foreignisms etc.; all the hallmarks of a default (Pinker (1998)). This example demonstrates that it is not so much ubiquity that determines the default – although this surely helps – but rather applicability. This is consistent with my assumption that the weak inflection assumes the position of default throughout, even when it is relatively obscure. One way to reconcile Pinker’s somewhat paradoxical finding that the ‘regular’ inflection may not in fact be the most regular inflection is to consider that whilst the designated default may not be a simple numbers game, popularity may nevertheless serve to modulate the strength of default. For example, it seems reasonable to assume that if the weak inflection is on parity with all others (as it is initially), agents will less readily resort to the default than if the weak inflection was dominant. I explain how I incorporate this assumption into the model in the next section. For now, I mainly wish to emphasize that the details of my agent’s thought process - memory, analogy and default - are well-grounded in psycholinguistic theory.

Before analysing how much these cognitive assumptions can explain, I think it is worth highlighting some of the more technical features of the model. The structure of the model is essentially the Utterance Selection Model (Baxter et al. (2009)) with the following simplifications. There is no population structure, so the probability with which an agent is selected to speak is independent of the identity of the listener, and vice versa; conversations are a one-word monologue and not a dialogue; there is never any mishearing on the part of the listener; and the word has no effect on the speaker’s lexicon. In spite of these simplifications, stochasticity creeps in at numerous junctures (when initializing the lexicons; when selecting a speaker and listener; when inflecting a word; and when forgetting a word). Consequently, each run is unique, and this has to be taken into account in the nature of the analysis.

nature” (Pinker (2007)).

5 Parametric Analysis

Given the inherent stochasticity in the model as well the sheer number of parameters, the key question is not simply whether the evaluation criteria are met, but whether they are met robustly. Let us focus on the parameters, which are as follows: number of agents; number of words; mean and variance of the vocabulary; token threshold; phonetic space; and phonetic search set. Most of these parameters are not important in and of themselves; their effect is relative to the others. As such, it is more fruitful to reduce them into two parametric combinations of interest.

Let us define:

$$\text{Normalized Token Threshold} = \text{Token Threshold} * (\text{Number of Words}/\text{Vocabulary})$$

This measures the fraction of tokens in an agent’s vocabulary that need to be activated in order to produce an inflection. It is perhaps more intuitive to consider the inverse of this function, which can be interpreted as **memory**. Memory is perfect if the normalized token threshold is such that an agent needn’t rely on analogy or default in order to produce any inflections.

Let us define:

$$\text{Normalized Search Set} = \text{Phonetic Search Set} * (\text{Number of Words}/\text{Phonetic Space})$$

This measures how many words we expect to find within a given phonetic search set. Since absolute values in phonetic space have no correlate in the real world they ought not to affect the normalized phonetic search set. I therefore assume that the phonetic search set increases linearly with phonetic space, thus neutralizing its effect. As argued above, I believe it is also reasonable to assume that the power of the default increases with the number of regular words. I therefore assume that the normalized phonetic search set decreases with the number of regular words in an agent’s lexicon. Together, these assumptions imply:

$$\text{Phonetic Search Set} = \text{Phonetic Space}/\text{Number of Regulars}$$

Substituting this into the first equation and simplifying, we get:

$$\text{Normalized Phonetic Search Set} = \text{Number of Words} / \text{Number of Regulars}$$

Again, it is more natural to consider to the inverse function, which I refer to as an agent’s **regularizing tendency**. The regularizing tendency is a measure of an agent’s readiness to resort to the default. The greater the proportion of regular words in an agent’s lexicon, the greater the regularizing tendency.

The first parametric combination, interpreted as memory, acts as a generally conserving force. A better memory means that a person is less likely to resort to analogy or default, and instead inflects words based more on how they’ve heard them inflected in the past. The second parametric combination of interest, the regularizing tendency, determines the relative strength of analogy and default. A higher regularizing tendency means that a person is more inclined to recourse to the default inflection. As agents have different lexicons, memory and regularizing tendency is specific each agent at each point in time. I will now argue that the evaluation criteria are borne out so long as memories and regularizing tendencies stay within certain reasonable bounds.

6 Results

I have chosen to analyse a simulation with 20 agents and 20 words. Since the game begins with just one weakly inflected word, the regularizing tendency is just 1/20 for all agents at this point in time. Memory is initialized such that the average agent can just about inflect every word solely based on how he has heard the word inflected in the past i.e. with no help from analogy or default.¹⁶

The language begins with as many inflections as words; a “primordial soup of conjugations” (Lieberman et al. (2007)). This is illustrated in the top right panel of figure 3 by the fact that the colour of a bar – corresponding to inflectional paradigms - is solely and uniquely determined by the word. At this point the inflections shouldn’t properly be called inflections at all since they do not signal grammar. The ‘inflections’ – whatever they may be - are meaningless as morphemes. Looking at the panel below we see that by 250,000 iterations words begin to share inflectional form: two words share the green inflection; three words share the turquoise inflection etc. These

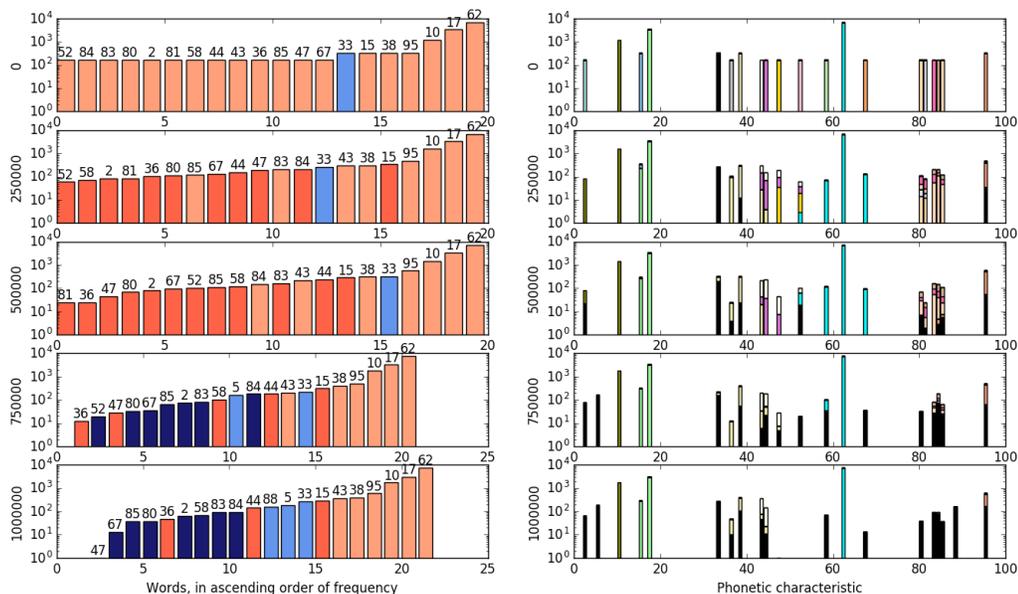
¹⁶ The peak of the bell-shaped vocabulary distribution is therefore determined by the lowest frequency word

words now stand in grammatical relation with one another; their morphemic features have become indicative of grammar, and morphologization is said to have taken place. With the regularizing tendency still low, mutant inflections are predominantly a product of analogy, and morphologization takes the form of phonetic clustering. Had regularizing tendencies been stronger, morphologization would have been less localised, but would have gone ahead nevertheless. The regularizing tendency therefore determines only the *nature* of morphologization. Whether morphologization takes place at all is determined by memory. Poor memory is the main impetus for the evolution of morphology; if people could perfectly recall any number of words regardless of their length and arbitrariness there would be no pressure for the language to be logical. But as a matter of fact humans find it difficult to remember strings of words if there is no morphemic logic between them (Nowak et al. (1999)). Agents speaking a language of uninflected content words are therefore especially likely to resort to analogy and default, which then produces variants that cohere with other words in the grammatical class. A language which consists of words with no morphemic relation to one another (as at the 0th iteration) therefore blends words of the same grammatical case (as at 250,000 iterations). These first instances of morphologization signify the transition from proto-language to “digital infinity”, in which humans combine a small set of phonemes to produce a potentially infinite number of meanings (Nowak et al. (1999)).

The language that emerges from this process of morphologization manifests the “jungle of endings” (Johnson (2016)), as seen in Latin or Greek. This is nicely illustrated by Figure 5, which shows many low frequency inflections floating around the primordial soup. This picture remains until around 700,000 iterations, at which point the weak (black) inflection begins to take hold. As in Pijpops et al. (2015), the rise is attributed to the weak inflection’s general applicability, and hence its availability as an inflection of last resort. In fact if $K = 1$ and $S = 1$ my model actually recovers that of Pijpops et al. (2015).¹⁷ Under these conditions incumbent agents would very rarely have recourse to the default, leading Pijpops et al. (2015) to systematically introduce newborn agents into their model. I tighten the constraint on memory ($K > 1$) and dispense of their assumption of agent replacement. In both models, the weak inflection is in some sense destined to rise; due to the phonetic constraints on analogy, it is the only inflection that *can* rise rapidly. However this does not

¹⁷ $K = 1$ means that agents need only one previous instance to produce an inflection, and $S = 1$ means that agents never inflect by analogy. Given these parametric combinations the only substantial difference between the thought process of agents in Pijpops et al. (2015) and my model is that they allow for multiple default inflections (agents may resort to strong inflections as well as the weak inflection).

Figure 3: Snapshots of the emergent language

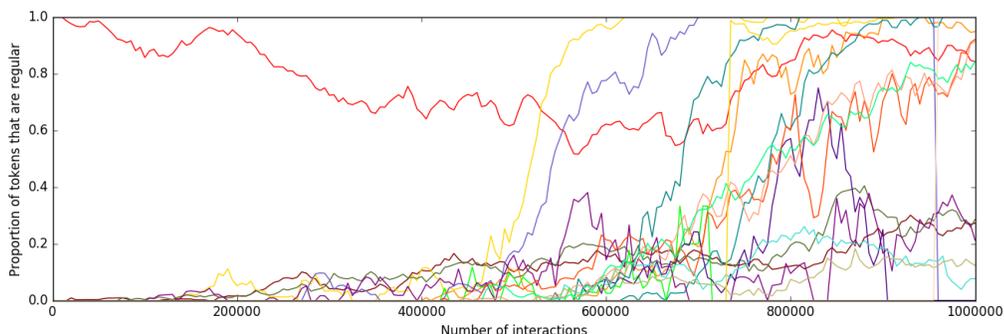


This figure illustrates the emergent language at five different snapshots in time. In the left hand panel each bar represents a single word in the emergent language. The height of the bar shows the frequency of the word and the colour of the bar shows the type of change the word has undergone. Light blue bars indicate that the word has retained a regular status and dark blue bars correspond to those that have been regularized. Likewise, light red bars correspond to words that have retained their original irregular status and darker red bars correspond to those that have undergone non-regularizing change. I define a word as 'regular' if it has more regular tokens than any other type of token. I have labelled each word with its phonetic position so that it can be cross-referenced with the right hand panel. Note, the bars in the last two snapshots have been shrunk by the entrance of new words. In the right hand panel the bars are positioned in phonetic space. Again their height represents the frequency but the colours within the stack represent the frequency of different token types within a given word. Black represents the weak inflection. A bar that is half black, for example, has half its tokens as regular. All other colours represent a strong inflection.

trivialize the conditions under which the weak inflection *will* rise. Regularization relies on a nucleus of 'defaulters'. Since the first word only begins to regularize after 500,000 iterations (see Figure 4), we can infer that this condition is not entailed by the initial memories and regularizing tendencies. Rather, it is satisfied only after around 500,000 iterations of 'neutral drift'. Perhaps low vocabulary agents spoke more than others; perhaps listeners disproportionately forgot an already low frequency word etc. All that matters is that the first word is regularized somehow and sometime. Key to the

analysis is what happens next. Following on from Pijpops et al. (2015), “the resulting decline in regularity of the strong system... fuel[s] the rise of the weak inflection, setting off a vicious cycle of weak ascension and strong decline.” In my model the cycle is driven by the regularizing tendencies, which increase with the number of regular words, hence increasing the number of speakers resorting to the default. Thus we do not have to exogenously assume a particularly bad memory or high regularizing tendency for the weak inflection to rise; neutral drift will ensure the regularization of the first word. The regularization of the rest is then caused by the self-intensifying regularizing tendency.

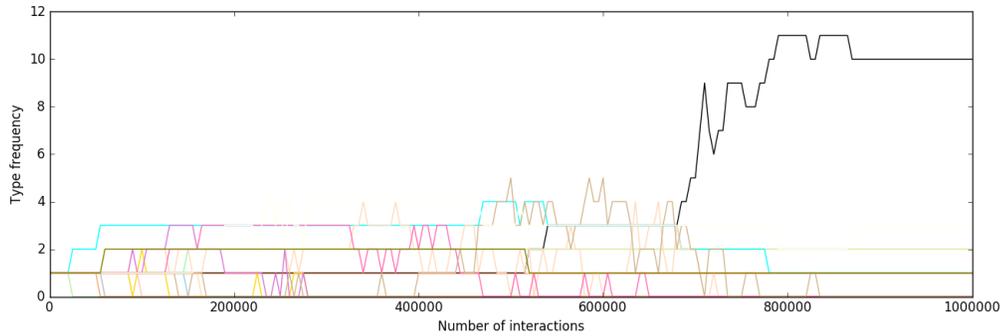
Figure 4: Proportion of regular tokens by word



Each line represents a single word in the emergent language. The lines show how the proportion of regular to irregular tokens change through time. An upwards-sloping line therefore indicates that the word is becoming increasingly regular.

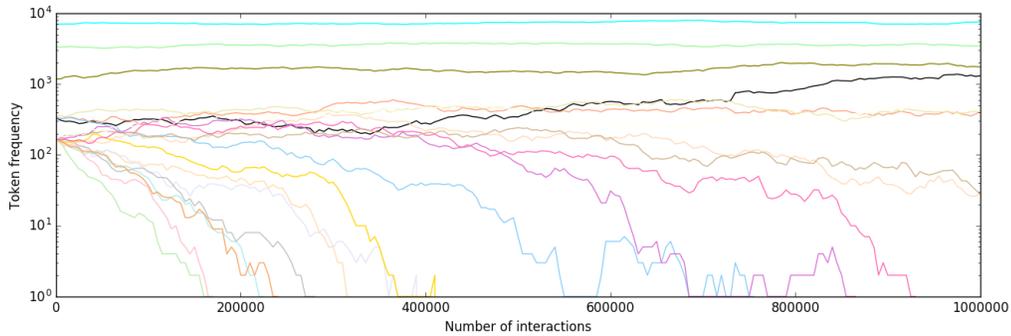
Regularization does not happen indiscriminately. Consistent with the conserving effect of frequency, it is rife amongst the least frequent words. This is most clearly illustrated in the left hand panel of Figure 3, which shows how words have changed over time. Observe that by the 750,000th iteration, six of the least frequent words have been regularized compared to just one from the rest. The same point can be made by comparing Figures 5 and 6, which show the weak inflection simultaneously dominating the word count despite having only the 4th most tokens. The frequency effect is best understood in terms of the token-based approach. The fewer tokens a word has in a speaker’s memory, the less likely the word is to meet the token threshold, and hence the more likely the agent is to produce a mutant inflection. Where this variant is the product of analogy, it generally catalyses non-regularizing change, as we saw was prevalent in the early stages of the game.

Figure 5: Word frequency of inflections



Each line represents a different inflection. The lines show the number of words for which a given inflection is most popular. Colours are consistent with the other figures, so the black line represents the number of regular words.

Figure 6: Token frequency of inflections



Each line represents an inflection. The lines show the number of tokens for a given inflection at each point in time. Again, the colour of each inflection is consistent with the other figures.

Where the variant is the product of default, it contributes to regularization. Both morphologization and regularization are therefore driven by the same underlying mechanism: a linguistic adaption to memory (or lack thereof). High frequency words which are easily recalled are resistant to either type of change. This explains the existence of suppletive verbs that were noted in the evaluation criteria: if a word is inflected so frequently that speakers do not even start by recalling the stem then the inflection can evolve independently of its base case (Bybee (2013)). Less frequent words for which speakers need more prompting track their root, and may also track the inflections of their

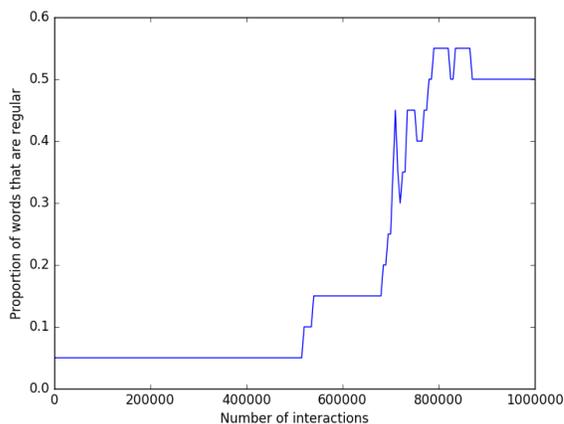
phonetic neighbours and global grammatical class, depending on frequency. A word’s ‘recollective autonomy’ is therefore the key to its stability. It will drive a wedge between the high and low frequency words so long as memories are neither perfect nor nonexistent.

The relationship between frequency and regularization is clearly not perfect. For example, w^{58} is regularized more slowly than some of its more frequent counterparts. Some of these anomalies may simply be down to the inherent stochasticity in the model, but most of them can be explained by phonology. A low frequency irregular word can resist regularization if it is supported by similar words of the same inflection. Take w^{36} , the lowest frequency word that is still in use by the 750,000th conversation. It is able to resist regularization because it neighbours a high frequency word (w^{38}) dominated by the same inflection. This means that even if a speaker doesn’t have enough w^{36} tokens to reach the token threshold (K), analogy fills the quota with neighbouring tokens of the same type. Consequently agents continue to inflect w^{36} irregularly, in spite of its frequency. If there are multiple low frequency words of the same inflection, the relationship can also be mutually supportive, as we see happening in the phonetic cluster $w^{80}, w^{81}, w^{83}, w^{84}, w^{85}$ in Figure 3. These relationships are part of a broader class of psycholinguistic ‘Frequency * Regularity’ interaction effects, in which low frequency words are only learnt and processed if part of a pattern (Christiansen Chater (2016)). Speakers are therefore able to inflect an uncommon word only if it conforms with the inflection of its phonetic neighbours. So long as there is the regularizing tendency does not altogether eliminate analogy, phonetic “islands of reliability” (Mailhammer (2007)) are thus translated into phonetic “islands of resilience”, the fourth of our evaluation criteria. To extend Mailhammer’s (2007) metaphor, the Frequency * Regularity interaction effect thus produces “islands of resilience”, that gain their strength from their uniformity (consistency with neighbouring words), and ‘tenacious trees’, that gain their strength is in their height (frequency).

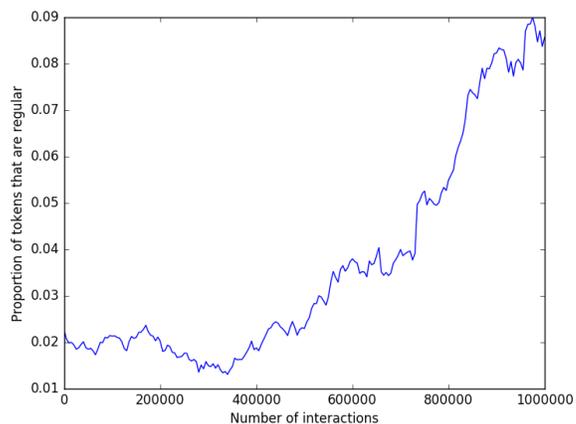
Consistent with Mailhammer’s study, my islands of resilience stand or fall on the basis of their internal consistency. Somewhere between the 500,000th and 750,000th conversation, w^{81} is forgotten altogether, and this catalyses the process of ‘paradigm levelling’ that was already underway by the 500,000th conversation. The reverse is also possible. Somewhere between the 750,000th and 1,000,000th iteration w^{47} is replaced by w^{88} . Had the $w^{80,82,83,84,85}$ cluster resisted regularization the island may have irregularized the newly formed word, in a process known as ‘analogical extension’. So whilst this specific run does not exhibit any clear-cut cases of irregularization, unidirectionality is by no means inherent to the model. In fact, after enough words are replaced, some

entrants will inevitably neighbour high frequency irregular incumbents that extend and ultimately irregularize the new word. The lack of irregularization may therefore be a mere sampling effect, especially if we consider that only around 20 out of the many hundreds of English verbs that exist have ever been irregularized. Whether the model properly explains irregularization is therefore open to investigation. Other simulations are encouraging, but even if we assume this run to be representative this may be as directing as it is dejecting. It points to other, external factors behind bidirectionality. Neogrammarian sound change, for example, may contribute greatly to phonetic consistency and subsequent irregularization. Sociolinguistic factors may also play a role. For example, one can imagine a certain amount of prestige associated with Latinate suffixes exemplified by “cacti”. Nevertheless, I satisfy myself with the fact that the model has exhibited morphologization, regularization and is at least consistent with irregularization. All that remains to be shown is that these changes obey realistic dynamics.

Figure 7: Proportion of regular words and tokens in the emergent language



(a) Proportion of words that are regular



(b) Proportion of tokens that are regular

To pass my final evaluation criterion, the changes must exhibit an S-curve. Figure 4 demonstrates that this is generally satisfied for the cases of regularization. Figure 3 demonstrates that this is also satisfied at the level of inflectional change, namely in charting the rise of the weak inflection. These dynamics have entirely disparate causes, so let us consider them in turn. I believe

the S-curve in individual word change is driven by the heterogeneity of vocabulary. This is because vocabulary size determines the effect hearing a word has on the likelihood of imitation. An additional token is proportionately more significant in the mind of a low vocabulary agent and therefore has a greater influence on the way the agent speaks in the future. Mutant inflections are therefore quick to dominate the mind of the lowest vocabulary agents and affect higher vocabulary agents in turn. The S-curve therefore reflects the bell-shaped distribution of vocabulary. Thus whilst memory and regularizing tendency dictate the general *direction* of change, the vocabulary distribution determines its *speed*. This mechanism is akin to that which Rogers (2010) identified as driving the diffusion of technological innovations. He segmented a bell-shaped population distribution into “innovators”, “early adopters”, “early majority”, “late majority” and “laggards” and demonstrated that if technology is adopted sequentially it will propagate in a logistic (S-shaped) fashion.¹⁸ In the case of regularization, Rogers’ ‘innovators’ are analogous to my defaulters.¹⁹ Regularization is therefore initiated by defaulters, and pervades speech according to an agent’s vocabulary.

The mechanism behind the inflectional S-curve has already been alluded to. The slow start and acceleration is the result of the self-intensifying regularizing tendency. Observe the situation after 500,000 iterations. With likely only one regular word in their lexicon, agents cast a very wide phonetic net before resorting to the default. By the 550,00th there are three regular words (corresponding to the red, yellow, purple lines in Figure 4). The regularizing tendency is therefore stronger and agents search fewer words before resorting to the default, leading some agents who previously drew upon phonetic neighbours to inflect the turquoise word now resorting to the default. These pivotal agents catalyse the regularization of the turquoise word, which in turn intensifies the global trend of regularization. Regularization is stoked further by the fact that the weak inflection is not produced only through default but also holds considerable analogical force at this point. As illustrated in Figure 6, this “wildfire-like” process plateaus at around 800,000 conversations after all that remains irregular are ‘islands of resilience’ and ‘tenacious trees’. Only stochasticity stops the system from stabilizing altogether. Especially destabilizing is the random exit and entrance of words, which by the end of the simulation contributes as much to global regularization than

¹⁸ One way to see why this is is to consider the cumulative probability distribution function that results from a bell-shaped frequency distribution.

¹⁹ By referring to agents in sufficiently abstract terms as to avoid taking sides in the much dichotomized debate on whether the source of linguistic variation lies in language use or acquisition. Innovators are simply ‘low vocabulary agents’ which could proxy for adults, teenagers, children, adult language learners, males, females, migrants, locals etc. All instances of language use are therefore seen as candidates for innovation.

endogenous regularization, consistent with the current state of English verbs (Cuskley et al. (2014)).

In summary, the selected run is consistent with all six evaluation criteria. As I have argued, this relies on memory and the regularizing tendency being bounded as follows. For change to be present, coherent and gradual, memories should not be so strong as to make morphology redundant; neither should they be so weak as to nullify the status quo. For change to be multidirectional, the regularizing tendencies should not be so strong as to immediately level all irregularity; neither should they be so weak as to as to make all change regular. And for the changes to follow realistic dynamics, vocabulary should exhibit a bell-shaped curve and the regularizing tendency should increase with the number of regular words in an agent’s vocabulary. If these conditions are met, so too are the evaluation criteria. Since the conditions seem to bound what can only realistically be expected, I conclude that the threefold thought process - memory, analogy and default - is sufficient to explain all six diachronic patterns in inflectional morphology.

7 Methodological Analysis

I do not have the space to defend against all challenges to the validity of my results, but I shall address one criticism that looms large in computational linguistics: that the desired results are built into the model. In my case, I take this criticism to mean that my evaluation criteria follow trivially from my assumptions. Rather than defend against this on a case by case basis I believe this criticism can be deflected altogether with the realization that my assumptions and evaluation criteria are taken from two entirely different disciplines. The assumptions are adapted from existing models in psycholinguistics; the evaluation criteria are assembled from research in historical linguistics. This means that the assumptions have been constructed without the evaluation criteria in mind and for a fundamentally different type of dataset (see van Noord (2015)). The ontologies to which they pertain are therefore non-overlapping, and so their logical relationship cannot be trivial. Indeed the very point of my emergentist approach is to see if *cognitive* assumptions can produce *diachronic* trends.

The emergentist approach is also well-suited to synchronic as well as diachronic analysis. My model lends itself especially naturally to the field of comparative linguistics. For example, it would be interesting to test its consistency with the “linguistic niche hypothesis” (Dale Lupyán (2012)) by varying the number of agents. Demographic theories behind the grammatical regularity of

English could also be tested by varying the variance of vocabulary, modelling the presence of (low vocabulary) second language learners. The model could also be used to analyse intra- as well as inter-language differences in inflectional morphology. The regularity of plurals compared to conjugations, for example, might be explained in terms of the number and frequency of nouns compared with verbs, which could be modelled by varying the number and frequency distribution of words and their effect on memory. Finally, there are countless ways in which the model could be extended. I have already suggested that phonological change may be necessary to fully account for the extent of regularization and irregularization. Population structure may also enrich the dynamics, if for example the likelihood that agents engage in conversation is based on their vocabularies. I have attempted to keep assumptions to a minimum in order to analyse the explanatory power of specifically cognitive assumptions. Extensions ought to proceed in light of my explanatory successes and failures.

8 Conclusion

In this paper I have attempted to explain six stylized facts of inflectional morphology as the emergent properties of cognitive constraints and biases. In my model, mutant inflections arise whenever an agent fails to recall an inflection. And since agents struggle to recall inflections affixed to rare and irregular words, mutation uproots all but the (high frequency) ‘tenacious trees’ and (phonetically consistent) ‘islands of resilience’. Variants are then selected by ‘analogy’ and ‘default’, which serve to perpetuate local and global regularities, hence explaining the direction of morphological change. Thus if my model is realistic, morphological change is modulated by memory and moulded by analogy and default.

Since it would be circular to argue that analogy and default both explain and can be explained by language, I have suggested that they are both facets of the inductive instinct animals evolved to make sense of the natural world. The same regularizing mechanisms that we evolved to patch up our visual, oral and other blind spots are therefore also responsible for patching up our linguistic memory. And whilst visual regularity may be just an illusion, the belief in linguistic regularity is self-fulfilling, for what people believe is true of the linguistic world comes to define morphological reality. Inflectional morphology has therefore been shaped by the brain, which has in turn been shaped by nature. It is therefore no surprise that linguistic structures are apt for metaphor: it is

precisely *because* we are used to finding the same trees on the same island that phonetic islands of the same inflection form. Morphology evolves to mirror ecology, and that explains why humans can grasp grammatical cases before they can tie their laces.

References

Baxter, Gareth J., et al. "Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English." *Language Variation and Change* 21.02 (2009): 257-296.

Baxter, Gareth J., et al. "Utterance selection model of language change." *Physical Review E* 73.4 (2006): 046118.

Blythe, Richard A. "Neutral evolution: a null model for language dynamics." *Advances in Complex Systems* 15.03n04 (2012): 1150015.

Blythe, Richard A. "Symmetry and Universality in Language Change." *Creativity and Universality in Language*. Springer International Publishing, 2016. 43-57.

Blythe, Richard A., and William Croft. "S-curves and the mechanisms of propagation in language change." *Language* 88.2 (2012): 269-304.

Bybee, Joan. *Language, use and cognition*. CUP, Cambridge (2010).

Bybee, Joan. *Language change*. Cambridge University Press, 2015.

Bybee, Joan L. "Usage-based theory and exemplar representations of constructions." The Oxford handbook of construction grammar. 2013.

Bybee, Joan L., and Paul J. Hopper, eds. *Frequency and the emergence of linguistic structure*. Vol. 45. John Benjamins Publishing, 2001.

Carroll, Ryan, Ragnar Svaré, and Joseph C. Salmons. "Quantifying the evolutionary dynamics of German verbs." *Journal of Historical Linguistics* 2.2 (2012): 153-172. Chomsky, Noam. "Rules and representations." *Behavioral and brain sciences* 3.01 (1980): 1-15.

Christiansen, Morten H., and Nick Chater. "Language as shaped by the brain." *Behavioral and brain sciences* 31.05 (2008): 489-509.

Christiansen, Morten H., Nick Chater, and Peter W. Culicover. *Creating language: Integrating evolution, acquisition, and processing*. MIT Press, 2016.

Coussé, Evie, and Ferdinand von Mengden, eds. *Usage-based approaches to language change*. Vol. 69. John Benjamins Publishing Company, 2014.

- Croft, William. *Explaining language change: An evolutionary approach*. Pearson Education, 2000.
- Cuskley, Christine F., et al. "Internal and external dynamics in language: evidence from verb regularity in a historical corpus of English." *PloS one* 9.8 (2014): e102882.
- Dale, Rick, and Gary Lupyan. "Understanding the origins of morphological diversity: The linguistic niche hypothesis." *Advances in Complex Systems* 15.03n04 (2012): 1150017.
- Darwin, Charles. *The descent of man, and selection in relation to sex*. Vol. 1. Murray, 1888.
- Dennett, Daniel C. "Darwin's dangerous idea." *The Sciences* 35.3 (1995): 34-40.
- Fertig, David. *Analogy and morphological change*. Edinburgh University Press, 2013.
- Fertig, David. *Are strong verbs really dying to fit in?*. 2009.
- Fertig, David. *Spreading like wildfire: Morphological variation and the dynamics of the Great English Verb Regularization*. 2016
- Garrett, Andrew. "Paradigmatic uniformity and markedness." *Explaining linguistic universals: Historical convergence and universal grammar* (2008): 125-143.
- Ghanbarnejad, Fakhteh, et al. "Extracting information from S-curves of language change." *Journal of The Royal Society Interface* 11.101 (2014): 20141044.
- Glushko, Maria. "Towards the quantitative approach to studying evolution of English verb paradigm." *Nordlyd* 31.1 (2004).
- Hadikin, Glenn. "Lexical selection and the evolution of language units." *Open Linguistics* 1.1 (2015): 458-466.
- Hare, Mary, and Jeffrey L. Elman. "Learning and morphological change." *Cognition* 56.1 (1995): 61-98.
- Hull, David L. "A mechanism and its metaphysics: An evolutionary account of the social and conceptual development of science." *Biology and Philosophy* 3.2 (1988): 123-155.
- Hume, David. *A treatise of human nature*. Courier Corporation, 2012.
- Hume, David. *Dialogues concerning natural religion*. William Blackwood, 1907.
- Johnson, Keith. "*The History of Early English: An Activity-based Approach*." Routledge (2016)
- Joseph, Brian D. "Morphologization from syntax." *The handbook of historical linguistics* 472 (2003): 492.
- Knooihuizen, Remco, and Oscar Strik. "Relative productivity potentials of Dutch verbal inflection patterns." *Folia Linguistica* 35.1 (2014): 173-200.

- Landsbergen, Frank. *Cultural evolutionary modeling of patterns in language change: exercises in evolutionary linguistics*. Netherlands Graduate School of Linguistics, 2009.
- Lieberman, Erez, et al. "Quantifying the evolutionary dynamics of language." *Nature* 449.7163 (2007): 713-716.
- Lounsbury, Thomas Raynesford. *The standard of usage in English*. Harper Brothers, 1908.
- Mailhammer, Robert. "Islands of resilience: the history of the German strong verbs from a systemic point of view." *Morphology* 17.1 (2007): 77-108.
- Nowak, Martin A., David C. Krakauer, and Andreas Dress. "An error limit for the evolution of language." *Proceedings of the Royal Society of London B: Biological Sciences* 266.1433 (1999): 2131-2136.
- Petersen, Alexander M., et al. "Statistical laws governing fluctuations in word use from word birth to word death." *Scientific reports* 2 (2012).
- Pijpops, Dirk, Katrien Beuls, and Freek Van de Velde. "The rise of the verbal weak inflection in Germanic. An agent-based model." *Computational Linguistics in the Netherlands Journal* 5 (2015): 81-102.
- Pinker, Steven, and Michael T. Ullman. "The past and future of the past tense." *Trends in cognitive sciences* 6.11 (2002): 456-463.
- Pinker, Steven, and Paul Bloom. "Natural language and natural selection." *Behavioral and brain sciences* 13.04 (1990): 707-727.
- Pinker, Steven. "Words and rules." *Lingua* 106.1 (1998): 219-242.
- Pinker, Steven. *Language learnability and language development, with new commentary by the author*. Vol. 7. Harvard University Press, 2009.
- Pinker, Steven. *The stuff of thought: Language as a window into human nature*. Penguin, 2007.
- Rogers, Everett M. *Diffusion of innovations*. Simon and Schuster, 2010.
- van Noord, Rik, and Jennifer K. Spenader. "Modeling the learning of the English past tense with memory-based learning." *Computational Linguistics in the Netherlands (CLIN)*, Antwerp 6 (2015).
- Zipf, George Kingsley. *"Selected studies of the principle of relative frequency in language."* (1932).