Proceedings of the Society for Computation in Linguistics

Volume 3 Article 40

2020

Stop the Morphological Cycle, I Want to Get Off: Modeling the Development of Fusion

Micha Elsner
The Ohio State University, melsner0@gmail.com

Martha B. Johnson
The Ohio State University, johnson.6713@osu.edu

Stephanie Antetomaso
The Ohio State University, antetomaso.2@osu.edu

Andrea D. Sims
The Ohio State University, sims.120@osu.edu

Follow this and additional works at: https://scholarworks.umass.edu/scil

Part of the Computational Linguistics Commons, and the Morphology Commons

Recommended Citation
DOI: https://doi.org/10.7275/33wn-nb03
Available at: https://scholarworks.umass.edu/scil/vol3/iss1/40

This Paper is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Proceedings of the Society for Computation in Linguistics by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
Stop the Morphological Cycle, I Want to Get Off: Modeling the Development of Fusion

Micha Elsner, Martha Booker Johnson, Stephanie Antetomaso, Andrea D. Sims
Department of Linguistics
The Ohio State University
elsner.14, johnson.6713, antetomaso.2, sims.120@osu.edu

Abstract

Historical linguists observe that many fusional (unsegmentable) morphological structures developed from agglutinative (segmentable) predecessors. Such changes may result when learners fail to acquire a phonological alternation, and instead, “chunk” the altered versions of morphemes and memorize them as underlying representations. We present a Bayesian model of this process, which learns which morphosyntactic properties are chunked together, what their underlying representations are, and what phonological processes apply to them. In simulations using artificial data, we provide quantitative support to two claims about agglutinative and fusional structures: that variably-realized morphological markers discourage fusion from developing, but that stress-based vowel reduction encourages it.

1 Introduction

While modern typologists reject the wholesale categorization of languages as isolating, agglutinative or fusional (Haskelmath, 2009), they still recognize a distinction between morphological structures which can be easily segmented and those which cannot (Plank, 1999). In ones with morphological fusion (or cumulation), multiple morphosyntactic properties (MSPs)1 are realized by a single morph with no immediately segmentable pieces.2 For instance, Turkish tarla-lar-ı and Old English feld-a both indicate ‘field-PL.ACC’ (Plank, 1999), but the Old English suffix cannot be further analyzed whereas the Turkish word has separate number and case morphemes.

Along with this taxonomic distinction comes a historical origin story, sometimes called the morphological cycle (Hock and Joseph, 1996)[183]. Through processes of phonological reduction, independent function words become attached to content words as agglutinative inflections. Further phonological reduction or sound changes blur the boundaries between morphemes, leading to fusion. Finally, affixes may become so non-transparent that their association with MSPs is lost (demorphologization) at which point new function words may be recruited to replace them, beginning the cycle anew.

Morphological change is more various and more complicated than this simple story suggests, and this cycle isn’t the only way in which fusion can arise (Grüenthal, 2007; Igartua, 2015; Karim, 2019). However, it is one way that has been observed. In this paper we focus on the role of phonological processes in the transition between agglutination and fusion. Morphological reanalysis often results from an interaction between the phonology of a language and the learning mechanism. Specifically in this context, morphemes are most likely to fuse if the environments in which they occur, and the phonological processes triggered by those environments, are vulnerable to reanalysis, which is to say, to mis-learning. The question becomes: which kinds of phonological processes are likely to make morphological constructions vulnerable to reanalysis, and which are not?

In order to test the role that phonological processes play in making agglutinative structures vulnerable to reanalysis, we provide a formal learning model3 for morphological systems whose internal representations clearly distinguish between agglutination and fusion. The model extends Cot-

1We use morphosyntactic category to refer to sets of properties; cross-linguistically common categories are TENSE, PERSON, NUMBER, etc., and morphosyntactic properties are PRESENT, PAST, etc.
2Following practice in morphology, we use the term morph to refer to (only) the form part of a morpheme.
3Code and data at github.com/melsner/scil2019-fusion.
terell et al. (2015), learning a Bayesian model which maps from sets of MSPs to surface forms in three steps: selection of a morphological template, concatenation of underlying forms, and phonology. We validate the model by testing on a series of artificial languages. The model recovers the expected analyses for prototypically agglutinative or fusional languages; for languages which can be analyzed in either way, we demonstrate in the first study that those with variably-realized morphological markers (i.e. ones that are sometimes present, sometimes absent) are less likely to be learned as fusional. In a second study, we show that languages with stress-based vowel reduction are more likely to be learned as fusional. Our model thus provides quantitative support for previous observations that languages with large proportions of agglutinative structures also frequently have large numbers of variably-realized morphs (Plank, 1999) and vowel harmony rather than stress-based reduction (Zingler, 2018).

2 Related work

<table>
<thead>
<tr>
<th>Indo-European</th>
<th>Ancient Greek</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRS</td>
<td>AOR</td>
</tr>
<tr>
<td>1SG</td>
<td>-m-i</td>
</tr>
<tr>
<td>2SG</td>
<td>-s-i</td>
</tr>
<tr>
<td>3SG</td>
<td>*-t-i</td>
</tr>
</tbody>
</table>

Table 1: Partial set of Indo-European and Ancient Greek (‘give’) person-number forms in present indicative and aorist

We begin with a concrete example of the kind of morphological change we are describing. In some Indo-European (IE) athematic verbs, person and number were expressed cumulatively but tense was realized via a separate morpheme: -i for present active indicative and zero for aorist active indicative (Table 1). (These endings are re-constructed for IE but attested in Sanskrit.) However, sound changes between IE and Proto-Greek obscured the unity of the person-number morphs across present and aorist. For example, word-final [m] turned into [n] as a result of sound change, resulting in different 1SG forms in Ancient Greek.4 These changes led speakers to reanalyze the formerly separate morphemes as fused (Brian Joseph, p.c.): 1SG.PRS -mi vs. 1SG.AOR -n. This reanalysis is evidenced by the fact that in Aeolic dialects, speakers extended the athematic ending -mi to verbs that did not historically have it, giving, e.g., fili-‘1SG.PRS’ where fili is expected etymologically. The fact that -mi was extended as a single unit indicates that it had undergone fusion.

The reanalysis of the Greek suffixes was thus driven by sound changes that introduced phonological alternations, and in the process introduced ambiguity regarding the morphological structure. In the wake of these changes, speakers were faced with an analytic choice, e.g.: is there one 1SG morpheme -m plus a phonological rule, or different 1SG endings -mi and -n that also express tense?

The extent to which sound change leads agglutinative structures to be reanalyzed as fusional has recently been questioned (Haspelmath, 2018).5 Nonetheless, this kind of ambiguity between analyses at different levels of representation is often a driver of language change (Bybee, 1999) and phonological reduction of agglutinative structures is widely cited as a source of fusionality (Bybee, 1997; Igartua, 2015, among others). Just as phonological rules and categories can arise when low-level phonetic processes like assimilation are reanalyzed as phonological, so fusion can appear when the effects of phonological process are “baked in” to the morphological representations. Bybee (2002) summarizes the idea (with reference mostly to syntax) with the catchphrase: “Items that are used together fuse together.”

Both Heath (1998) and Zingler (2018) point out the implication that agglutinative constructions must have “barriers”—typological features which prevent them from becoming fusional.6 Zingler makes a specific proposal, that fixed (lexical) stress systems tend to encourage fusion, while vowel harmony discourages it. This builds on a typological observation: the kinds of phonological alternations that occur in agglutinative and fusional systems tend to differ, “...with vowel harmony tending to imply agglutination” (Plank, 1999)[310].7 Zingler argues that fixed stress leads

---

4Also, prior to Proto-Greek [t] deleted in some contexts, affecting the 3SG.AOR, and between Proto-Greek and attested Greek [t] → [s] (Brian Joseph, p.c.). Both were regular sound changes but had consequences for morphology.

5In fact, Haspelmath states (pp107-8) that “...we do not know how it is that robust inflectional patterns with cumulative and suppletive affixes arise”. Our paper offers a partial answer.

6The argument of Heath (1998) applies to the first (isolating-agglutinative) step of the cycle, rather than the second (agglutinative-fused) as discussed here: he suggests that established agglutinative systems grammaticalize independent function words into morphemes more quickly, due to their analogical similarity to existing morphemes.

7An anonymous reviewer questioned the basis for this
to reduction in unstressed syllables, which over time may lose their vowels, placing their consonants in new environments with varied phonological effects. Harmony, on the other hand, prevents the loss of vowels, while at the same time indicating that bound elements are part of the phonological word (since they undergo harmonic changes based on the word stem).\(^8\)

One question here has to do with the relationship between language-level and construction-level properties. From Haspelmath’s perspective, individual constructions may be agglutinative or fusional, but it is not clear that languages as a whole fall into cleanly defined types. However, Zingler’s proposal is rooted in phonology-morphology interactions. Phonological processes generally operate across a range of constructions in a language. Phonological properties are language-level and thus might be expected to have an across-the-board effect on morphological structure. Moreover, accumulation of effects on individual constructions may result in a disproportionate number of constructions of the same type (agglutinate, fusional, etc.) in a given language. In other words, there is no expectation that the ways constructions develop historically will be fully independent of each other. To the extent that the phonological context is the same for different morphological constructions, we might expect similar pressures in and outcomes of language change. While Zingler himself does not say so, his ideas stand as an implicit challenge to Haspelmath’s questioning of the validity of morphological types at the language level.

We argue below that the presence of variably-realized morphological marking is also a protective factor against fusion. Many agglutinative languages have position classes that are sometimes filled by an overt morph, and sometimes not; this is what we mean by ‘variably-realized’ morphological marking. Examples include morphosemantic markers such as causatives, desideratives or negatives, whose position class slots are filled only when that meaning occurs, and optional agreement marking (Plank, 1999). Polysynthetic languages, which are invariably mostly agglutinative, contain even more variably-realized elements, such as incorporated objects (Comrie, 1989). We suggest that, because variably-realized elements break up sequences of morphemes that would otherwise always appear next to one another, they render fusional analyses less appealing to the learner. Our argument not only explains the previous observation that variably-realized marking and agglutination correlate, but might also help to explain where and how fusionality develops.

Caballero and Kapatsinski (to appear) quantify fusionality in the Uto-Aztecan polysynthetic language Choguita Raramuri. They show that morphemes exhibit some fusion, especially close to the stem. Their research focus is similar to ours in examining how learners might infer morphological boundaries. However, their approach differs from our own in two ways. First, it provides a description of how much fusion is present based on the Naive Discriminative Learner (Baayen et al., 2011) and some variant models, but not a causal model of how language properties encourage or discourage fusion. Second, it lacks an explicit model of phonological rules. Caballero and Kapatsinski point out that if learners can mentally “undo” the effects of regular phonological rules, the Naive Learner will overestimate the degree of fusionality. The model we present below is designed to test causal mechanisms underlying the development of fusionality, and specifically the role of phonological rules.

---

\(^8\)Plank (1998)\(^9\) points out that this idea of vowel harmony ‘cementing’ the internal cohesion of agglutinative word structure goes back to Baudouin de Courtenay (1876), but is not unproblematic in its reasoning. In our work, nothing depends on vowel harmony creating greater word-internal cohesion.

---

---

Figure 1: Overall architecture of our model, consisting of three finite-state transducers, producing two intermediate layers of latent representation (in gray).

<table>
<thead>
<tr>
<th>Transducer 1: fusion</th>
<th>Transducer 2: lexicon</th>
<th>Transducer 3: phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input MSPs (M_1=1, M_3=1, \text{STEM}=1)</td>
<td>Abstract ms (M_1=1) (M_3=1, \text{STEM}=1)</td>
<td>Underlying mwi-mela</td>
</tr>
<tr>
<td>Surface form mwimela</td>
<td>Surface form mwimela</td>
<td>Surface form mwimela</td>
</tr>
</tbody>
</table>

---
3 Model

Our model is intended to capture the first stage of the transition from agglutination to fusion, in which the learner reanalyzes an ambiguous polymorphemic structure as monomorphic. This re-analysis is covert, affecting only the learner’s mental representation; in order for the system to become unambiguously fusional (i.e., for the change to become actualized, in the terminology of historical linguistics), the reanalyzed marker must generalize to other words, as we saw above for Greek, or undergo further diachronic changes. We leave modeling such changes for future work.

The model (Figure 1) formalizes our intuitions about agglutinative and fusional analyses of morphological systems. In order to do so, it represents morphemes as invariant underlying representations and applies phonological processes that transform them into surface forms. Because the popular sequence-to-sequence framework for inflection (Kann and Schütze, 2016) conflates these processes within a single neural network, we choose instead to extend an older model, Cotterell et al. (2015), in which these components are separate. While this model may be less capable overall, it is more interpretable in terms of the theoretical questions we are trying to answer.

Cotterell et al. model the correspondence between sequences of abstract morphemes and surface strings. The term “abstract morpheme” refers to a set of MSPs that already reflect the effects of fusion—in the context of agglutination, each abstract morpheme is a single MSP, whereas for fusion, the abstract morphemes bundle together many MSPs. The model maps abstract morphemes to surface strings in the following steps: first, each abstract morpheme is assigned an underlying phonological form; next, these forms are concatenated to yield an underlying inflected form; finally, this form is passed through a finite-state transducer which applies (stochastic) phonological rules. (Lines 2-4 of Figure 1.)

Our model differs from theirs primarily in adding a new initial step, which maps a sequence of atomic MSPs into a corresponding sequence of abstract morphemes. This is the step at which fusion occurs. For instance, a sequence STEM=give, NUM=PL, TENSE=PRS could be output as three separate symbols, or as STEM=give, NUM=PL|TENSE=PRS, where we use the | notation to indicate that two MSPs are fused into a single abstract morpheme. The model simplifies slightly by requiring uniformity at the level of morphosyntactic categories; in our illustrating example, either all combinations of number and tense MSPs would be fused or none would be.

For simplicity, we also modify the model so that it consists of a cascade of relatively small finite-state transducers (FSTs) (Mohri et al., 2002) which we can implement using the Carmel package (Graehl, 1997). This necessitates some changes and simplifications to the model, but allows us to use Carmel’s built-in Bayesian inference (Chiang et al., 2010) rather than belief propagation as in Cotterell et al. (2015).

As stated, the first transducer in the cascade maps a sequence of MSPs into a sequence of abstract morphemes (without specifying any phonological detail). For computational convenience, we make two simplifying assumptions: The input MSPs are provided in a fixed, templatic order (Stump, 1997), in which only contiguous subsequences can be fused. MSPs are not allowed to fuse with the stem (that is, there is no MSP-conditioned stem allomorphy), even though this occurs in real languages. The transducer (Fig 2) first chooses an allowable fusion template via epsilon transition and then deterministically transforms the input sequence.

The second transducer is a lexicon (Figure 3) which maps each abstract morpheme to a phonological underlying form. Cotterell et al. implement this as a distribution of point masses on strings, which is intractable and must be approximated. We use a simpler solution which is finite-
state and tractable. Each word in the lexicon has an initial state with two outgoing epsilon transitions; one leading back to the start state (thus producing a null morpheme) and another leading to a linear chain of 15 states. Each state in the chain can produce any non-null character, or transition back to the start. This transducer can produce any string up to 15 characters long; the posterior tends to concentrate around a single underlying form per morpheme. We set the prior odds ratio for the two initial transitions so that the null morpheme is 100 times more likely a priori than the linear chain. This prior biases the model toward parsimonious analyses with smaller morpheme inventories, provided they can satisfactorily account for the data.

The third transducer (Figure 4) implements phonological rules. While Cotterell et al. supply a full finite-state phonology (Riggle, 2004; Hayes and Wilson, 2008, and others), in our experiments below, we use a custom machine implementing only the specific rules which actually exist in our artificial language. However, the machine executes the rules non-deterministically; the system must learn the true probability with which the rules occur. Again, we use prior parameters to determine how much evidence is necessary to convince the system that a phonological rule is justified. In our experiments below, we set the prior odds ratio of the rule applying to 1:100. In simulation C, we vary the strength of the prior (by multiplying the prior counts by a constant \( \alpha \)) and report results as a function of this parameter.

We perform posterior inference using blocked Gibbs sampling (Chiang et al., 2010). For each language, we run 20 Markov chains with random starting points, annealing linearly from temperature 4 to 1 over 200 iterations. We average the final counts from each chain to obtain the posterior.

4 Case study 1: Variably-realized marking

In this section, we run a series of simulations on artificial languages, intended to be reminiscent of the Bantu language Kihehe, spoken in Tanzania (Lewis, 2009). Simulations \( A \) – \( B \) show that the model can learn both agglutinative and fusional systems; \( C \) shows that the model’s preference for fusionality is dependent on the phonological prior weight \( \alpha \). \( D \) gives the main conclusion, that the presence of a variably-realized marker between two obligatory ones can block the emergence of fusion.

We first give a brief overview of Kihehe itself. Kihehe verbs are marked for person-number agreement with the subject; the form of the agreement marker reflects the noun class of the subject. This marker is sometimes followed by a tense marker. Although Kihehe has morphemes which begin with vowels, its phonological rules act to prevent onsetless syllables from surfacing, by transforming the first vowel in a VV sequence into a glide, or deleting one vowel, and in both cases, lengthening the remaining vowel (Odden and Odden, 1999).\(^{11}\) This creates a system in which agreement and tense markers are arguably fused on the surface (Table 2). In 3SG and 3PL, where vowel deletion occurs, segmentation is impossible. In the other cells, segmentation of the surface form is possible but gliding prevents postulation of a single, invariant form of each agree-

---

\(^{11}\)We present these phonological processes here as SPE rules, although of course other theoretical frameworks like OT could derive the same results.

Table 2: Conjugation of a Kihehe verb in the present tense (Johnson, 2015).

<table>
<thead>
<tr>
<th>Underlying</th>
<th>Surface</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndi-i-ko:mala</td>
<td>ndi:ko:mala</td>
<td>“I am sitting”</td>
</tr>
<tr>
<td>u-i-ko:mala</td>
<td>wi:ko:mala</td>
<td>“You, SG are sitting”</td>
</tr>
<tr>
<td>a-i-ko:mala</td>
<td>i:ko:mala</td>
<td>“S/he is sitting”</td>
</tr>
<tr>
<td>tu-i-ko:mala</td>
<td>twi:ko:mala</td>
<td>“We are sitting”</td>
</tr>
<tr>
<td>mu-i-ko:mala</td>
<td>mwi:ko:mala</td>
<td>“You, PL are sitting”</td>
</tr>
<tr>
<td>va-i-ko:mala</td>
<td>vi:ko:mala</td>
<td>“They are sitting”</td>
</tr>
</tbody>
</table>
Table 3: Morphophonology of four simulated languages (case study 1).

<table>
<thead>
<tr>
<th>Name</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>Phonology</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{A}$</td>
<td>${ta, ko, he}$</td>
<td>$-$</td>
<td>${i, a}$</td>
<td>$-$</td>
<td>koimela, muimela</td>
</tr>
<tr>
<td>$\mathcal{B}$</td>
<td>${mu, gu, si}$</td>
<td>${y}$</td>
<td>${de}$</td>
<td>$-$</td>
<td>dunumela, yamela</td>
</tr>
<tr>
<td>$\mathcal{C}$</td>
<td>as $\mathcal{A}$</td>
<td>$-$</td>
<td>as $\mathcal{A}$</td>
<td>as $\mathcal{C}$</td>
<td>mwwimela ($&lt; mu-i-mela$), kamela ($&lt; ko-a-mela$)</td>
</tr>
<tr>
<td>$\mathcal{D}$</td>
<td>as $\mathcal{A}$</td>
<td>${sa}$</td>
<td>as $\mathcal{A}$</td>
<td>as $\mathcal{D}$</td>
<td>mwwimela ($&lt; mu-i-mela$), musimela ($&lt; mu-sa-i-mela$)</td>
</tr>
</tbody>
</table>

Figure 5: Probability of fusion in $\mathcal{A}$ vs $\mathcal{B}$. 

As in Kihehe, our artificial languages have stems made up of CV syllables. We use an inventory of 5 vowels and 15 consonants; for each language, we generate 200 unique random stems, with length $\min(5, \text{Geom}(0.5))$, which we use to create a corpus of 1000 inflected forms. Each language has two required morphosyntactic categories, $M_1$ and $M_3$ (e.g., person and tense), realized as prefixes, with uniformly distributed values (MSPs). In simulation $\mathcal{D}$, we explore the impact of a variably-realized category $M_2$ which appears between the two. Table 3 shows the realizations of $M_1$, $M_2$ and $M_3$ and the phonology in each simulation.

Language $\mathcal{A}$ is prototypically agglutinative. Each category:property (MSP) pair licenses a unique, segmentable morph in the surface string. (The morphs that realize $M_1$ contain equal numbers of high and low vowels, and for $M_3$ contain equal numbers of vocalic and consonantal onsets.) Language $\mathcal{B}$ is prototypically fusional. While words inflect for the same categories as in language $\mathcal{A}$, each $M_1, M_2$ value pair licenses a unique morph that realizes both categories (a sampled string of one or two syllables). We expect the model to analyze $\mathcal{A}$ as agglutinative, due to the prior preference for a small morpheme inventory (the agglutinative analysis has $6+4=10$ morphemes while the fusional analysis has $6*4=24$), and $\mathcal{B}$ as fusional; this is the actual result (Figure 5).

Language $\mathcal{C}$ has the same underlying properties as language $\mathcal{A}$, but is subject to phonological rules which result in non-isomorphic relationships between form and meaning in the surface forms. (The surface prefixes are thus segmentable, but not into invariant forms; for example, ko- alternates with k- and mu- with mw-, conditioned on their phonological environment.) We use language $\mathcal{C}$ to explore the effects of the prior parameter $\alpha$, which encodes our bias against using the phonological rule; larger $\alpha$ means that more evidence is required to justify the rule’s existence. Not surprisingly, small $\alpha$ leads to agglutinative analyses, while large $\alpha$ leads to fusion (Figure 6, top).

Finally, we investigate the effects of $M_2$, a variably-realized category between $M_1$ and $M_3$, using language $\mathcal{D}$. For this simulation, we set $\alpha = 1000$, a setting which we found in the previous experiment would result in a fusional analysis. We do so because we are interested in whether $M_2$ can prevent fusion from occurring; thus, it makes sense to start from a setting in which fusion is expected. All versions of language $\mathcal{D}$ have the category $M_2$ between $M_1$ and $M_3$, but we vary the
probability with which it takes its non-zero value (realized as sa-). We find (Figure 6, bottom) that when sa- always or never occurs, the posterior mode is a fully fused system, \(M_1|M_2|M_3\). But when sa- is variably realized, full fusion essentially never occurs. Instead, we find either agglutination (\(M_1-M_2-M_3\), the plurality outcome when \(p(\text{sa}) = .25\)) or partial fusion, in which \(M_2\) is realized jointly with one of its neighbors.

Thus, the important result is that in the context of phonological rules that create surface-ambiguous word-forms, variably-realized morphemes decrease the likelihood of agglutinative morphemes being reanalyzed as fusional.

5 Case study 2: Stress-based vowel reduction

Our next study addresses Zingler’s claims about Turkish agglutination. Zingler argues (p422) that languages have various mechanisms for articulatory reduction of vowels. One of these is vowel harmony, which replaces some distinctive features of a vowel with those of its neighbor, and another is durational reduction, which reduces a vowel’s absolute length, and tends to erode its features by centralizing it. These mechanisms are complementary; harmony correlates with syllable-timed languages and with systems that assign stress to a fixed syllable relative to the word boundary. Durational reduction correlates with stress-timed languages and with systems in which the stressed syllable is lexically determined. Zingler’s hypothesis is that durational reduction leads to fusion, and that vowel harmony, as an alternative way to ease articulation without durational reduction, is what prevents Turkish from becoming fusional.\(^{12}\)

In this section, we validate Zingler’s claim that vowel reduction tends to encourage the development of fusion, and add to his idea by showing that this is especially so when the position of reduction is predictable within particular morphemes. Simulation \(\mathcal{E}\) investigates the case where stress is predictable within morphemes, and \(\mathcal{F}\) the case where it is not. As above, we use artificial languages in which stems consist of CV syllables. Languages in this section have two required categories, \(M_1\) and \(M_2\), realized as suffixes. Table 1 shows the realizations of \(M_1\) and \(M_2\) and the phonology.

We next apply vowel reduction. In simulation \(\mathcal{E}\) we apply final stress and then alternate strong and weak syllables moving left; the reduction rule deletes each weak vowel with some probability.\(^{13}\) So, the word \(\text{dite-ko-de}\) in fully reduced form would become \(\text{dtekde}\). Simulation \(\mathcal{F}\) is similar, but with initial stress, so \(\text{dite-ko-de}\) would become \(\text{ditkod}\). Such stress rules follow from the core predictions of metrical stress theory (Hayes, 1995).\(^{14}\) Within each simulation, we compare languages with varying rates of reduction, ranging from no reduction to all unstressed vowels reduced.

Although neither \(\mathcal{E}\) nor \(\mathcal{F}\) has a true lexical stress system, the varying stress rules have implications for the predictability of stress placement.

\(^{12}\)Zingler also argues that vowel harmony helps maintain a morpheme minimality criterion. He does not consider whether morpheme minimalism plays a role in preventing fusion, but we believe this could also be relevant and could be simulated in our model, with suitable alterations to the lexicon. But we leave doing so for future work.

\(^{13}\)This approximates the ‘fall of the jers’, a sound change in the history of the Slavic languages (Kiparsky, 1979).

\(^{14}\)Kager (1995) gives example languages which have the stress systems described here. Weri parses feet from right-to-left, with final stress; Hungarian parses from left to right with initial stress.
on morphemes. Because each word has two obligatory suffixes, final stress (simulation $E$) means that the second suffix, corresponding to $M_2$, will always be pronounced with a full vowel, while the suffix for $M_1$ will be probabilistically reduced. The same condition would hold in a true lexical stress language, although in such a language it would also hold if the number of suffixes were variable. In $F$, however, stress lands on the suffix realizing $M_1$ when the length of the stem is even, on the suffix for $M_2$ when it is odd. Thus, each suffix appears in both strong and weak positions.

Vowel reduction disrupts the original CV structure of our languages, allowing consonant clusters to appear on the surface. It is extremely common for such clusters to simplify for articulatory reasons (Brohan and Mielke, 2018) — we apply only one simplification rule, progressive voicing assimilation. Thus, $dtekde$ would surface as $ddekte$. In a real language, we might expect further simplifications to apply to prevent, for instance, geminate $dd$ at the beginning of a word; for our purposes, however, a single assimilation rule is sufficient.

We apply the same learning procedure as in the previous section. The feature and lexicon transducers are unchanged. The transducer for vowel reduction is shown as Figure 7; the transducer for assimilation resembles the one in Figure 4. We use $\alpha = 1000$ as a bias parameter to penalize both phonological rules (vowel reduction and consonant assimilation).

Figure 8 (top) shows the results for language $E$. With reduction rate 0 (no reduction), the posterior mode is an agglutinative system. Optional vowel reduction (25-75%) produces mixed systems in which both agglutination and fusion are recognized as possible analyses, although the posterior probability of fusional analyses climbs slightly as reduction increases. With 100% reduction, the posterior strongly prefers fusion.

The orange line shows the posterior probability of vowel reduction. The system always underestimates the true probability of reduction — when the true probability is 50%, for instance, the posterior is only 20% — and counterintuitively, learns that reduction is absent when its true probability is 100%. This reflects the influence of the prior bias against the reduction rule, but also the fact that the system learns some cases of reduction as variant lexical items. Table 5 shows one Markov chain’s final learned representations for two values of $M_1$ (-ta) and $M_2$ (-de) as a function of reduction rate. With no reduction, the system learns only agglutinative analyses; intermediate systems learn underlying forms for both fused and unfused morphemes, including multiple variant forms of each one. The system with 100% reduction learns only a fused morpheme, -tte, which incorporates the result of both vowel reduction and assimilation. With no evidence for an overt vowel between the $t$s, the system has no reason to learn the rule.

Figure 8 (bottom) shows the results for lan-
Table 5: Underlying forms learned for two morphemes in variants of language $\mathcal{F}$. First entry is the posterior mode, (parentheses) show alternatives with $p > .01$.

| Rate | $M_1$=I | $M_2$=III | $M_1$=I|$M_2$=III |
|------|---------|-----------|-------------|
| 0    | ta      | de        | -           |
| 25   | ta (t, te) | de (te)   | tade        |
| 50   | ta (t)   | te (de)   | tade        |
| 75   | ta (t, t) | te (de)   | tte         |
| 100  | -        | -         | tte         |

As predicted, the probability of fusion increases again with the rate of reduction, but the results are less extreme, since stress placement on the suffixes varies depending on the stem. For this language, agglutination is always the plurality outcome, but intense reduction increases the probability that some fusional analyses will be produced.

Returning to Zingler’s argument, Turkish is similar to the case in which the probability of reduction is 0, a case which in our simulations is indeed strongly agglutinative. Because Turkish is syllable-timed and has vowel harmony, it is unlikely to develop the alternate pattern of stress-timing and durational reduction which Zingler argues could lead it to develop more fusion. We have shown that stress-timing and durational reduction does favor fusional analyses. It is tempting to speculate that the same argument might help to explain the differences between Finnish (vowel harmony and agglutination) and Estonian (no harmony and limited fusion); Estonian historically had a more agglutinative structure. In particular, we note that Estonian has word-initial stress (Lippus et al., 2014), which simulation $\mathcal{F}$ shows is predictive of a mixed rather than entirely fusional system.

6 Conclusion

Our results show that, at least in principle, pre-existing typological features can help to determine whether an agglutinative construction evolves into a fusional one, or remains stable. In particular, we present firm evidence that variably-realized marking makes fusion less likely while durational vowel reduction has the opposite effect. While authors like Plank (1999) have listed many independent features or elements which characterize prototypically “fusional” morphology, these have typically been discussed as typological clusters, without necessarily providing a causal explanation. Our modeling results give a mechanism in which some of these features precede, and give rise to others.

A variety of researchers have noted (Greenberg, 1966) and attempted to discover (Murawaki, 2018; Bjerva et al., 2019) correlations between typological features. Harris (2008) suggests that in many cases, such correlations reflect precisely this kind of historical mechanism—the likelihood that a language will develop in some typological direction is dependent on the features it already has, some of which may encourage a particular change while others tend to reinforce existing patterns. While the simulations presented here use artificial data, we hope to apply this model to real corpus data from languages in which fusion might be developing, in order to isolate particular changes in the phonology as the “triggers” of ongoing morphological change, or explain distributionally why one set of morphemes appears more fusional than another. In doing so, we can discover how theoretical explanations of language change, such as the morphological cycle, might be realized in the minds of language users.

Acknowledgments

We thank the members of the Fall 2018 Seminar on Models of Morphological Learning and Change at OSU for their comments on preliminary versions of this project, Brian Joseph for discussion of historical examples, as well as three anonymous reviewers.

References


