

Morphological Complexity Outside of Universal Grammar

Jirka Hana & Peter W. Culicover

April 28, 2006

Abstract

There are many logical possibilities for marking morphological features. However only some of them are attested in languages of the world, and out of them some are more frequent than others. For example, it has been observed (Sapir, 1921; Greenberg, 1957; Hawkins & Gilligan, 1988) that inflectional morphology tends to overwhelmingly involve suffixation rather than prefixation. This paper proposes an explanation for this asymmetry in terms of acquisition complexity. The complexity measure is based on the Levenshtein edit distance, modified to reflect human memory limitations and the fact that language occurs in time. This measure produces some interesting predictions: for example, it predicts correctly the prefix-suffix asymmetry and shows mirror image morphology to be virtually impossible.

1. Background

We address here one aspect of the question of why human language is the way it is. It has been observed (Sapir, 1921; Greenberg, 1957; Hawkins & Gilligan, 1988) that inflectional morphology tends overwhelmingly to be suffixation, rather than prefixation, infixation, reduplication or other logical possibilities that are quite rare if they exist at all. For this study, we presume that the statistical distribution of possibilities is a consequence of how language is represented or processed in the mind. That is, we rule out the possibility that the distributions that we find are the result of

contact, genetic relatedness, or historical accidents (e.g. annihilation of speakers of languages with certain characteristics), although such possibilities are of course conceivable and in principle might provide a better explanation of the facts than the one that we presume here.

The two possibilities that we focus on concern whether the preference for suffixation is a property of the human capacity for language *per se*, or whether it is the consequence of general human cognitive capacities. Following common practice in linguistic theory, let us suppose that there is a part of the human mind/brain, called the Language Faculty, that is specialized for language (see e.g. Chomsky, 1973). The specific content of the Language Faculty is called Universal Grammar. We take it to be an open question whether there is such a faculty and what its specific properties are; we do not simply stipulate that it must exist or that it must have certain properties, nor do we deny its existence and assert that the human capacity for language can be accounted for entirely in terms that do not appeal to any cognitive specialization. The goal of our research here is simply to investigate whether it is possible to account for a particular property of human language in terms that do not require that this property in some way follows from the architecture of the Language Faculty.

1.1. Types of inflectional morphology

Inflectional morphology is the phenomenon whereby the grammatical properties of a word (or phrase) are expressed by realizing the word in a particular form taken from a set of possible forms. The set of possible forms of a word is called its **paradigm**.¹ A simple example are the English nominal paradigms distinguishing singular and plural. The general rule is that the singular member of the paradigm has nothing added to it, it is simply the stem, while the plural member has some

¹The word *paradigm* is used in two related, but different meanings: (1) all the forms of a given lemma; (2) in the original meaning, referring to a distinguished member of an inflectional class, or more abstractly to a pattern in which the forms of words belonging to the same inflectional class are formed. In this paper, we reserve the term paradigm only for the former meaning and use the phrase “paradigm pattern” for the latter.

variant of *s* added to the end of the stem.²

- (1) Singular: book patch tag
Plural: book·s patch·es tag·s

Other, more complex instances of inflectional morphology involve morphological case in languages such as Finnish and Russian, and tense, aspect, modality, etc. in verb systems, as in Italian and Navajo. For a survey of the various inflectional systems and their functions, see (Spencer & Zwicky, 1998).

It is possible to imagine other ways of marking plural. Imagine a language just like English, but one in which the plural morpheme precedes the stem.

- (2) Singular: book patch tag
Plural: s·book s·patch s·tag

Or imagine a language in which the plural is formed by reduplicating the entire stem –

- (3) Singular: book patch tag
Plural: book·book patch·patch tag·tag

– or a language in which the plural is formed by reduplicating the initial consonant of the stem and following it with a dummy vowel to maintain syllabic well-formedness.

- (4) Singular: book patch tag
Plural: be·book pe·patch te·tag

Many other possibilities come to mind, some of which are attested in languages of the world, and others of which are not. A favorite example of something imaginable that does not occur is that of pronouncing the word backwards. The pattern would be something like

- (5) Singular: book patch tag
Plural: koob tchap gat

²Throughout this paper, we mark relevant morpheme boundaries by ‘·’, e.g. *book·s*.

1.2. A classical example: Prefix-suffix asymmetry

Greenberg (1957) finds that across languages, suffixing is more frequent than prefixing and far more frequent than infixing. This tendency was first suggested by Sapir (1921). It is important that the asymmetry holds not only when simply counting languages, which is always problematic, but also in diverse statistical measures. For example, Hawkins and Gilligan (1988) suggest a number of universals capturing the correlation between affix position in morphology and head position in syntax. The correlation is significantly skewed towards preference of suffixes. For example, postpositional and head-final languages use suffixes and no prefixes; while prepositional and head-initial languages use not only prefixes, as expected, but also suffixes. Moreover, there are many languages that use exclusively suffixes and not prefixes (e.g. Basque, Finnish), but there are very few that use only prefixes and no suffixes (e.g. Thai, but in derivation, not in inflection). There have been several attempts to explain the suffix-prefix asymmetry, using processing arguments, historical arguments, and combinations of both.

1.2.1. Processing explanation

Cutler, Hawkins, and Gilligan (1985); Hawkins and Gilligan (1988) offer an explanation based on lexical processing. They use the following line of reasoning: It is assumed that lexical processing precedes syntactic processing and affixes usually convey syntactic information. Thus listeners process stems before affixes. Hence a suffixing language, unlike a prefixing language, allows listeners to process morphemes in the same order as they are heard. The preference is a reflection of the word-recognition process.

In addition, since affixes form a closed class that is much smaller than the open class of roots, the amount of information communicated in the same time is on average higher for roots than for affixes. Therefore, in a suffixing language, the hearer can narrow down the candidates for the

current word earlier than in a prefixing language. Moreover, often (but not always) the inflectional categories can be inferred from context.³

1.2.2. Historical explanation

Givón (1979) argues that the reason for suffix preference is historical. He claims that (1) bound morphemes originate mainly from free morphemes and that (2) originally all languages were SOV (with auxiliaries following the verb). Therefore verbal affixes are mostly suffixes since they were originally auxiliaries following the verb. However, assumption (2) of the argument is not widely accepted (see, for example, Hawkins & Gilligan, 1988, p. 310 for an opposing view). Moreover, it leaves out open the case of non-verbal affixes.

1.2.3. Processing & Historical explanation

Hall (1988) tries to integrate the historical explanation offered by Givón (1979) (§1.2.2) and the processing explanation by Hawkins and Gilligan (1988) (§1.2.1). He adopts Givón's claim that affixes originate mainly from free morphemes, but he does not need the questionable assumption about original SOV word-order; he uses Hawkins and Gilligan's argument about efficient processing to conclude that prefixes are less likely than suffixes because free morphemes are less likely to fuse in pre-stem positions.

Although the work above correctly explains suffix-prefix asymmetry, it has two disadvantages: (1) it relies on several processing assumptions that are not completely independent of the explained problem, (2) there are many other asymmetries in the distribution of potential morphological systems, (3) as stated above, it addresses only verbal morphology. In the rest of the paper, we

³For example, even though in free word-order languages like Russian or Czech it is not possible to predict case endings in general, they can be predicted in many specific cases because of agreement within the noun phrase, subject-verb agreement, semantics, etc.

develop an alternative measure that we believe addresses all of these issues.

2. Our approach

As noted, the question of why some possibilities are more frequent than others and why some do not exist has two types of answers, one narrowly linguistic and one more general. The linguistic answer is that the Language Faculty is structured in such a way as to allow some possibilities and not others, and the preferences themselves are a property of Universal Grammar. This is in fact the standard view in Mainstream Generative Grammar, where the fact that rules of grammar are constrained in particular ways is taken to reflect the architecture of the Language Faculty; the constraints are part of Universal Grammar (Chomsky, 1973; Wexler & Culicover, 1980) and prevent learners from formulating certain invalid hypotheses about the grammars that they are trying to acquire.

The alternative, which we are exploring in our work, is that the possibilities and their relative frequencies are a consequence of relative computational complexity for the learner of the language. On this view, morphological systems that are inherently more complex are not impossible, but less preferred. Relatively lower preference produces a bias against a particular hypothesis in the face of preferred competing hypotheses. This bias yields a distribution in which the preferred option is more widely adopted, other things being equal. See (Culicover & Nowak, 2002) for a model of such a state of affairs.

If we simply observe the relative frequencies of the various possibilities we will not be able to confirm the view that we have just outlined, because it relies on a notion of relative complexity that remains undefined. We run the risk of circularity if we try to argue that the more complex is less preferred, and that we know what is more complex by seeing what is less preferred, however relative preference is measured. Therefore, the problem that we focus on in this paper is that of developing a measure of complexity that will correctly predict the clear cases of relative preference, but that will also

be independent of the phenomenon. Such a measure should not take into account observations about preference *per se*, but rather formal properties of the systems under consideration. On this approach, if a system of Type I is measurably more complex than a system of Type II, we would predict that Type I systems would be more commonly found than Type II systems.

2.1. Complexity

We see basically two types of measures as the most plausible accounts of relative morphological complexity, learning and real-time processing. Simplifying somewhat, inflectional morphology involves adding a morpheme to another form, the stem. From the perspective of learning, it may be more difficult to sort out the stem from the inflectional morpheme if the latter is prefixed than if it is suffixed. The other possibility is a processing one: once all of the forms have been learned, it is more difficult to recognize forms and distinguish them from one another when the morphological system works a particular way, e.g. uses inflectional prefixes.

We do not rule out the possibility of a processing in principle, although we do not believe that the proposals that have been advanced (see §1.2) are particularly compelling or comprehensive. The types of measures that we explore here (see §4) are of the learning type.

2.2. Acquisition complexity – the dynamical component

We assume that the key determinant of complexity is the transparency or opacity of the morphological system to the learner. If we look at a collection of data without consideration of the task of acquisition, but just consider the overall transparency of the data, there is no apparent distinction between suffixation, prefixation, or a number of other morphological devices that can be imagined. However, language is inherently temporal, in the sense that expressions are encountered and processed in time. At the beginning of an unknown word, it is generally hard for a naïve learner to predict the entire form of the word. Given this, our question about relative complexity may be

formulated somewhat more precisely as follows: Assuming the sequential processing of words, how do different formal morphological devices contribute to the complexity of acquiring the language? The intuition of many researchers is that it is the temporal structure of language that produces the observed preference for suffixation; see §3 for a range of proposals. We adopt this insight and make it precise. In particular, we compute for all words in a lexicon their relative similarity to one another as determined by a sequential algorithm. Words that are identical except for a single difference are closer to one another if the difference falls towards the end of the words than if it comes at the beginning, a reflection of the higher processing cost to the learner of keeping early differences in memory versus the lower processing cost of simply checking that early identities are not problematic. We describe the algorithm in detail in §4 and justify some of the particular choices that we make in formulating it.

An important consequence of the complexity measure is that it correctly yields the desired result, i.e. that inflectional suffixation is less costly to a system than is inflectional prefixation. Given this measure, we are then able to apply it to cases for which it was not originally devised, e.g. infixation, various types of reduplication, and templatic morphology.

3. Relevant studies in acquisition and processing

In this section, we review several relevant studies.

3.1. Lexical Processing

A large amount of psycholinguistics literature suggests that lexical access is generally achieved on the basis of the initial part of the word:

- the beginning is the most effective cue for recall or recognition of a word, cf. (Nooteboom, 1981)

- word-final distortions often go undetected, cf. (Marslen-Wilson & Welsh, 1978; Cole, 1973; Cole & Jakimik, 1978, 1980)
- speakers usually avoid word-initial distortion, cf. (Cooper & Paccia-Cooper, 1980)

An example of a model based on these facts is the cohort model of Marslen-Wilson and Tyler (1980). It assumes that when an acoustic signal is heard, all words consistent with it are activated; as more input is being heard, fewer words stay activated, until only one remains activated. This model also allows easy incorporation of constraints and preferences imposed by other levels of grammar or real-world knowledge.

Similarly, as (Connine, Blasko, & Titone., 1993; Marslen-Wilson, 1993) show, changes involving non-adjacent segments are generally more disruptive to word recognition than changes involving adjacent segments.

3.2. The acquisition of word segmentation

Studies of word boundary acquisition are also very relevant, especially in the case of concatenative morphology. We think it is reasonable to assume that the learner can extrapolate at least some of the techniques used in word segmentation to morpheme segmentation. According to Jusczyk and Aslin (1995), the ability to segment a sound stream into words starts to develop around 6 to 8 months of age. Since fluent speech contains few pauses between adjacent words, children must rely on other cues to find word boundaries – stress, phonotactic constraints, degree of coarticulation and statistical structure. Saffran, Aslin, and Newport (1996) showed that 8 month old infants are sensitive to transitional probabilities in utterances of an artificial language. The transitional probabilities reflect the fact that certain segments tend not to co-occur adjacently unless separated by word-boundary, i.e., word internal segments are more predictable on the basis of the previous segments than a word initial syllable.

Johnson and Jusczyk (2001) show that although all of the cues are used, in case of a conflict, stress or coarticulation wins over statistics. Still statistics is sufficient if no other cues are present. Jusczyk (1995) suggests that infants may first learn to recognize isolated words and then match these sound patterns to ones in fluent speech. This cannot explain all word segmentation, but may provide a starting point for other strategies.

3.3. *External Cues for Morphology Acquisition*

Language contains many cues on different levels that a speaker can exploit when processing or acquiring morphology. None of these cues is 100% reliable. It is questionable whether they are available to their full extent during the developmental stage when morphology is acquired.

1. Phonotactics. It is often the case that a certain segment combination is impossible (or rare) within a morpheme but does occur across the morpheme boundary. Saffran et al. (1996) showed that hearers are sensitive to phonotactic transition probabilities across word boundaries (see §3.2). The results in (Hay, Pierrehumbert, & Beckman, 2003) suggest that this sensitivity extends to morpheme boundaries. Their study found that clusters infrequent in a given language tend to be perceived as being separated by a morpheme boundary.⁴
2. Syntactic cues. In some cases, it is possible to partially or completely predict inflectional characteristics of a word based on its syntactic context. For example in English, knowing what the subject is makes it possible to know whether or not the main verb will have the 3rd person singular form.
3. Semantic cues. Inflectionally related words (i) share certain semantic properties (e.g. both *walk* and *walked* refer to the same action), (ii) occur in similar contexts (*eat* and *ate* occur

⁴The study explores the perception of nonsense words containing nasal-obstruent clusters. Words containing clusters rare in English (e.g. /np/) were rated as potential words more likely when the context allowed placing a morpheme boundary in the middle of the cluster, e.g. *zan·plirshdom* was rated better than *zanp·lirshdom*.

with the same type of objects, while *eat* and *drink* occur with different type of objects). Similarly, words belonging to the same morphological category often share certain semantic features (e.g. referring to multiple entities).

4. Distributional cues. According to Baroni (2000), distributional cues are one of the most important cues in morphology acquisition. Morphemes are syntagmatically independent units – if a substring of a word is a morpheme then it should occur in other words. A learner should look for substrings which occur in a high number of different words (that can be exhaustively parsed into morphemes). He also claims that distributional cues play a primary role in the earliest stages of morpheme discovery. Distributional properties suggest that certain strings are morphemes, making it easier to notice the systematic semantic patterns occurring with certain of those words. Longer words are more likely to be morphologically complex.

3.4. Computational acquisition

In computational linguistics, there has been ongoing interest in the acquisition of morphology using both supervised and unsupervised methods. Although it is very likely that the processes in human acquisition are quite different from the computational algorithms, the type of information these algorithms exploit may be available to a human.

3.4.1. Clustering

Several algorithms exploit the fact that forms of the same lemma⁵ are likely to be similar in multiple ways. For example (Yarowsky & Wicentowski, 2000) assume that forms belonging to the same lexeme are likely to have similar orthography and contextual properties, and that the distribution of forms will be similar for all lexemes. In addition they combine these similarity measures with

⁵For example, the forms *break*, *breaks*, *broke*, *broken*, *breaking*, have the same lemma (citation form, canonical form) *break*.

iteratively trained probabilistic grammar generating the word forms. Similarly Baroni, Matiasek, and Trost (2002) use orthographical and semantic similarity.

Formal similarity. The usual tool for discovering similarity of strings is the Levenshtein edit distance (Levenshtein, 1966). The advantage is that it is extremely simple and is applicable to concatenative as well as nonconcatenative morphology. Some authors (Baroni et al., 2002) use the standard edit distance, where all editing operations (insert, delete, substitute) have a cost of 1. Yarowsky and Wicentowski (2000) use a more elaborated approach. Their edit operations have different costs for different segments and the costs are iteratively re-estimated; initial values can be based either on phonetic similarity or a related language.

Semantic similarity. In most of the applications, semantics cannot be accessed directly and therefore must be derived from other accessible properties of words. For example, Jacquemin (1997) exploits the fact that semantically similar words occur in similar contexts.

Distributional properties. Yarowsky and Wicentowski (2000) acquire morphology of English irregular verbs by comparing the distributions of their forms with regular verbs, assuming they are distributed equally.⁶ They also note that forms of the same lemma have similar selectional preferences. For example, related verbs tend to occur with similar subjects and objects. The selectional preferences are usually even more similar across different forms of the same lemma than

⁶Obviously, this approach would have to be significantly modified for classes other than verbs and/or for highly inflective languages. Let's consider for example Czech nouns. Not all nouns have the same distribution of forms. For example, many numeral constructions require the counted object to be in genitive. Therefore, currency names are more likely to occur in genitive than, say, proper names. Proper nouns occur in vocative far more often than inanimate objects, words denoting uncountable substances (e.g. *sugar*) occur much more often in singular than in plural, etc. Therefore, we would have to assume that there is not just a single distribution of forms shared by all the noun lemmas, but several distributions. The forms of currency names, proper names and uncountable substances would probably belong to different distributions.

The algorithm in Yarowsky and Wicentowski (2000) is given candidates for verbal paradigms and it discards those whose forms do not fit into the required uniform distribution. The algorithm for discovering Czech nouns could use the same technique, but (i) there would not be just one distribution but several, (ii) the algorithm would need to discover what those distributions are.

across synonyms. For this case, they manually specify regular expressions that (roughly) capture patterns of possible selectional frames.

3.4.2. Minimum description length

Minimum description length (MDL; (Rissanen, 1989); see also (Kazakov, 1997; Marcken, 1995)) is based on the insight that a grammar can be used to compress a corpus; the better is the grammar the better is the compression. MDL considers both the size of the grammar and the size of the compressed corpus. Goldsmith (2001) used the MDL approach in an algorithm acquiring concatenative morphology in a completely unsupervised manner from a raw corpus (5K to 500K words). MDL is used to accept or reject hypotheses proposed by a set of heuristics. He achieves 86% precision and 90% recall for English and similar results for French. The algorithm tends to correctly identify large regular paradigms.

3.4.3. Prior knowledge

Some algorithms can capitalize on prior knowledge if it is available – for example Yarowsky and Wicentowski (2000) can add lists of functional words, phonetic similarities table for a related language, lists of nouns/verb/adjective roots and some rules for identifying the part-of-speeches of the rest.

3.4.4. Neural networks

Most of the research on using neural or connectionist networks for morphological acquisition is devoted to finding models that are able to learn both rules and exceptions (Cf. Rumelhart & McClelland, 1986; Plunkett & Marchman, 1991; Prasada & Pinker, 1993, etc.). Since we are interested in comparing morphological systems in terms of their typological properties, this research is not directly relevant.

However, there is also research comparing the acquisition of different morphological types. Gasser (1994) shows that a simple modular recurrent connectionist model is able to acquire various inflectional processes and that different processes have a different level of acquisition complexity. His model takes phones (one at a time) as input and outputs the corresponding stems and inflections. During the training process, the model is exposed to both forms and the corresponding stem-inflection pairs. This is similar (with enough simplification) to our idealization of a child being exposed to both forms and their meanings.

Many of the results are in concord with the preferences attested in real languages (see §1.2) – it was easier to identify roots in a suffixing language than in a prefixing one, the templates were relatively easy and infixes were relatively hard.⁷ In a similar experiment Gasser and Lee (1991) showed that the model does not learn linguistically implausible languages – Pig Latin or language mirror image language (see (5)). The model was unable to learn any form of syllable reduplication. A model enhanced with modules for syllable processing was able to learn a very simple form of reduplication – reduplicating onset or rime of a single syllable. It is necessary to stress that the problem addressed by Gasser was much simpler than real acquisition: (1) at most two inflectional categories were used, each with only two values, (2) each form belonged only to one paradigm, (3) there were no irregularities, (4) only the relevant forms with their functions were presented (no context, no noise).

4. The complexity model

We turn next to our approach to the issue. For the comparison of acquisition complexity of different morphological systems, we assume that morphology acquisition has three consecutive stages as

⁷The accuracy of root identification was best in the case of suffixes, templates and umlaut (ca 75%); in the case of prefixes, infixes and deletion it was lower (ca 50%); all above the chance baseline (ca 3%) The accuracy of the inflection identification showed a different pattern – the best were prefix and circumfix (95+%), slightly harder were deletion, template and suffix (90+%), and the hardest were umlaut and infix (ca 75%); all above the chance baseline (50%).

follows:⁸

1. forms are learned as suppletives,
2. paradigms (i.e., groups of forms sharing the same lemma) are discovered and forms are grouped into paradigms,
3. regularities in paradigms are discovered and morphemes are identified (if there are any).

The first stage is uninteresting for our purpose; the complexity of morphological acquisition is determined by the complexity of the second and third stages. To simplify the task, we focus on the second stage. This means that we estimate the complexity of morphology acquisition in terms of the complexity of clustering words into paradigms: the easier it is to cluster words into paradigms, the easier, we assume, it will be to acquire their morphology.⁹

We assume that this clustering is performed on the basis of the semantic and formal similarity of words; words that are formally and semantically similar are put into the same paradigm and words that are different are put into distinct paradigms. We employ several simplifications: we ignore most irregularities, we assume that there is no homonymy and no synonymy of morphemes and we also disregard phonological alternations. Obviously, a higher incidence of any of these makes the acquisition task harder.

4.1. Semantic similarity

Our model simplifies the acquisition task further by assuming that the semantics is available for every word. We believe that this is not an unreasonable assumption since infants are exposed

⁸A more realistic model would allow iterative repetition of these stages. Even after establishing a basic morphological competence, new forms that are opaque for it are still learned as suppletives. The output of Stage 3 can be used to improve the clustering in Stage 2.

⁹Of course, it is possible to imagine languages where Stage 2 is easy and Stage 3 is very hard. For instance, in a language where plural is formed by some complex change of the last vowel, Stage 2 is quite simple (words that differ only in that vowel go into the same paradigm), while Stage 3 (discovering the rule that governs the vowel change) is hard.

to language in context. If they have limited access to context, their language development is very different, as Peters and Menn (1993) show in their comparison of morphological acquisition in a normal and a visually impaired child. Moreover, as computational studies show, words can be clustered into semantic classes using their distributional properties (Yarowsky & Wicentowski, 2000).

4.2. *Similarity of forms*

As noted earlier, we assume that ease of morphological acquisition correlates with ease of clustering forms into paradigms using their formal similarity as a cue. We propose a measure called paradigm similarity index (PSI) to quantify the ease of such clustering. A low PSI means that (in general) words belonging to the same paradigm are similar to each other, while they are different from other words. The lower the index, the easier it is to correctly cluster the forms into paradigms.

If L denotes the set of words (types, not tokens) in a language L and $\text{prdgm}(w)$ is a set of words belonging to the same paradigm as the word w , then we can define PSI as:

$$\text{PSI}(L) = \text{avg}\{\text{ipd}(w)/\text{epd}(w) \mid w \in L\} \quad (1)$$

where epd is the average distance between a word and all other words:

$$\text{epd}(w) = \text{avg}\{\text{ed}(w, u) \mid u \in L\} \quad (2)$$

and ipd is the average distance between a word and all words of the same paradigm:

$$\text{ipd}(w) = \text{avg}\{\text{ed}(w, u) \mid u \in \text{prdgm}(w)\} \quad (3)$$

Finally, `ed` is a function measuring the similarity of two words (similarity of their forms, i.e., sounds, not of their content). In the subsequent models, we use various variants of the Levenshtein distance (LD), proposed by Levenshtein (1966), as the `ed` function.

4.3. Model 0 – Standard Levenshtein distance

The Levenshtein distance defines the distance between two sequences s_1 and s_2 as the minimal number of edit operations (substitution, insertion or deletion) necessary to modify s_1 into s_2 . For an extensive discussion of the original measure and a number of modifications and applications, see (Sankoff & Kruskal, 1999).

The algorithm of the Model 0 variant of the `ed` function is in Fig. 1. The pseudocode is very similar to functional programming languages like Haskell or ML. The function `ed` accepts two strings and returns a natural number – the edit distance of those strings. The function is followed by several templates introduced by ‘|’ selecting the proper code depending on the content of the arguments. The edit distance of

- two empty strings is 0,
- a string from an empty string is equal to the length of that string – the number of DELETES or INSERTS necessary to turn one into the other.
- two nonempty strings is equal to the cost of the cheapest of the following three possibilities
 - cost of MATCH or SUBSTITUTE on the current characters plus the edit distance between the remaining characters.
 - the cost of DELETing the first character of the first string (`u`), i.e., 1, plus the edit distance between the remaining characters (`us`) and the second string (`v:vs`)
 - the cost of INSERTing the first character of the second string (`v`) at the beginning of the

first string, i.e., 1, plus the edit distance between the first string (u:us) and the remaining characters of the second string (vs)

The standard Levenshtein distance is a simple and elegant measure that is very useful in many areas of sequence processing. However, for morphology and especially acquisition, it is an extremely rough approximation. It does not reflect many constraints of the physical and cognitive context the acquisition occurs in. For example, the fact that some mutations are more common than others (e.g. vowels are more volatile than consonants) is not taken into account.

What is most crucial, however is that the standard LD does not reflect the fact that words are perceived and produced in time. The distance is defined as the minimum cost over all possible string modifications. This may be desirable for many applications and is even computable by a very effective dynamic programming algorithm (Cf. Sankoff & Kruskal, 1999). However the limitations of human memory make such a computational model highly unrealistic. In the subsequent models, we modify the standard Levenshtein distance measure in such a way that it reflects more intuitively the physical and cognitive reality of morphology acquisition. Some of the modifications are similar to edit distance variants proposed by others, some we believe are original.

4.3.1. *Suffix vs. prefix*

Unsurprisingly, our Model 0 (based on the standard Levenshtein distance) treats suffixing and prefixing languages as equally complex. Consider the two “languages” in Table 1, or more formally in (6), differing only in the position of the affix.

$$(6) L = \{kuti, norebu, \dots\}, A = \{ve, ba\}, L_P = A \cdot L, L_S = L \cdot A.$$

For both languages, the cheapest way to modify any singular form to the corresponding plural form is to apply two substitution operations on the two segments of the affix. Therefore, the edit cost is

2 in both cases, as Table 2 shows. The same is true in the opposite direction (Plural \rightarrow Singular). Therefore the complexity index is the same for both languages. Similarly, the result for languages with different length of affixes (*ve·kuti* vs. *uba·kuti*) or languages where one of the forms is a bare stem (*kuti* vs. *ba·kuti*) would be the same for both affix types – see Table 3. Of course, this is not the result we are seeking.

Mirror image Obviously, the model (but also the standard Levenshtein distance) predicts that reversal as a hypothetical morphological operation is extremely complicated to acquire – it is unable to find any formal similarity between two forms related by reversal.

4.4. Model 1 – matching strings in time

In this and subsequent models, we modify the standard edit distance to better reflect the linguistic and psychological reality of morphological acquisition – especially the fact that language occurs in time, and that human computational resources are limited.

Model 1 uses an incremental algorithm to compute similarity distance of two strings. Unlike Model 0, Model 1 calculates only one edit operation sequence. At each position, it selects a single edit operation. The most preferred operation is MATCH. If MATCH is not possible, another operation (SUBSTITUTE, DELETE or INSERT) is selected randomly.¹⁰ The edit distance computed by this algorithm is larger or equal to the edit distance computed by Model 0 algorithm (Fig. 1). It cannot be smaller, because Model 0 computes the optimal distance. It can be larger because the operation selected randomly does not have to be optimal.

The algorithm for computing such edit distance is spelled out in Fig. 2. The code for the first three cases (two empty strings, or a nonempty string and an empty string) is the same as in the Model 1 algorithm. The algorithms differ in the last two cases covering nonempty strings: MATCH

¹⁰A more realistic model could (1) adjust the preference in the operation selection by experience; (2) employ a limited look-ahead window. For the sake of simplicity, we ignore these options.

is performed if possible, a random operation is selected otherwise.

4.4.1. Prefixes vs. Suffixes.

Other things being equal, Model 1 considers it easier to acquire paradigms of a language with suffixes than of a language with prefixes. Intuitively, the reason for the higher complexity of prefixation is as follows: When a non-optimal operation is selected, it negatively influences the matching of the rest of the string. In a prefixing language, the forms of the same lemma differ at the beginning and therefore a non-optimal operation can be selected earlier than in a suffixing language. Thus the substring whose matching is negatively influenced is longer.

Let L_P be a prefixing language, L_S the analogous suffixing language, $w_p \in L_P$ and w_s the analogous word $\in L_S$.¹¹ Obviously, it is more probable that $\text{ipd}(w_p) \geq \text{ipd}(w_s)$ than not. Asymptotically, for infinite languages, the $\text{epd}(w_p) = \text{epd}(w_s)$. Therefore, for such languages $\text{PSI}(L_P) > \text{PSI}(L_S)$. We cannot assume infinite languages, but we assume that the languages are large enough to avoid pathological anomalies.

Consider Fig. 3. It shows all the possible sequences of edit operations for two forms of a lemma from both prefixing (A) and suffixing (B) languages L_P and L_S . The best sequences are on the diagonals.¹² The best sequences (SSMMMM, or 2 SUBSTITUTES followed by 4 MATCHES, for L_P and MMMMSS for L_S) are of course the same as those calculated by the standard Levenshtein Distance. And their costs are the same for both languages. However, the paradigm similarity index PSI is not defined in terms of the best match, but in terms of the average cost of all possible sequences of edit operations – see (1). The average costs are different; they are much higher for L_P than for L_S . For L_S , the cost is dependent only on the cost of matching the two suffixes. The stems are always

¹¹If S is a set of stems, A a set of affixes, then $L_P = A \cdot S$ and $L_S = S \cdot A$. If $s \in S$ and $a \in A$, then $w_p = a \cdot s$ and $w_s = s \cdot a$. The symbol \cdot denotes both language concatenation and string concatenation.

¹²Note that this is not the general case, e.g. for words of different length there is no diagonal at all – cf. Fig.3 C or D.

matched by the optimal sequence of MATCH operations. Therefore a deviation from the optimal sequence can occur only in the suffix. In L_P , however, the uncertainty occurs at the beginning of the word and a deviation from the optimal sequence there introduces uncertainty later that cause further deviations from the optimal sequence of operations. The worst sequences for L_S contain 4 MATCHES, 2 DELETES and 2 INSERTS; the cost is 4. The worst sequences for L_P contain 6 DELETES and 6 INSERTS; the cost is 12.

In case of languages using zero affixes, the difference is even more apparent, as C & D in Fig. 3 show. Model 1 allows only one sequence of edit operations for words *kuti* and *kuti-ve* of the suffixing language $L_S^0 - \text{MMMMII}$.¹³ The cost is equal to 2 and since there are no other possibilities, the average cost of matching those two words is trivially optimal. The optimal sequence for words *kuti* and *ve-kuti* of the prefixing language L_P^0 (IIMMMM) costs also 2. However, there are many other nonoptimal sequences. The worst ones contain 6 INSERTS and 4 DELETES and have a cost of 10.¹⁴

4.4.2. Evaluation

We randomly generate pairs of languages in various ways. The members of the pair are identical except for the position of the affix. There is no homonymy in the languages. For each such pair we calculated the following ratio:

$$\text{sufPref} = \frac{\text{PSI}(L_P)}{\text{PSI}(L_S)} \quad (4)$$

If $\text{sufPref} > 1$ Model 1 considers the suffixing language L_S easier to acquire than the prefixing language L_P .

¹³Note that DELETE or INSERT operations cannot be applied if MATCH is possible.

¹⁴In a model using a look-ahead window, the prefixing language would be still more complex, but the difference would be smaller.

We generated 100 such pairs of languages with the parameters summarized in Table 4, calculating statistics for suffPref. The alphabet can be thought of as a set of segments, syllables or other units. Before discarding homonyms, all distributions are uniform. As can be seen from Table 5, Model 1 really considers the generated suffixing languages much simpler than the prefixing ones.

4.4.3. Other processes

Infixes. Model 1 makes an interesting prediction about the complexity of infixes. It considers infixing languages to be more complex than suffixing languages, but less complex than prefixing languages. The reason is simple – the uncertainty is introduced later than in case of a prefix, therefore the possibly the string whose matching can be influenced by a non-optimal operation selection is shorter.

This prediction contradicts the fact that infixes are much rarer than prefixes (§1.2). Note, however that the prediction concerns simplicity of clustering word forms into paradigms. According to the model, it is easier to cluster forms of an infixing language into paradigms than those of a prefixing language. It may well be the case that infixing languages are more complex from another point of view, that of identification of morphemes: other things being equal, a discontinuous stem is probably harder to identify than a continuous one.

Metathesis. The model prefers metathesis occurring later in a string for the same reasons as it prefers suffixes over prefixes. This prediction is in accord with data (see §A.2). However, the model also considers metathesis (of two adjacent segments) to have the same cost as an affix consisting of two segments and even cheaper than an affix with more segments. This definitely does not reflect the reality. In 4.5.2, we suggest how to rectify this.

Mirror image Similarly as Model 0, this model considers mirror image to be extremely complicated to acquire.

Templatic morphology. As we note in Appendix §A.1, templatic morphology does not have to be harder to acquire than morphology using continuous affixes. Following Fowler (1983), it can be claimed that consonants of the root and vowels of the inflection are perceptually in different “dimensions” – consonants are modulated on the basic vowel contour of syllables – and therefore clearly separable.

4.5. Possible further extensions

4.5.1. Model 2 – morpheme boundaries and backtracking

In this section we suggest extending Model 1 by a notion of a probabilistic morpheme boundary to capture the fact that, other things being equal, exceptions and high number of paradigm patterns make a language harder to acquire. This is just a proposal; we leave a proper evaluation for a future paper.

Intuitively, a morphological system with a small number of paradigmatic patterns should be easier to acquire than a system with large number of paradigms (or a lot of irregularities). However the measure in previous models is strictly local. The cost depends only on the matched pair of words, not on global distributional properties. This means that words related by a rare pattern can have the same score as words related by a frequent pattern. For example, Model 1 considers, *foot* [fut] / *feet* [fit] to be equally similar as *dog* [dag] / *dogs* [dagz], or even more similar than *bench* [bentʃ] / *benches* [bentʃɪs]. Thus a language with one paradigmatic pattern is assigned the same complexity as a language where every lemma has its own paradigm (assuming the languages are otherwise equal, i.e., they are of the same morphological type and morphemes have the same length).

Model 2 partially addresses this drawback by enhancing Model 1 with probabilistic morpheme

boundaries and backtracking. Probabilistic morpheme boundaries are dependent on global distributional properties, namely syllable predictability. Which syllable will follow is less predictable across morphemes than morpheme internally. This was first observed by Harris (1955), and is usually exploited in computational linguistics in unsupervised acquisition of concatenative morphology (see §3.4). Johnson and Jusczyk (2001) show that the degree of syllable predictability is one of the cues used in word segmentation (see §3.2). Since acquisition of word segmentation occurs before morphology acquisition, it is reasonable to assume that this strategy is available in the case of morphological acquisition as well. Hay et al. (2003) suggest that this is in fact the case. They found that clusters that are infrequent in a given language tend to be perceived as being separated by a morpheme boundary. The transitional probabilities for various syllables¹⁵ are more distinct in a language with few regular paradigms. Thus in such a language morpheme boundaries are easier to determine than in a highly irregular language.

In Model 2, the similarity distance between two words is computed using a stack and backtracking. Each time when there is a choice of operation (i.e., anytime MATCH operation cannot be applied), a choice point is remembered on the stack. This means that Model 2 makes it possible to correct apparent mistakes in matching that Model 1 was not able to do. The new total similarity distance between two words is a function of (1) the usual cost of edit operations, (2) the size of the stack in all steps (3) the cost of possible backtracking. Each of them is adding to the memory load and/or slowing processing.

Matching morpheme boundaries increases the probability that the two words are being matched the “right” way (i.e., that the match is not accidental). This means that it is more likely that the choices of edit operations made in the past were correct, and therefore backtracking is less likely to occur. In such case, Model 2 flushes the stack. Similarly, the stack can be flushed if a certain number of matches occurs in a row, but a morpheme boundary contributes more to the certainty

¹⁵It is probable that learners extract similar probabilities on other levels as well – segments, feet, etc.

of the right analysis. In general, we introduce a notion of *anchor*, that is, a sequence of matches of certain weight when the stack is flushed. This can be further enhanced by assigning different weights to matching of different segments (consonants are less volatile than vowels). Morpheme boundaries would then have higher weight than any segment. Moreover, more probable boundaries would have higher weights than less probable ones.

Thus in general, a regular language with more predictable morpheme boundaries needs a smaller stack for clustering words according to their formal similarity.

Suffix vs. prefix. It is evident that Model 2 also considers prefixing languages more complex than suffixing languages for two reasons. First, the early uncertainty of a prefixing language leads to more deviations from the minimal sequence of edit operations in the same way as in Model 1. Second, the stack is filled early and the information must be kept there for a longer time, therefore the memory load is higher.

Infixes. Our intuitions tell us that Model 2, unlike Model 1, would consider an infixing language more complex than a prefixing language. The reason is that predicting morpheme boundaries using statistics is harder in an infixing language than in the corresponding prefixing language. However we have not worked out the formal details of this.

4.5.2. *Other possibilities*

Variable atomic distances. A still more realistic model would need to take into consideration the fact that certain sounds are more likely to be substituted for one another than other sounds. The model would reflect this by using different SUBSTITUTE costs for different sound pairs. For example, substituting [p] for [b], which are the same sounds except voicing, would be cheaper than substituting [p] for [i], which differ in practically all features. This would reflect (i) language

independent sound similarities related to perception or production (e.g. substituting a vowel by a vowel would be cheaper than replacing it by a consonant), (ii) sound similarities specific to a particular language and gradually acquired by the learner (e.g. [s] and [ʃ] are allophones, and are therefore often substituted one for the other, in Korean, but not in Czech). An iterative acquisition of these similarities was successfully used by (Yarowsky & Wicentowski, 2000, Cf. 3.4.1)

More realistic insert. The model could also employ more realistic INSERT operations, one referring to a lexicon of acquired items and one referring to the word to be matched. The former INSERT would allow the insertion of units recognized as morphemes in the previous iterations of the second (paradigm discovery) and third stages (pattern discovery) of the acquisition process. This INSERT is much cheaper than the normal INSERT. A model containing such INSERT would consider metathesis much more complex than, for example, concatenative morphology. The latter INSERT would work like a copy operation – it would allow inserting material occurring at another place in the word. This INSERT would make reduplication very simple.

5. Conclusion

In this paper, we showed that it is possible to model the prevalence of various morphological systems in terms of their acquisition complexity. Our complexity measure is based on the Levenshtein edit distance modified to reflect external constraints – human memory limitations and the fact that language occurs in time. Such a measure produces some interesting predictions; for example it predicts correctly the prefix-suffix asymmetry and shows mirror image morphology to be virtually impossible.

A. Templatic morphology, Metathesis

A.1. Templatic morphology

In template morphology, both the roots and affixes are discontinuous. Only Semitic languages belong to this category. Semitic roots are discontinuous consonantal sequences formed by 3 or 4 consonants (l-m-d – ‘learn’). To form a word the root must be interleaved with a (mostly) vocalic pattern.

	lomed ‘learn _{masc} ’	shatak ‘be-quiet _{pres.masc} ’
(7)	lamad ‘learnt _{masc.sg.3rd} ’	shatak ‘was-quiet _{masc.sg.3rd} ’
	limed ‘taught _{masc.sg.3rd} ’	shitek ‘made-sb-to-be-quiet _{masc.sg.3rd} ’
	lumad ‘was-taught _{masc.sg.3rd} ’	shutak ‘was-made-to-be-quiet _{masc.sg.3rd} ’

Phonological alternations are possible – e.g. stop alternating with fricatives ([b]/[v]). Semitic morphology is not exclusively templatic, some processes are also concatenative.

Processing Template morphology. From the processing point of view, template morphology may seem complicated. However, if we assume that consonants of the root and vowels of the inflection are perceptually in different “dimensions” and therefore clearly separable, it would not be more complicated than morphology using continuous affixes or suprasegmentals. Fowler (1983) convincingly argues on phonetic grounds for such assumption – consonants are modulated on the basic vowel contour of syllables.

Ravid’s study (2003) also suggests that template morphology is not more difficult to acquire than a concatenative one. He finds that in case of forms alternatively produced by template and concatenative processes, children tend to acquire the template option first. He also claims that young Israeli children rely on triconsonantal roots as the least marked option when forming certain verbs. Three year old children are able to extract the root from a word – they are able to interpret novel

root-based nouns.

A.2. *Metathesis*

In morphological metathesis, the relative order of two segments encodes a morphological distinction. For example, in Rotuman (Austronesian family, related to Fijian), words distinguish two forms, called the complete and incomplete phase¹⁶ by Churchward (1940), and in many cases these are distinguished by metathesis (examples due to Hoeksema and Janda (1988)):¹⁷

	aíε	aiér	‘fish’	
(8)	púε	puér	‘rule, decide’	(Rotuman)
	tíkɔ	tjók	‘flesh’	
	séma	séám	‘left-handed’	

Although phonological metathesis is not rare, it is far less common than other processes like assimilation. As a morphological marker (i.e., not induced by phonotactics as a consequence of other changes) it is extremely rare – found in some Oceanic (incl. the above mentioned Rotuman) and North American Pacific Northwest languages (e.g. Sierra Miwok, Mutsun) (Becker, 2000). According to Janda (1984), it is probable that in such cases of metathesis, originally, some other means marked the morphological category and metathesis was only a consequence of phonotactic constraints, and only later it became the primary marker.

Mielke and Hume (2001) examined 54 languages involving metathesis and found that it is very rare word/root-initially or with non-adjacent segments. They found only one language (Fur) with a fully productive root-initial metathesis involving wide variety of sounds. Apparent cases of non adjacent

¹⁶According to Hoeksema and Janda (1988), the complete phase indicates definiteness or emphasis for nouns and perfective aspect or emphasis for verbs and adjectives; while the incomplete phase marks words as indefinite/imperfective and nonemphatic.

¹⁷In many cases, subtraction (*rako* vs. *rak* ‘to imitate’), subtraction with umlaut (*hoti* vs. *höt* ‘to embark’) or identity (*rī* vs. *rī* ‘house’) is used instead. See (McCarthy, 2000) for more discussion.

metathesis can be usually analyzed as two separate metathesis, each motivated by an independent phonological constraint.

Processing Metathesis. Mielke and Hume (2001) suggest that the reasons for the relative infrequency of metathesis are related to word recognition – metathesis impedes word recognition more than other frequent processes, like assimilation. Word recognition (see §3.1) can also explain the fact that it is even rarer (or perhaps nonexistent) word/root-initially or with non-adjacent segments: since (i) lexical access is generally achieved on the basis of the initial part of the word and (ii) since phonological changes involving non-adjacent segments are generally more disruptive to word recognition.

References

- Baroni, M. (2000). *Distributional cues in morpheme discovery: A computational model and empirical evidence*. Unpublished doctoral dissertation, UCLA.
- Baroni, M., Matiasek, J., & Trost, H. (2002). Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. In *Proceedings of the Workshop on Morphological and Phonological Learning of ACL/SIGPHON-2002* (p. 48-57).
- Becker, T. (2000). Metathesis. In G. Booij, C. Lehmann, & J. Mugdan (Eds.), *Morphology: A handbook on inflection and word formation* (pp. 576–581). Berlin: Mouton de Gruyter.
- Chomsky, N. (1973). Conditions on Transformations. In S. Anderson & P. Kiparsky (Eds.), *Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.
- Churchward, C. M. (1940). *Rotuman Grammar and Dictionary*. Sydney: Australasia Medical Publishing Co. (Reprint New York: AMS Press, 1978)
- Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception and Psychophysics*(13), 153–156.

- Cole, R. A., & Jakimik. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America*(67), 965–970.
- Cole, R. A., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Information processing strategies*. London: Academic Press.
- Connine, C. M., Blasko, D. G., & Titone., D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*(32), 193-210.
- Cooper, W., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Culicover, P. W., & Nowak, A. (2002). Markedness, antisymmetry and complexity of constructions. In P. Pica & J. Rooryk (Eds.), *Variation yearbook. vol. 2 (2002)* (pp. 5–30). Amsterdam: John Benjamins.
- Cutler, A., Hawkins, J., & Gilligan, G. (1985). The suffixing preference: a processing explanation. *Linguistics*, 23, 723-758.
- Fowler, C. A. (1983). Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic production of Vowels in Monosyllabic Stress Feet. *Journal of Experimental Psychology*, 386-412.
- Gasser, M. (1994). Acquiring receptive morphology: a connectionist model. In *Annual Meeting of the Association for Computational Linguistics*, 32 (p. 279-286).
- Gasser, M., & Lee, C.-D. (1991). A short-term memory architecture for the learning of morpho-phonemic rules. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems 3* (p. 605-611). San Mateo, CA: Morgan Kaufmann.
- Givón, T. (1979). *On Understanding Grammar*. New York: Academic Press.
- Goldsmith, J. (2001). Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics*, 27(2), 153–198.
- Greenberg, J. H. (1957). *Essays in Linguistics*. Chicago: University of Chicago Press.

- Hall, C. J. (1988). Integrating Diachronic and Processing Principles in Explaining the Suffixing Preference. In J. A. Hawkins (Ed.), *Explaining Language Universals* (chap. 12).
- Harris, Z. S. (1955). From phoneme to morpheme. *Language*, 31(2), 190-222.
- Hawkins, J. A., & Gilligan, G. (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua*, 74, 219-259.
- Hay, J., Pierrehumbert, J., & Beckman, M. (2003). Speech Perception, Well-formedness and the Statistics of the Lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Laboratory Phonology VI*. CUP.
- Hoeksema, J., & Janda, R. D. (1988). Implications of Process-morphology for Categorical Grammar. In R. T. Oehrle, E. Bach, & D. Wheeler (Eds.), *Categorical Grammars and Natural Language Structures* (pp. 199-247). Academic Press.
- Jacquemin, C. (1997). Guessing Morphology from Terms and Corpora. In *SIGIR '97: Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 27-31, 1997, Philadelphia, PA, USA* (p. 156-165). ACM.
- Janda, R. D. (1984). Why Morphological Metathesis Rules Are Rare: On the Possibility of Historical Explanation in Linguistics. In C. Brugman & M. Macaulay (Eds.), *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society* (pp. 87-103). Berkeley, CA: Berkeley Linguistics Society.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, 44(4), 548-567.
- Jusczyk, P. W. (1995). Language Acquisition: Speech Sounds and the Beginning of Phonology. In J. Miller & P. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, communication*. Academic Press.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29.

- Kazakov, D. (1997). *Unsupervised learning of naive morphology with genetic algorithms*.
- Levenshtein. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*, 10(8), 707–710.
- Marcken, C. de. (1995). *Unsupervised Language Acquisition*. Unpublished doctoral dissertation, MIT, Cambridge, MA.
- Marslen-Wilson, W. D. (1993). Issues of process and representation in lexical access. In G. Altmann & R. Shillcock (Eds.), *Cognitive Models of Language Processes: The Second Sperlonga Meeting*. Hove: Lawrence Erlbaum Associates.
- Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*(8), 1–71.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*(10), 29–63.
- McCarthy, J. (2000). The prosody of phase in Rotuman. *Natural language and linguistic theory*, 18(1), 147–197.
- Mielke, J., & Hume, E. (2001). Consequences of Word Recognition for Metathesis. In E. Hume, N. Smith, & J. van de Weijer (Eds.), *Surface Syllable Structure and Segment Sequencing*. Leiden: HIL.
- Nooteboom, S. (1981). Lexical retrieval from fragments of spoken words: beginnings versus endings. *Journal of Phonetics*, 9, 407–424.
- Peters, A. M., & Menn, L. (1993). False starts and filler syllables: ways to learn grammatical morphemes. *Language*, 69(4), 742–777.
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition*, 38, 73–193.
- Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8, 1–56.

- Ravid, D. (2003). A developmental perspective on root perception in Hebrew and Palestinian Arabic. In Y. Shimron (Ed.), *The processing and acquisition of root-based morphology*. Amsterdam: Benjamins.
- Rissanen, J. (1989). *Stochastic complexity in statistical inquiry*. Singapore: World Scientific Publishing Co.
- Rumelhart, D., & McClelland, J. (1986). On learning the past tense of english verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing* (Vol. II). Cambridge: MIT Press.
- Saffran, J., Aslin, R., & Newport, E. (1996). *Science*, 274(5294), 1926-1928.
- Sankoff, D., & Kruskal, J. B. (Eds.). (1999). *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*. CSLI.
- Sapir, E. (1921). *Language, an introduction to the study of speech*. New York: Harcourt.
- Spencer, A., & Zwicky, A. M. (Eds.). (1998). *The Handbook of Morphology*. Oxford, UK and Malden, MA: Blackwell.
- Wexler, K., & Culicover, P. W. (1980). *Formal principles of language acquisition*. Cambridge, Mass: MIT Press.
- Yarowsky, D., & Wicentowski, R. (2000). Minimally supervised morphological analysis by multi-modal alignment. In *Proceedings of the 38th Meeting of the Association for Computational Linguistics* (p. 207-216).

Figure 1: Model 0 (Levenshtein) Edit Distance Algorithm

Figure 2: Model 1 Edit Distance Algorithm

Figure 3: Comparing words in Model 1

Table 1: Sample prefixing and suffixing languages

Table 2: Comparing prefixed and suffixed words in Model 0

Table 3: Comparing prefixed and suffixed words in Model 0

Table 4: Experiment: Parameters

Table 5: Experiment: Results

Abstract

There are many logical possibilities for marking morphological features. However only some of them are attested in languages of the world, and out of them some are more frequent than others. For example, it has been observed (Sapir, 1921; Greenberg, 1957; Hawkins & Gilligan, 1988) that inflectional morphology tends to overwhelmingly involve suffixation rather than prefixation. This paper proposes an explanation for this asymmetry in terms of acquisition complexity. The complexity measure is based on the Levenshtein edit distance, modified to reflect human memory limitations and the fact that language occurs in time. This measure produces some interesting predictions: for example, it predicts correctly the prefix-suffix asymmetry and shows mirror image morphology to be virtually impossible.

```

ed :: String, String -> Integer
| [], []      = 0
| u, []      = length u          // DELETE u
| [], v      = length v          // INSERT v
| u:us, v:vs = min [
    (if u == v then 0 else 1) + ed(us, vs), // MATCH/SUBST
    1 + ed(us, v:vs),                       // DELETE u
    1 + ed(u:us, vs) ]                       // INSERT v

```

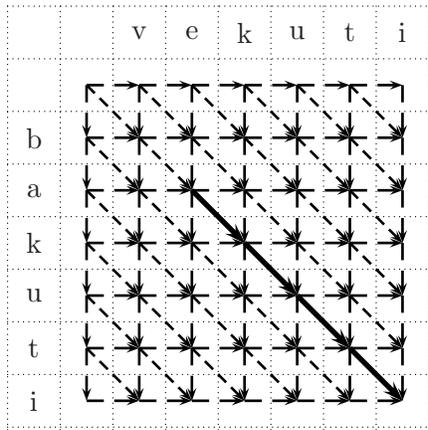
Figure 1: Model 0 (Levenshtein) Edit Distance Algorithm

```

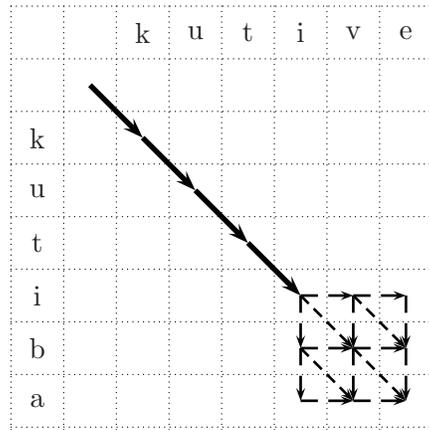
ed :: String, String -> Integer
| [], []      = 0
| u, []      = length u          // DELETE u
| [], v      = length v          // INSERT v
| u:us, u:vs = ed(us, vs)        // MATCH
| u:us, v:vs = 1 + random [ed(us, vs), ed(us, v:vs), ed(u:us, vs)]
                                // SUBSTITUTE, DELETE, INSERT

```

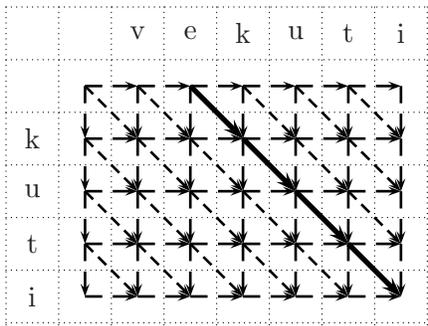
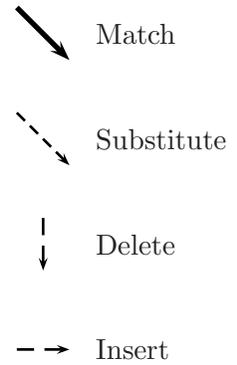
Figure 2: Model 1 Edit Distance Algorithm



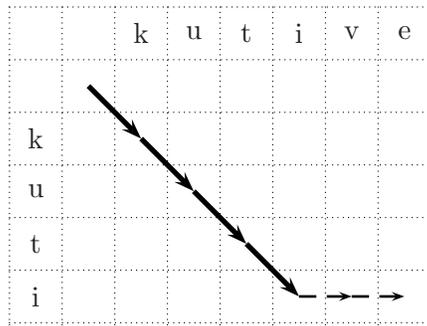
A. A prefixing language in M1



B. A suffixing language in M1



C. Zero prefixes in M1



D. Zero suffixes in M1

Figure 3: Comparing words in Model 1

Table 1: Sample prefixing and suffixing languages

	Prefixing language (L_P)	Suffixing language (L_S)
Singular	ve·kuti	kuti·ve
Plural	ba·kuti	kuti·ba
Singular	ve·norebu	norebu·ve
Plural	ba·norebu	norebu·ba
	⋮	⋮

Table 2: Comparing prefixed and suffixed words in Model 0

Prefixing language (L_P)				Suffixing language (L_S)			
		operation	cost			operation	cost
v	b	substitute	1	k	k	match	0
e	a	substitute	1	u	u	match	0
k	k	match	0	t	t	match	0
u	u	match	0	i	i	match	0
t	t	match	0	v	b	substitute	1
i	i	match	0	e	a	substitute	1
Total cost			2	Total cost			2

Table 3: Comparing prefixed and suffixed words in Model 0

Prefixing language (L'_P)				Suffixing language (L'_S)			
		operation	cost			operation	cost
	u	insert	1	k	k	match	0
v	b	substitute	1	u	u	match	0
e	a	substitute	1	t	t	match	0
k	k	match	0	i	i	match	0
u	u	match	0	v	u	substitute	1
t	t	match	0	e	b	substitute	1
i	i	match	0		a	insert	1
Total cost			3	Total cost			3

Table 4: Experiment: Parameters

Number of languages	100
Alphabet size	25
Number of stems in a language	50
Shortest stem	1
Longest stem	6
Number of affixes in a language	3
Shortest affix	0
Longest affix	3

Table 5: Experiment: Results

mean	1.29
standard deviation	0.17
Q1	1.16
median	1.27
Q3	1.33