Cahiers du CLSL, n° 69, 2025, pp. 57-71

Exploring oppositions in morphology

John A. Goldsmith & Marina Ermolaeva
University of Chicago; Moscow State University
goldsmith@uchicago.edu; marinkaermolaeva@gmail.com

Abstract

In this paper, we define a group structure over strings and note that by applying this computation to words, we obtain major steps towards a method for identifying allomorphy and learning morphophonemics. First order differences among a set of words forming a paradigm identify *morphs*, while second order differences identify *allomorphy*. When this allomorphy appears at morpheme boundary, this can in a wide range of cases be identified as *morphophonology*.

1 Introduction

For those who delve deeply into it, language, like music, seems to reveal complex patterns of varying scales. What is the language of patterns? At its most abstract, the language is transformations, and at its more concrete, it is differences and samenesses. The student who learns a new European language has to learn various verbal paradigms, with the understanding that once we learn to conjugate *parler*, for example, that knowledge will directly extend to a large number of other verbs, because the differences among the various inflected forms of *parler* are exactly matched by the differences among the inflected forms of *sauter*, and many, many other verbs. Our goal in the work we describe here is based on the belief that by carefully building a series of notions of difference that are useful in describing language, we may achieve greater insight into the structure of language. If we let ourselves be inspired by the calculus, we are led to ask whether our notions of difference can be extended to second-order differences, just as first

derivatives are extended to second derivatives. This paper describes certain initial steps along the way to that end, and is part of a larger project in progress.[1]

2 Strings, and a group structure for multisets

Discussions of formal languages often begin by assuming an alphabet and a semigroup formed from the alphabet with concatenation, and just as often an identity element is assumed, forming a monoid. To talk about particular languages, it is necessary in addition to include the notion of *set* (or, as we will see, something *like* the notion of set, such as a *multiset*). A regular language can be defined as a semiring with two operations, *set union* and *string concatenation*, defined both with respect to pairs of strings and to sets of strings (with closure under both operations as well as closure under Kleene star operation, and complementation, though we do not discuss these latter two properties).

In this paper we are interested, however, in exploring the notion of *difference* of strings, and of sets of strings, and this requires that an *inverse* be available for at least one of the operations. In the context of arithmetic, the difference between two numbers is typically defined in a way that takes advantage of the existence of the inverse of every number under addition. The difference of 5 and 3 is 5 + (-3), and the difference of 3 and 5 is 3 + (-5). But more generally, the notion of the difference of two elements a and b can be formalized in the context of a group as the operation of the group on a and b^{-1} (the inverse of b with respect to the operation of the group).

When we turn to the semiring of strings, we find (to our surprise) that we have an embarrassment of riches: we can define inverse elements *either* with respect to the concatenation operator, *or* with respect to the set union operator – if we replace the sets of our ring by multisets. The inverse elements with respect to concatenation we must introduce in any event, but its use can be restricted to a smaller set of cases if we

1 We are pleased to dedicate this paper to François Bavaud, in the belief that our efforts to learn more about the nature of natural language through mathematical formalization will resonate with his own interest in exploring many aspects of natural language through mathematics over the course of a long and productive career.

introduce the use of multisets. This leads us to more than one solution for some of the areas which we explore, and we are led to compare the approaches to the problems we consider.[2]

We will briefly outline these two accounts. In the account in which the elements of the alphabet have inverses with respect to the operation of *concatenation*, we do not need multisets; we can restrict our attention to traditional sets. Each element of the alphabet a_i is allowed an inverse element which we denote a_i^{-1}. Thus $walkingg^{-1}n^{-1}i^{-1} = walk$, and $(the)^{-1} = e^{-1}h^{-1}t^{-1}$.

The *second* account of the notion of *difference* employs not an inverse for concatenation, but rather for union with the extension from sets to *multisets*, by which we mean sets in which each element is associated with a *multiplicity*, which we define to be a number in Z (and for the sake of brevity, we will typically refer to multisets as *m-sets*). This provision allows for an element to be a member of a multiset with negative multiplicity (or, much less interestingly, zero multiplicity). The multiplicity of a_i in A is indicated $\mu_A(a_i)$, where the subscript of μ can be omitted if context makes it clear which multiset we are concerned with. The union of two m-sets A and B is defined as the set union of the elements with non-zero multiplicity in either A or B, and the multiplicity of an element a in $A \cup B$ is defined as the sum of the multiplicities of a in A and in B. We adopt the convention that we have employed so far: the name of an individual string is expressed by a lower case letter, such as a or a_j, while the name of an m-set is expressed by a capital letter, such as A. It is both convenient and natural to indicate elements whose multiplicity is 1 with no special marking, and to indicate elements whose multiplicity is -1 with a prefixed minus sign: $-a$. A multiset $\{a, -b\}$ should *not* be thought of as containing an element denoted "$-b$"; rather, it contains an element b of multiplicity -1. This may lead to some confusion if we do not keep the definition clearly in mind. Thus it is true that $--A = A$, for any m-set A, but $\{--a\}$ does not equal $\{a\}$, and neither $--a$ nor $\{--a\}$ is a

2 There is no reason to be uncomfortable with the introduction of inverses with respect to both operations on the base set, as we do when we construct a field.

meaningful expression.[3]

We distinguish very sharply between the inverse for concatenation, expressed x^{-1}, and the inverse for multiset union, expressed with a leading minus sign, "$-x$". Let us say a few words about each approach before continuing

3 Concatenative inverses

In this approach, we augment the alphabet by adding for each letter in A an inverse. The inverse of a will be indicated a^{-1}, that of b as b^{-1} and so on: $aa^{-1} = a^{-1}a = \varepsilon$. The set of all inverses of the letters in A is noted as A^{-1}. We will use the capital script $\mathcal{A}$ to indicate the original alphabet A augmented with both the null symbol ε and the set of all inverses A^{-1}. $\mathcal{A}$ with the concatenation operator is thus isomorphic to the free group over the original alphabet A. The string cat has an inverse $(cat)^{-1}$, which equals $(t^{-1}a^{-1}c^{-1})$, and in similar fashion, $walking(ing)^{-1} = walk$. We can easily define a notion of difference of two strings, but we must distinguish between left difference and right difference. The *right difference* between *walks* and *walking* is $(walking)^{-1}walks = (ing)^{-1}s$, while the *left difference* between *walks* and *walking* is $walks(walking)^{-1}$, which does not simplify algebraically. The left difference between *view* and *preview* is $(pre)^{-1}$.

4 Multiset inverses

Multisets as we have defined them form a group under the operation of union, since any m-set has an inverse, in a natural way. If an m-set A consists of elements a_i, with corresponding multiplicities μ_i, then A's inverse is the m-set consisting of the same elements with multiplicities $-1 \cdot \mu_A$, and we will indicate the inverse of a multiset A as $-A$. We may also define $A - B$ as $A \cup -B$. For example, if $A = \{a, b, c\}$ and $B = \{a, b, d\}$, then $A - B = \{c, -d\}$. $B - A$, the difference between B and A, is $\{-c, d\}$.

3 The intersection operator from the algebra of sets does not carry over naturally to this version of multisets.

string s	string t	$\Delta_{\text{R}}(s,t) = \frac{s}{t}\text{R}$	$\Delta_{\text{L}}(s,t) = \frac{s}{t}\text{L}$
$walked$	$walking$	$\frac{ed}{ing}$	$\frac{walked}{walking}$
$walk$	$walking$	$\frac{\varepsilon}{ing}$	$\frac{walk}{walking}$
$walk$	$jump$	$\frac{walk}{jump}$	$\frac{walk}{jump}$
$walked$	$jumped$	$\frac{walked}{jumped}$	$\frac{walk}{jump}$
$remind$	$mind$	$\frac{remind}{mind}$	$\frac{re}{\varepsilon}$

FIGURE 1 – Some examples of string difference.

5 Pairwise differences

We return to our discussion of both approaches. In many of the cases that will be of interest, we are interested in the difference between two strings, or between two multisets, each consisting of just one string. The two approaches yield slightly different results for strings x and y. In the case of the concatenation inverse, we must specify whether we intend a left difference or the right difference of x and y; these are, respectively, xy^{-1} and $y^{-1}x$. We indicate these as $\Delta_{\text{L}}(x,y)$ and $\Delta_{\text{R}}(x,y)$. $\Delta_{\text{R}}(walks, walking) = \Delta_{\text{R}}(s, ing) = (ing)^{-1}s$. See figure 1 for further examples.

In the case of multiset union, the difference between the multiset $A = \{x\}$ and $B = \{y\}$ consists of a multiset with one element of multiplicity 1 and one of multiplicity -1: $\{x, -y\}$, and there is no distinction made between left and right difference. However, it is also true that the difference between *walks* and *walking* is $\{walks, -walking\} = \{walk\}\{s, -ing\}$, and the difference between *walking* and *walking* is $\{walk, -walk\}\{ing\}$.

It is convenient to express this more directly, and to use a notation that can be used for either approach to differences, and for this we use the notation $\frac{x}{y}$, $\frac{x}{y}\text{L}$, and $\frac{x}{y}\text{R}$. In the context of multiset inverses, we define $\{\frac{a}{b}\}$ as $\{a, -b\}$. In the context of concatenation inverse, we define $\frac{x}{y}\text{L}$ as $\Delta_{\text{L}}(x,y)$, and $\frac{x}{y}\text{R}$ as $\Delta_{\text{R}}(x,y)$.

We define the operation of concatenation of multisets of strings in the natural way, given a definition of concatenation of strings, in order

to ensure that concatenation distributes properly over the operation of (multiset) union. When it is helpful, we use the symbol $\times$ to mark concatenation, either of strings or of m-sets; when it leads to no confusion, we omit that symbol. The concatenation of two multisets of strings A and B is defined as the m-set of all strings of the form $a_i b_j$, $a_i \in A$, $b_j \in B$, and the multiplicity of $a_i b_j$ is $\mu_A(a_i) \cdot \mu_B(b_j)$. The following true statements illustrate the sort of descriptions we will explore:

1. $\{walks, jumps\} = \{walk, jump\}\{s\}$

2. $\{walk,\ jump\}\ \{s,\ ed,\ ing\} = \{walks,\ walked,\ walking,\ jumps,\ jumped, jumping\}$

3. $\{walks\} - \{jumps\} = \{walk, -jump\}\{s\} = \{\frac{walk}{jump}\}\{s\}$

4. $\{outperform\} - \{perform\} = \{out, -\varepsilon\}\{perform\} = \{\frac{out}{\varepsilon}\}\{perform\}$

5. $\{hard,\ soft\}\{\ en\ \}\{\ \varepsilon,\ s,\ ed,\ ing\ \} = \{harden,\ hardens,\ hardened,\ hardening,\ soften,\ softens,\ softened,\ softening\}$

6. $\{hard,\ harder,\ hardest,\ harden,\ hardens,\ hardened,\ hardening\} = \{hard\}\left\{\left\{\begin{array}{c}\varepsilon\\ er\\ est\end{array}\right\} \cup \{en\}\{\varepsilon, s,\ ed,\ ing\}\right\}$

7. $\{sing\} - \{sang\} = \{s\left\{\frac{i}{a}\right\} ng\}$

Observe that in the traditional semiring with concatenation of strings, union plays the role of addition and concatenation plays the multiplicative role, because concatenation is distributive over union, while union is not distributive over concatenation.[4] For a semiring to be a ring, there must be inverses for each element under the additive operation, which in our case reduces to the m-set union. Thus if we introduce m-set

4 That is, $\{a\} \cup (\{b\} \times \{c\}) = \{a, bc\}$ does not in general equal $(\{a\} \cup \{b\}) \times (\{a\} \cup \{c\}) = \{a, b\} \times \{a, c\} = \{aa, ac, ba, bc\}$.

inverses, the structure we are exploring is a ring, while if we adopt only concatenation inverses, we are not exploring a ring.

The examples in this paper are all taken from the standard orthography of English, but all of it applies equally to a transcription of a language in a phonetic or phonological form.

The ultimate goal of this work is to establish a counterpoint to generative grammar, in the following sense. Generative grammar takes as its task to elucidate the principles, both universal and language-particular, which accounts for the observed data on the basis of knowledge of the smallest relevant units in the language (typically morphemes and/or words). One analyzes a chosen utterance on the basis of the smallest underlying forms, which are already known to the linguist, employing the universal/language-particular principles that are called upon. This generative analysis, however, says nothing about the (epistemological) origin of these units and thereby leaves half of the account of the language untouched. American descriptivists, such as Zellig Harris, often referred to the tasks of segmentation and classification, with these tasks in mind, and the present paper aims to better understand what those principles would be when viewed from today's computational point of view.

6 Oppositions

6.1 The origins of the concept

The term *opposition* was proposed by Trubetzkoy in his major work, *Grundzüge der Phonologie* (Trubetzkoy, 1939), and it is closely related to the notion of difference between two elements in a language. Trubetzkoy proposed that two elements in a language can be put into a relationship with each other called an *opposition*, which consists of two things: first, a statement as to what they share in common, and secondly, a statement of what the first element possesses that the second does not, and of what the second possesses that the first does not, and the two of them together constitute a difference of the sort we have been discussing. If what we have called "properties" can be thought of as members of an m-set along the lines we described in the preceding section, then the statement of the properties that A possesses but that

B does not can be expressed as a multiset in which A's distinctive properties have a positive multiplicity and B's distinctive properties have a negative multiplicity. For Trubetzkoy and the structuralists who followed him, different items in the grammar of a language could be put into oppositions with one another: one phoneme could be put into opposition with another, one word with another, one case with another, and so on.

Consider a simple case, such as the phonological opposition $Opp(pat, mat)$ between the words pat and mat. What these two words have in common is $\_at$, which is to say, the sequence at positioned to the right of something else (the element of positioning is what is indicated by the underscore "_"), and their difference is the difference between a p and an m. What is the opposition $Opp(p, m)$ between p and m? That question gives rise to a second order opposition, that is, an opposition that arises because of the definition of a first order opposition; we shall see other sorts of second order oppositions below. The second order opposition here consists of a commonality (of p and m), and the differences. To describe this, we need recourse to features; features are the linguist's way of describing second order oppositions in phonology. What p and m have in common is a point of articulation – in particular, labial point of articulation. How they differ is that the first is voiceless and oral, and the second is nasal.

In this paper, we focus on oppositions between strings, and just stray briefly into the elements that permit us to discuss oppositions between individual elements, which are the "phonological features" that grew out of Trubetzkoy's and Jakobson's conception of phonology.

We may return now to this question: what *is* an opposition between two strings? By definition it is two things: a statement of what the pair of words have in common, and a statement of how each differs from the other. In the case of strings, a natural statement (but not the only reasonable statement) of what they have in common is to produce, first, a *substring* present at the left edge of both strings, or at the right edge of both strings, along with a statement as to where that common substring appears in the two words; and second, a statement of how they *differ*, which is a string difference of what remains if we remove the

string they have in common.[5] In the case we are considering, there is a natural connection between this and the distributivity of concatenation over m-set union, since extracting a maximal common substring at an edge is no different from using the distributive identity in a maximal fashion; that is, just as $\{walked, walking\} = \{walk\}\{ed, ing\}$, so $\{walked\} - \{walking\} = \{\frac{walked}{walking}\} = \{walk\}\{\frac{ed}{ing}\}$.

When we go beyond phonology, we compare more than simply strings or structures of sounds; we compare words with word-particular information (which are called **semantic** and **morphological** much of the time), since even the information that we consider syntactic we usually refer to as *morphosyntactic* when we consider words out of syntactic context.

When we consider the opposition $(walks/walking)$, we ask what they have in common and what they do not have in common. What they have in common is a string, $walk$, which precedes their difference, and semantic information, as well as the morphosyntactic information that we call "Verb". They differ with the string opposition (s/ing) and morphosyntactic features that differ between these two suffixes.

In this case, we do not get much in return for asking what (s/ing) have in common and how they differ (this is different from the phonological case, where there is a difference between information in a string and information "inside" a phoneme, as we noted when asking what the difference was between p and m.) In this case, if the two suffixes have some morphosyntactic information in common, it is natural to associate it with the stem, i.e., the commonality.

As Trubetzkoy notes, there are cases where there is a natural ordering of the two elements in the difference. In some cases, the opposition-difference is something versus nothing $(walking/walk)$ and in some cases the opposition difference is "less" versus "more". This latter kind of opposition is natural and even central in semantics, and more limited in the domain of phonology.

Thus we return to the central example of a binary word-based

5 Lee (2002) provides a detailed examination of alternative definitions of string commonality in this context. An important point to bear in mind is that while asking what two objects have in common is a perfectly meaningful question, it may have more than one reasonable answer.

opposition as a common stem plus an ordered pair of two affixes, specified either as prefixal or as suffixal (i.e., indicating if they precede or follow the common stem). An opposition is an ordered pair, then, of two things: a commonality and a difference, and so from a logical point of view an opposition is an ordered pair of a commonality and a difference. This definition is intended to be extremely general, and by no means restricted to strings. Given its importance here, we propose to indicate this with its own notation: in an opposition between X and Y, indicated $[A, \frac{B}{C}]_{Opp}$, A is what X and Y have in common, B is what X has and Y does not, and C designates what Y has and X does not. That statement summarizes the result of this section.

Before continuing, let us review some classic observations about morphological (i.e., word-internal) structure. The set of words $\{truck, train, travel, trip\}$ equals $\{tr\}\{uck, ain, avel, ip\}$ (and, as one can easily see, all of the words share a semantic component somehow relating to locomotion, to put it awkwardly). But this analysis is amusing rather than insightful (or grammatical). There are two ways of expressing why this analysis is not significant. In the terms that we are proposing, it is because the commonalities of any pair of these four words is different (the opposition of $truck$ and $train$ begins with the commonality of the two words, which includes the shared meaning of "vehicle", while the opposition of $truck$ and $trip$ includes nothing of that sort in the commonality). A related but nonetheless distinct way of expressing the irrelevance lies in Greenberg's condition on word analysis (morphemic analysis), which requires that a morphemic analysis contains minimally two items, as expressed in the classic Greenberg rectangle (see Greenberg, 1960).

$walk$	ed
$jump$	ing

which could also be expressed as $\left\{ \begin{array}{c} walk \\ jump \end{array} \right\} \left\{ \begin{array}{c} ed \\ ing \end{array} \right\}$.

We believe that morphology emerges with the presence of at least two oppositions which share the same difference. It is natural to define

an operation of union on oppositions, but only on pairs (or sets) of oppositions that have identical differences, such as $[W, \frac{Y}{Z}]_{Opp}$ and $[X, \frac{Y}{Z}]_{Opp}$. Here we define their union as:

$$[W, \frac{Y}{Z}]_{Opp} \cup [X, \frac{Y}{Z}]_{Opp} := [W \cup X, \frac{Y}{Z}]_{Opp}$$

Given the two oppositions $[walk\_, \frac{ed}{ing}]_{Opp}$ and $[jump\_, \frac{ed}{ing}]_{Opp}$, their union is $[walk\_ \cup jump\_, \frac{ed}{ing}]_{Opp}$. With this, we turn to the notion of a paradigm in a morphology, where that property of opposition union plays a central role.

6.2 Oppositions within a paradigm

Let us begin by defining a paradigm as simply a set of words, recognizing that in the real world a good deal more is intended when one speaks of a paradigm. Two typical paradigms that we will be interested in is $P_{walk} = \{walk, walks, walked, walking\}$ and $P_{move} = \{move, moves, moved, moving\}$.

Let us first consider the opposition of a paradigm with itself, a *self-opposition*, which can be naturally thought of as an array of all of the pairwise oppositions of distinct members of the paradigm, as in figure 2 (where we have stacked the arguments for ease of presentation), a sort of outer product of the paradigm with itself, where we use Φ to indicate the opposition of something with itself, an essentially useless object.

P_{walk}	$walk$	$walks$	$walked$	$walking$
$walk$	Φ	$Opp\begin{pmatrix} walk, \\ walks \end{pmatrix}$	$Opp\begin{pmatrix} walk, \\ walked \end{pmatrix}$	$Opp\begin{pmatrix} walk, \\ walking \end{pmatrix}$
$walks$	$Opp\begin{pmatrix} walks, \\ walk \end{pmatrix}$	Φ	$Opp\begin{pmatrix} walks, \\ walked \end{pmatrix}$	$Opp\begin{pmatrix} walks, \\ walking \end{pmatrix}$
$walked$	$Opp\begin{pmatrix} walked, \\ walk \end{pmatrix}$	$Opp\begin{pmatrix} walked, \\ walks \end{pmatrix}$	Φ	$Opp\begin{pmatrix} walked, \\ walking \end{pmatrix}$
$walking$	$Opp\begin{pmatrix} walking, \\ walk \end{pmatrix}$	$Opp\begin{pmatrix} walking, \\ walks \end{pmatrix}$	$Opp\begin{pmatrix} walking, \\ walked \end{pmatrix}$	Φ

FIGURE 2 – Self-opposition of a paradigm.

In looking at an array of pairwise oppositions, it is natural to separate it into two arrays, one for the commonalities, and one for the differences, and we have done just this, showing the commonalities in figure 3, and the differences in figure 4 (we omit diagonal elements throughout). (Here as elsewhere, the expression $\frac{a}{b}$ denotes $\{a, -b\}$, in the multiset interpretation, or $\{a, b^{-1}\}$ in the concatenation interpretation.)

P_{walk}	$walk$	$walks$	$walked$	$walking$
$walk$	Φ	$walk\_$	$walk\_$	$walk\_$
$walks$	$walk\_$	Φ	$walk\_$	$walk\_$
$walked$	$walk\_$	$walk\_$	Φ	$walk\_$
$walking$	$walk\_$	$walk\_$	$walk\_$	Φ

FIGURE 3 – Left-edge commonalities in the paradigm P_{walk}.

P_{walk}	$walk$	$walks$	$walked$	$walking$	
$walk$	Φ	$\frac{\varepsilon}{s}$	$\frac{\varepsilon}{ed}$	$\frac{\varepsilon}{ing}$	$\boldsymbol{\varepsilon}$
$walks$	$\frac{s}{\varepsilon}$	Φ	$\frac{s}{ed}$	$\frac{s}{ing}$	$\boldsymbol{s}$
$walked$	$\frac{ed}{\varepsilon}$	$\frac{ed}{s}$	Φ	$\frac{ed}{ing}$	$\boldsymbol{ed}$
$walking$	$\frac{ing}{\varepsilon}$	$\frac{ing}{s}$	$\frac{ing}{ed}$	Φ	$\boldsymbol{ing}$
	$\boldsymbol{\varepsilon}$	$\boldsymbol{s}$	$\boldsymbol{ed}$	$\boldsymbol{ing}$	

FIGURE 4 – Right differences in P_{walk}.

Thus prefixes or suffixes emerge from the description of the differences between members of a paradigm in an outer product of the oppositions of the members of the paradigm.

An opposition is by its nature a relation between two objects, but a paradigm is in general a larger set of forms (larger than two, that is), and part of what holds it together conceptually is that all of its members share something in common – which we often call its stem. It is thus natural to expand the concept of a binary opposition to a paradigm-like

set of forms under the condition that each pair of items is analyzed as an opposition, but one in which the all pairs share the same commonality (here, the stem). We will refer to such cases, the sort in figures 3 and 4, as *pure paradigms.*

e-final verbal pattern		
$move$	$love$	$hate$
$moves$	$loves$	$hates$
$moved$	$loved$	$hated$
$moving$	$loving$	$hating$

FIGURE 5 – English e-final verb stems.

In the case of verbs in English such as $walk$ or $jump$, the array of commonalities is constant throughout, but in the case of other verbal paradigms, the commonalities in some pairs is different from the commonalities in other pairs. We consider first e-final stems, as illustrated in figure 5, and analyzed in figure 6, and we see that in some cases, the commonality is mov and in others it is $move$. From a simple logical point of view, satisfying the condition that all commonalities be the same appears to be met by making the stem smaller and smaller, so to speak: in this case, making it mov, and changing the analysis to figure 7.

In work not reported here, we extend the computation of oppositions to the case of opposition between two self-oppositions of paradigms; that is, in the case illustrated here, we define the opposition between (for example) P_{move} and P_{walk}.

7 Conclusion

We hope to have provided a small view of a mathematical way of understanding differences that arise in a systematic way in natural language. In work in progress, we extend this notion of difference from differences within paradigms to differences across paradigms, to better understand how languages employ large families of pairs of differences, where the oppositions within the families are constant, and where the

P_{move}	*move*	*moves*	*moved*	*moving*
move		*move*	*move*	*mov*
moves	*move*		*move*	*mov*
moved	*move*	*move*		*mov*
moving	*mov*	*mov*	*mov*	

P_{move}	*move*	*moves*	*moved*	*moving*	
move	Φ	$\frac{\varepsilon}{s}$	$\frac{\varepsilon}{d}$	$\frac{e}{ing}$	$\boldsymbol{e, \varepsilon}$
moves	$\frac{s}{\varepsilon}$	Φ	$\frac{s}{d}$	$\frac{es}{ing}$	$\boldsymbol{s, es}$
moved	$\frac{d}{\varepsilon}$	$\frac{d}{s}$	Φ	$\frac{ed}{ing}$	$\boldsymbol{d, ed}$
moving	$\frac{ing}{e}$	$\frac{ing}{es}$	$\frac{ing}{ed}$	Φ	$\boldsymbol{ing}$
	$\boldsymbol{e, \varepsilon}$	$\boldsymbol{e, es}$	$\boldsymbol{d, ed}$	$\boldsymbol{ing}$	

FIGURE 6 – Analysis 1 of an *e*-final stem: *not* a pure paradigm (mixed stems).

M_{move}	*move*	*moves*	*moved*	*moving*
move		*mov*	*mov*	*mov*
moves	*mov*		*mov*	*mov*
moved	*mov*	*mov*		*mov*
moving	*mov*	*mov*	*mov*	

M_{move}	*move*	*moves*	*moved*	*moving*	
move	Φ	$\frac{e}{es}$	$\frac{e}{ed}$	$\frac{e}{ing}$	$\boldsymbol{e}$
moves	$\frac{es}{e}$	Φ	$\frac{es}{ed}$	$\frac{es}{ing}$	$\boldsymbol{es}$
moved	$\frac{ed}{e}$	$\frac{ed}{es}$	Φ	$\frac{ed}{ing}$	$\boldsymbol{ed}$
moving	$\frac{ing}{e}$	$\frac{ing}{es}$	$\frac{ing}{ed}$	Φ	$\boldsymbol{ing}$
	$\boldsymbol{e}$	$\boldsymbol{es}$	$\boldsymbol{ed}$	$\boldsymbol{ing}$	

FIGURE 7 – Analysis 2 of *e*-final stem, with stem *mov*-, a pure paradigm.

cross-family differences are themselves bounded in certain respects. For example, there are two distinct families of inflections for French verbs including *choisir* and *partir*, respectively. Each family is defined by the pairwise differences within it, but a higher order set of differences can be computed that deals with the differences across the two families. We have employed some of this work in software for learning morphology, employing Minimum Description Length methods for calculating the information content of oppositions and sets of oppositions.

Acknowledgements

Thanks to Paul Goldsmith-Pinkham and Jason Riggle for helpful comments on an early version of this paper.

References

Greenberg, J. H. (1960). A quantitative approach to the morphological typology of language. *International Journal of American Linguistics*, 26(3):178–194.

Lee, J. L. (2002). *Morphological paradigms: Computational structure and unsupervised learning*. PhD thesis, University of Chicago.

Trubetzkoy, N. (1939). *Grundzüge der Phonologie*. Number 7 in Travaux du Cercle Linguistique de Prague. Cercle linguistique de Prague, Prague.