14th International Morphology Meeting

Budapest, Hungary
13-16 May 2010

Workshop
Stems in Inflection and Lexeme Formation

ABSTRACTS

organization:

Olivier Bonami
Gilles Boyé
Fabio Montermini

http://www.nytud.hu/imm14/stemsworkshop.pdf
## Contents

Stems, roots and exponents in the verbal system of Lithuanian  
*Peter M. Arkadiev*  

*entia non sunt multiplicanda praeter necessitatem* (Occam), *sed aliquando necesse est*: an argument for morphological catégories  
*Mark Aronoff*  

Stems in a database of morphological complexity  
*Matthew Baerman & Scott Collier*  

Stem alternations and multiple exponence  
*Matthew Baerman & Greville G. Corbett*  

Paradigm projections in the French verb system  
*Basilio Calderone*  

Stem space and derivation in French  
*Bernard Fradin & Delphine Tribout*  

What distinguishes stem allomorphy? A masked priming study with French stimuli  
*Hélène Giraudo & Madeleine Voga*  

Morphotactic transparency. Stems and suffixes in English and Polish inflection  
*Paula Orzechowska*  

On the intraparadigmatic allomorphy of derivational affixes (the case of Lithuanian verbs)  
*Jurgis Pakerys*  

Stem dependency and irregular verbs in Tamil  
*Bhavani Saravanann*  

Maximising stems  
*Andrew Spencer*
Stems, roots and exponents in the verbal system of Lithuanian

Peter M. Arkadiev

Center for Balto-Slavic research, Institute of Slavic Studies,
Russian Academy of Sciences

1. The verbal system of Lithuanian is fairly complex and admits of cross-classifications according to different formal criteria (see Ambrasas (ed.) 1997 for a traditional analysis, and Dressler et al. 2006 for a novel one). The most important two of these are the choice of the marker of the Present tense (sometimes somewhat controversially called “thematic vowel”), yielding -a, -i, and -o conjugations, and the presence of special vocalic suffices in different parts of the paradigm, yielding the classification of verbs into “primary” (no such suffices in any cell of the paradigm), “mixed” (a “thematic suffix” is present in the Infinitive stem and/or in the Past stem, but absent in the Present stem), and “suffixal” (with genuine derivational suffixes present in all cells of the paradigm).

2. Lithuanian verbal system is organized around three “principal parts” or stems, whose form is in many cases lexically specified (the greatest variety of patterns is attested with the “primary” verbs), but from which all the other verbal forms can be constructed by the application of more or less straightforward general rules. These are the Infinitive, the Present, and the Past stems. In addition to the three stems, the verbal root must be also considered, since it is from the root that the stems are build, and because it serves as the base of many derivational processes. It is important to note that the root is not necessarily identical to any of the stems. In Fig. 1, the examples of different Lithuanian verb classes are given, showing the types of stem-formation.

Fig. 1. Lithuanian verbal stems and classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Root</th>
<th>Infinitive</th>
<th>Present</th>
<th>Past</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>-bėg-</td>
<td>bėg-ti</td>
<td>bėg-a</td>
<td>bėg-o</td>
<td>‘run’</td>
</tr>
<tr>
<td>(a-present only)</td>
<td>-vilk-</td>
<td>vilk-ti</td>
<td>velk-a</td>
<td>vilk-o</td>
<td>‘drag’</td>
</tr>
<tr>
<td>-dauž-</td>
<td>dauž-ti</td>
<td>daužį/-a</td>
<td>daužį/-ė</td>
<td>‘break’ (tr.)</td>
<td></td>
</tr>
<tr>
<td>-vog-</td>
<td>vog-ti</td>
<td>vog/-a</td>
<td>vog/-ė</td>
<td>‘steal’</td>
<td></td>
</tr>
<tr>
<td>-gim-</td>
<td>gim-ti</td>
<td>gim-a</td>
<td>gim/-ė</td>
<td>‘defend’</td>
<td></td>
</tr>
<tr>
<td>-et-</td>
<td>et-ti</td>
<td>etn-a</td>
<td>etn/-o</td>
<td>‘go’</td>
<td></td>
</tr>
<tr>
<td>-alp-</td>
<td>alp-ti</td>
<td>alp-st-a</td>
<td>alp-o</td>
<td>‘faint’</td>
<td></td>
</tr>
<tr>
<td>-leid-</td>
<td>leis-ti</td>
<td>leidži-a</td>
<td>leid-o</td>
<td>‘let’</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>a-pres.</td>
<td>-gied-</td>
<td>giedo-ti</td>
<td>gied-o</td>
<td>‘sing’</td>
</tr>
<tr>
<td>i-pres.</td>
<td>-myl-</td>
<td>myl-ti</td>
<td>myl-i</td>
<td>mylė-o</td>
<td>‘love’</td>
</tr>
<tr>
<td>o-pres.</td>
<td>-raš-</td>
<td>raš-ti</td>
<td>raš-o</td>
<td>raš/-ė</td>
<td>‘write’</td>
</tr>
<tr>
<td>Derived</td>
<td>-bal+in-</td>
<td>balin-ti</td>
<td>balin-a</td>
<td>balin-o</td>
<td>‘whitewash’</td>
</tr>
<tr>
<td>(a-present only)</td>
<td>-važ+iuo-</td>
<td>važiuo-ti</td>
<td>važiuo-a</td>
<td>važiau-o</td>
<td>‘drive’</td>
</tr>
</tbody>
</table>

3. The three main stems do not actually share the same status within the verbal system. Fig. 2 schematizes the relations between the stems and the different parts of the paradigm based on the respective stems.
Fig. 2. Lithuanian verb stems and forms based on them

<table>
<thead>
<tr>
<th>Present finite forms</th>
<th>Past finite forms</th>
<th>Infinitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present Active and</td>
<td>Past Active Participle</td>
<td>Habitual Past</td>
</tr>
<tr>
<td>Passive Participles</td>
<td>Verbal Noun</td>
<td>Future</td>
</tr>
</tbody>
</table>

From Fig. 2 it is evident that the Present stem, whose formation rules are particularly diverse and include suffixation, infixation, qualitative and quantitative alternations of vowels etc., can be considered to be the exponence of the Present tense, since all and only Present tense formations are based on this stem. The Past stem, though clearly related to the expression of the Past tense, already is to some degree “morphemic”, because at least one of the Past forms (the Past Passive Participle) is not derived from the Past stem, while a form with no clear relation to the Past tense (the Verbal Noun) is based on the Past stem. Finally, the Infinitive stem is clearly “morphemic”, since the forms derived from do not form a morphosyntactically coherent set.

4. The Infinitive stem might be considered simply an “elsewhere” option covering those forms which failed to be derived from the other independently motivated stems. However, it is not the whole story, since the relation between the Infinitive stem and the forms based thereon is not straightforward. First of all, on the basis of the morphophonological and accentual peculiarities of the forms based on the Infinitive stem it is possible to argue that the latter is derived by subtraction of -ti from the Infinitive (and not vice versa). The Infinitive is thus derived not from “its own” stem, but directly from the root. Second, the same kind of data suggest that the Past Passive Participle, despite its superficial formal similarity to the Infinitive, should also be treated as directly based on the root rather than on the Infinitive stem.

5. Thus, the following picture of the Lithuanian verbal system emerges. First, there is the root, from which at least two forms (Infinitive and Past Passive Participle) and two stems (Present and Past) are directly derived by various rules (from fairly general to completely idiosyncratic). Second, the Infinitive stem is derived from the Infinitive by subtraction. Third, various finite and non-finite forms are derived from the three stems by general rules applying to all types of verbs.

To conclude, I argue that the notion of stem plays a central role in the Lithuanian verbal inflection. The Lithuanian data clearly show that the relation between the more abstract stems and the particular verbal forms may be quite complex and not necessarily uniform, and also that such property of stems as their being “morphemic” in Aronoff’s (1994) sense can be a matter of degree.

References
entia non sunt multiplicanda praeter necessitatem
(Occam), sed aliquando necessesse est:
an argument for morphological categories

Mark Aronoff
Stony Brook University

Occam’s razor is especially apt in these newly frugal times: if you don’t need it, don’t get it! In his wisdom, Occam never defined necessitas, because one person’s need is another’s excess (speak to any teenager about this point). So why do I feel that morphologists need stems rather than just want them?

In any discipline, we posit a new entity because we need to make a distinction that is otherwise difficult or impossible to make. The initial distinction that we need stems for is descriptive: in order to describe certain morphological phenomena, we need to distinguish the form to which a particular morphological operation applies from the lexeme of which that form is a manifestation. At this point, we ask whether this descriptive necessity is reflective of a deeper theoretical necessity: do stems exist as independent rather than purely derived objects and does positing the existence of stems, this act of multiplying entities, lead to a better understanding of language. I will show that it does, that we really do need stems.

Finally, we must ask whether stems can be reduced to more basic entities, as some have argued. I will show that arguments against positing stems are misconceived: stems can be done away with only if are willing to ignore certain properties that hold of stems but not of their parts. The attempt to do away with stems is thus similar in kind to attempts to do away with other complex entities: elements, molecules, or species. Unless we can explain all the properties of elements, molecules, or species in terms of their parts, we are forced to multiply entities. So too with stems.
Stems in a database of morphological complexity

Matthew Baerman  
Scott Collier  
Surrey Morphology Group

This paper describes a database on morphological complexity currently under construction, and in particular the role that stem alternations and stem classes play within it. Morphological complexity is understood here as structure found in morphological forms that is not accounted for by a simple mapping from morphosyntactic/morphosemantic feature values to morphology. Key examples of this are inflectional classes, which introduce arbitrary distinctions in form, and syncretism, which introduces arbitrary classes of forms within the paradigm.

In what we may call a canonical system, inflection is the job of affixes while stems remain invariant. If this were always true then stems would play no meaningful role in the database. But stems may display much the same behaviour as affixes, varying across lexemes (stem classes) as well marking feature values (stem alternations), and so are evaluated in the database along exactly the same parameters as affixes.

The type of data we are interested in is represented in the table below, showing (a fragment of) Latvian noun paradigms.

<table>
<thead>
<tr>
<th>example forms</th>
<th>analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>infl. class</td>
<td>A</td>
</tr>
<tr>
<td>gender</td>
<td>m</td>
</tr>
<tr>
<td>‘father’</td>
<td>tēvs</td>
</tr>
<tr>
<td>sg nom</td>
<td>tēvu</td>
</tr>
<tr>
<td>sg gen</td>
<td>tēva</td>
</tr>
<tr>
<td>pl nom</td>
<td>tēvi</td>
</tr>
<tr>
<td>pl gen</td>
<td>tēvus</td>
</tr>
<tr>
<td>infl. class</td>
<td>A</td>
</tr>
<tr>
<td>gender</td>
<td>m</td>
</tr>
<tr>
<td>inflections</td>
<td>stem</td>
</tr>
<tr>
<td>sg nom</td>
<td>PLAIN s</td>
</tr>
<tr>
<td>sg acc</td>
<td>PLAIN u</td>
</tr>
<tr>
<td>sg gen</td>
<td>PLAIN a</td>
</tr>
<tr>
<td>pl nom</td>
<td>PLAIN i</td>
</tr>
<tr>
<td>pl gen</td>
<td>PLAIN u</td>
</tr>
</tbody>
</table>

The structure of the database itself is simple, distinguishing (i) FORMS (ordered symbols within the cells), and (ii) VALUES (indexes/labels given to the rows and columns). Values are of four types:

(i) MORPHOSYNTACTIC VALUE (e.g. plural accusative)  
(ii) INFECTION TYPE, namely an inflectional position class or slot, e.g. prefix, suffix, stem, tone etc. – however many parallel subsystems one chooses to recognize.  
(iii) LEXICAL TYPE: a property of the lexeme which is constant throughout the paradigm and may have correlations with other aspects of inflectional behaviour, e.g. gender or some phonological feature of the stem.  
(iv) INFECTION CLASS is an index which compiles the components of the paradigm for inflectionally distinct classes of lexemes. Thus in the Latvian example, each unique pairing of stem forms and affix forms is assigned an inflection class index.
A database query for a sample stem form (PAL, the palatalized stem) would return:

<table>
<thead>
<tr>
<th>form</th>
<th>(i) morphosyntactic value</th>
<th>(ii) inflection type</th>
<th>(iii) lexical type</th>
<th>(iv) inflection class</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAL sg gen</td>
<td>stem</td>
<td>masculine</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>PAL pl nom</td>
<td>stem</td>
<td>masculine</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>PAL pl acc</td>
<td>stem</td>
<td>masculine</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>PAL pl gen</td>
<td>stem</td>
<td>masculine</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>PAL pl gen</td>
<td>stem</td>
<td>feminine</td>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

The main interest in the database lies in the classes which are dynamically generated by having identical cell contents within a given column. This can be modified or restricted by combining information from multiple columns. For example, we can look for all the values that may be realized by a palatalized stem, or restrict the query to the distribution of the palatalized stem within individual paradigms. Further, we can look for correlations between stem alternations and other properties (e.g. the correlation shown above of genitive plural stem palatalization and feminine gender).

While the notion of stem has a long tradition, and is borne out by examples such as Latvian, in many cases inflectional elements display a range of patterns that do not easily lend themselves to clear-cut identification. The database separates these out under the rubric ‘inflection type’, where labels such as ‘affix’ or ‘stem’ need only be understood as labels of convenience. This is illustrated in the multi-layered inflectional patterns of Mazatec, illustrated below by a fragment of a verb paradigm:

<table>
<thead>
<tr>
<th>example forms</th>
<th>analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>inflection class</td>
<td>1</td>
</tr>
<tr>
<td>‘dust’</td>
<td></td>
</tr>
<tr>
<td>1sg</td>
<td>bu(^1) hnû(^1)</td>
</tr>
<tr>
<td>2sg</td>
<td>ntu(^2) hni(^2)</td>
</tr>
<tr>
<td>3</td>
<td>bu(^3) hnû(^3)</td>
</tr>
<tr>
<td>1incl pl</td>
<td>ntu(^2) hnû(^2)</td>
</tr>
<tr>
<td>1excl pl</td>
<td>ntu(^2) hni(^24)</td>
</tr>
<tr>
<td>2pl</td>
<td>ntu(^2) hnû(^2)</td>
</tr>
</tbody>
</table>

The inflected forms consist of (at least) three elements alternating according to different patterns (which in turn are divided into different lexical classes). It is unclear what exactly to call ‘stem’: the invariant \(hn\) or that plus the prefix (which at least diachronically can be seen as a stem formative)? Is tone, which is realized over the whole word form, an autosegmental element or does it belong to the stem? The database does not take a position on these questions, but does provide a way of measuring the contribution of these various elements and so will contribute to a comprehensive typology of inflectional formatives. We are keen to have reactions and input from potential users before implementing the database.
Stem alternations and multiple exponence

Matthew Baerman
Greville G. Corbett
Surrey Morphology Group

Stem alternations can be studied in their own right, but we should not overlook the fact that they typically co-occur with other inflectional marking such as affixation, and thus constitute one element in a network of multiple exponence. How multiple exponence is modelled is a point where morphological theories differ substantially: in some (e.g. Paradigm Function Morphology) it is unconstrained, in others (e.g. Distributed Morphology) there are tightly defined universal constraints, while still other theories (e.g. Network Morphology) take a middle course, allowing dependencies between multiple exponence to be defined on a language-specific basis. Our knowledge of the role of stem alternations within this broader paradigmatic context remains fairly spotty, so we propose a typological framework to help us understand the differing contributions of stems and affixes to the complete paradigm. We distinguish four criteria:

1. Do stem and affix mark different or the same features? An example where they mark different features is Greek verbs (Figure 1), where the stem marks aspect and the affixes person, number and tense. This means that we have distributed rather than multiple exponence. A common assumption in such cases is that the features reserved for stem exponence will be more ‘relevant’ in Bybee’s (1985) terms, as in Greek, with aspect versus subject person-number. Where stem and affix mark the same feature, yielding multiple exponence, further questions arise:

2. Do stem and affix distinguish the same values of the feature? In principle they could redundantly mark exactly the same feature values, but in practice this is restricted to small systems, as in Dholuo (Figure 2), where both stem and affix mark singular versus plural. Otherwise some feature values will be conflated, and here we can distinguish various subtypes. Thus conflation can be bidirectional, as in Chichimec (Figure 3), where both the affix and the stem conflate person values, but according to distinct patterns, giving us a skewed pattern. Or the distinctions found in one component can subsume those found in another, where we can further ask:

3. Is the set of conflated values morphologically defined, or can they be described as a natural class in terms of morphosyntactic features? In Russian (Figure 4) the stem alternation collapses a set of person-number values which is purely morphological (morphemic). Conversely, if the conflated values comprise a natural class, we can ask:

4. Does the conflation embrace an entire feature, or does it correspond to a natural class of values within a feature? For example, in Gothic (Figure 5) stems conflate all the values of the person feature. That is, affixes cumulatively mark a bundle of three features (person, number, and tense), while stems mark only two of these: tense (present nim- versus preterite nam-/nēm-), and within the preterite, number (singular nam- versus plural nēm-). In Slovene (Figure 6) the conflation occurs within a feature, but the pattern is not random. Nominal affixes distinguish three number values: singular, dual and plural, while stems distinguish only two, conflating dual and plural. If one admits decomposition of feature structure, this can be construed as a natural class ‘non-singular’ embracing dual and plural.

While the study of stem alternations has a long history, we suggest this is a novel approach. Prior work has focused on the morphological or phonological operations that derive stem alternations, or on the categories they express (in particular, where suppletion is involved). We propose a typology of the relative contribution of the stem as one element in a network of multiple exponence. While tacit assumptions may exist as to how inflectional distinctions should be distributed amongst the components of the paradigm, no serious cross-linguistic typological study has yet been conducted.
Figure 1: Modern Greek (Holton, Mackridge & Philippaki-Warburton 1997)

Figure 2: Dholuo (Okoth-Okombo 1982, Tucker 1994)

Figure 3: Chichimec (de Angulo 1933)

Figure 4: Russian (Zaliznjak 1977)

Figure 5: Gothic (Wright 1910)

Figure 6: Slovene (Priestly 1993)

References
Paradigm projections in the French verb system

Basilio Calderone
Modyco, CNRS & Université Paris Ouest Nanterre

Though central in most recent morphological theories, the notion of ‘paradigm’ remains unclear from a phenomenal point of view. It seems undeniable that a paradigmatic dimension acts as enforcement of distributed regularities over the lexicon, thus defining morphological micro-structures that account for most aspects of stem-affix conjunction. In this sense, paradigms have a clear ‘emergent’ dimension whose distributional outcomes are easily perceivable, yet an unequivocal definition of paradigms per se is still lacking. In other words, paradigms are most often conceived of in terms of distributional (i.e., statistical) regularities over the lexicon, but in purely linguistic (functional) terms they remain evanescent objects. Possibly such ontological weakness is grounded in the nature of paradigm which is local and global at the same time: contrastive actions at the word level are concretely manifested at a local, sub-lexical level.

In this research, an attempt is made at modeling the French verbal system by means of a topological multi-layered organization of verbal forms appealing solely to distributional information. The computational model exploited here is based on adaptive self-organizing memories, which proved able to organize the inflected forms in input over a topological grid of both paradigmatic (i.e., contrastive) oppositions and syntagmatic (i.e., formal) similarities. Different outputs in the abstraction organization of paradigms will be evaluated from a quantitative point of view. At the same time, the obtained paradigmatic-inflectional knowledge will be tested against generalization processes with respect to inedited input forms, with the specific aim of evaluate the predictive performance of the learning algorithm.

References
Aronoff, M. (1994), Morphology by itself. The MIT Press. 1
1. **Goal** One of the main source of irregularity in the inflectional morphology of Romance languages lies in the fact that the stems used as bases by inflectional rules often change within the same paradigm e.g. Fr. /bwa/ *(I) drink’, /byv-3/ *(we) drink’, /bwav/ *(they) drink’. Introducing Stem space was an attempt to deal in a principled way with the systematic variation introduced by stem changing ((Boyé 2000), (Bonami and Boyé 2003)). Originally, Stem space denotes the zones of the inflectional paradigm (of a lexical category C) within which stems are likely to show alternation. In French, there are 12 such zones in the paradigm of verbs. (Bonami, Boyé, and Kerleroux 2009) proposes to extend to 13 the Stem space of verbs in French, in order to cope with the fact that many derived lexemes must be formed on a base which is not included in the set of the 12 types provided by the original Stem space. This 13th stem type is therefore restricted to derivation.

The communication we propose focuses on the role of stem suppletion in lexeme formation. The phenomena we investigate are conversion from V to N (e.g. marcher>marche ‘to walk’, *a walk’) on a one hand, derivation of Agent Nominals from verbs through -eur suffixation (e.g. voyag-eur, ‘traveller’) on the other hand. These lexeme formation rules are interesting because each of them may choose a different stem from the stem set associated to the verb they apply to. These stem selection processes will be compared with what happens for other lexeme formation rules, since slight differences come out, some of them systematic, others anecdotal. One of the goals of the communication is precisely to give a detailed picture of the situation in order to be able to formulate hypotheses on a sufficiently secure basis. The forms that will be investigated come from a list of lemmas constituting an augmented version of the TLF’s nomenclature. The topics mentioned in the following sections will be addressed in turn.

2. **Extending the Stem space** It will be argued that the stem space of French verbs must be extended one step further and that a 14th stem type has to be postulated in order to account for several lexeme formation processes such as A>V conversion (e.g. rouge>rougir ‘red’, ‘to turn red’), N>V conversion (e.g. fleur>fleurir ‘flower’, ‘to flower’), V>N conversion (e.g. bondir>bon’d ‘to leap’, *(a) leap’) and rules forming verbs from adjectives by prefixation (e.g. pauvre>appauvrir ‘poor’, ‘to impoverish’ riche>enrichir ‘rich’, ‘to en-"

rich’). This stem type will be dubbed ‘zero stem’ for reasons that will be explained in due time. Insofar as postulating this zero stem is essentially motivated by the behaviour of verbs belonging to the so-called 2nd conjugation, it amounts to acknowledge the necessity of stem-based classes in French, instead of the classical Inflectional classes. This issue will be discussed further in the communication.
3. Stem allotting and stem selection Although many deverbal lexeme formation rules take stem 1 (= IMPF or PRS.1[2PL]) as base, VN composition takes stem 3 (= PRS.SG), as shown in table 1. However (Bonami, Boyé, and Kerleroux 2009) have shown that -ion suffixation is based on stem 13. Beside, stem 12 (= PST.PTCP) arguably seems to be the appropriate stem for deriving by conversion a subset of adjectives, and V N conversion requires stem 0, as illustrated in table 2. In the end, stems 0, 1, 3, 12 and 13 are likely to be used as bases for deverbal lexeme formation rules. Given this situation, the following question arises: may any lexeme formation rule choose any of these stem types or are they hard-wired preferences for each rule? The latter option seems to be the right one, since abstract deverbal nouns in -ion are exclusively formed on stem 13, while agent deverbal nouns may select either stem 1 e.g. lav-eur ‘washer’, stem 13 direct-eur ‘director’. The way the relevant stem is chosen by each lexeme formation rule will be discussed and a formalism will be proposed to account for the data. The empirical soundness of triggering features such as ‘Latinate’ (Dell and Selkirk 1978) – or their equivalent – is an issue that will also be addressed.

4. Correlations In a few cases, the choice of a stem seems to be correlated with stable options, which need to be indicated as such in any account of the data. For example, the -eur/-rice alternation (in contradistinction to -eur/-euse suffixation) is correlated with the selection of stem 13, be it selected by a rule forming an adjective or a noun. As for conversion, converted lexemes formed upon stem 13 have a result meaning in their majority, while those built upon stem 12 can never denote an agent. If we turn to derived agent nouns in -eur, we see that whenever doublets exist, the one formed upon stem 13 is pragmatically unmarked e.g. lect-eur ‘reader’, whereas the one formed upon stem 1 tends to be marked, generally as specific to a given domain of activity e.g. lis-eur ‘the one who reads aloud in a community (monks, etc.).’ A tentative account of these data will be proposed.

<table>
<thead>
<tr>
<th>Lexeme</th>
<th>Stem 0</th>
<th>Stem 1</th>
<th>Stem 12</th>
<th>Stem 13</th>
<th>Derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>remplir</td>
<td></td>
<td>remplis</td>
<td>remplis</td>
<td>remplissage</td>
<td>remplis-aż</td>
</tr>
<tr>
<td>ralementhir</td>
<td></td>
<td>ralatis</td>
<td>ralati</td>
<td>ralentissement</td>
<td>ralatis-mâ</td>
</tr>
<tr>
<td>abattre</td>
<td></td>
<td>abat</td>
<td>aba</td>
<td>abattoir</td>
<td>abat-war</td>
</tr>
<tr>
<td>abattre</td>
<td></td>
<td>abat</td>
<td>aba</td>
<td>abat-jour</td>
<td>aba-5ur</td>
</tr>
<tr>
<td>tordre</td>
<td></td>
<td>tord</td>
<td>tord</td>
<td>tord-boyaux</td>
<td>tord-bwajo</td>
</tr>
</tbody>
</table>

Tab. 1: Lexeme formation rules based on stem 1 or 3

<table>
<thead>
<tr>
<th>Lexeme</th>
<th>Stem 0</th>
<th>Stem 1</th>
<th>Stem 12</th>
<th>Stem 13</th>
<th>Derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>construire</td>
<td></td>
<td>köstrui</td>
<td>köstruiz</td>
<td>köstruiz</td>
<td>köstrykt</td>
</tr>
<tr>
<td>découvrir</td>
<td></td>
<td>dekuvu</td>
<td>dekuvu</td>
<td>dekuvu</td>
<td>dekuvu-sit</td>
</tr>
<tr>
<td>bondir</td>
<td></td>
<td>bôd</td>
<td>bôdis</td>
<td>bôdi</td>
<td>bôdit</td>
</tr>
</tbody>
</table>

Tab. 2: Lexeme formation rules based on other stems
5. References


What distinguishes stem allomorphy?
A masked priming study with French stimuli

Hélène Giraudo
CLLE-ERSS, CNRS & Université de Toulouse

Madeleine Voga
Université Paul Valéry – Montpellier III

Since Rumelhart & McClelland (1986) first presented their connectionist model of the English past tense system, the question of the nature of morphological representation has divided psycholinguists. This is a central question in the debates on the nature of cognition, since it concerns the understanding of how the lexicon is organized in terms of structural units, and how these units interact with each other during lexical access. One of the important controversies in this domain is about the description of the core units of the lexicon, namely the morpheme versus lexeme problem. Regarding the later, as Aronoff pointed out (1994), it is better to speak of lexeme-based morphology, because the term “word-based” has led to the misunderstanding that the concrete form of a word might be the basis of morphological operations. However, it is often an abstract stem form of a lexeme, which does never surface as a concrete word form, that constitutes the basis for morphology, and hence, the term “lexeme-based” is more appropriate. This lexeme-based view of morphology is shared by many morphologists (Bybee, 1988; Booij, 2002): morphology is not the “syntax of morphemes” but the extension of patterns of existing systematic form-meaning correspondences between words.

Aronoff (1994) and Stump (2001) have shown that stems in inflectional morphology are morphemic; a morphemic stem is that part of a word to which inflectional material is added and has no meaning of its own. This view is not easily defendable for allomorphic stems, especially so under the light of psycholinguistic data showing that morphological processing is not an “all or none” phenomenon, but that different levels of semantic, orthographic or phonological similarity induce graded effects of morphological facilitation (at least as far as the priming technique is concerned). For example, Rueckl, Mikolinski, Raveh, Miner & Mars, (1997, see also Plaut & Gonnerman, 2000) put forward evidence of graded effects among irregular inflections (e.g., make-made primes better than) take-took). In the same vein, Pastizzo & Feldman find graded effects within a masked priming procedure where morphological effects are estimated relatively to an orthographic control
(ex. hatch – HATCHED > fall – FELL > teach – TAUGHT). Graded effects do not necessarily mean that there is no morphology in the lexicon, nevertheless they are often taken as evidence that there is no explicit morphological level: morphology is thus diluted into formal and semantic properties of lexical units.

The experiments we present here aim to show that, in a view where the distinction between regularity and irregularity has no psychological reality (for Greek: Tsapkini, Jarema & Kehayia, 2000; Voga & Grainger, 2004; for English: Pastizzo & Feldman, 2002) the differences in effect amplitude between allomorphic and non allomorphic inflections should be attributed to the circumstances under which these effects are traditionally measured, rather than to the organisation of the inflectional paradigm, the existence or inexistence of several stems within the same lexeme or the morphomic nature of allomorphic stems, at least as far as masked priming is concerned.

One of the difficulties of the study of morphology for alphabetic languages – in which the vast majority of research is conducted – is that not only morphology is correlated with semantic, orthographic and phonological factors, but also that stems and inflected words exist also as free word-forms, entertaining with each other different relations. Giraudo & Grainger (2001) proposed a supra-lexical approach of morphological processing, in which abstract morphemic representations (in the sense of Aronnoff, 1994) receive activation from whole-word form representations, in such a way that word recognition enables the activation of the morphological level, and not the contrary. The key notion here is lexical competition, central for interactive activation models (e.g. McClelland & Rumelhart, 1981), meaning that the presentation of the stimulus at the entry of the cognitive system (i.e. prime) will produce multiple activations, namely activation of all lexical entries that share formal characteristics with the prime. These multiple representations enter into a competition phase, and identification is achieved when a single word exceeds a given threshold becoming more active than its competitors. The central assumption of this model is that if lexical competition processes affect strongly the identification system, they should also have an impact on morphological effects.

The first experiment replicates the classic regularity effect: regular inflections of French verbs prime their infinitive form targets, whereas allomorphic inflections fail to prime the infinitive forms (ex. pouvons primes pouvoir whereas pu does not). In the second experiment, this classic advantage of non allomorphic over allomorphic inflections is reversed, by modifying the relative frequencies between primes and targets. In fact, in the first experiment, the target is the infinitive form of
the verb, which is the member of the paradigm that has the most elevated residual activation (due to its elevated frequency), and thus, the lowest activation threshold. In the second experiment, targets are no longer the easiest to activate paradigm member, but another inflection, which because of its low frequency, has a higher activation threshold. At the same time, the allomorphic inflections used in the second experiment as primes are rather frequent ones, in any case much more frequent than 2\textsuperscript{nd} plural inflections targets. We obtain thus a reversed “regularity effect” that is a frequent allomorphic prime like \textit{pu} primes its target (\textit{pouvez}) whereas non allomorphic, less frequent inflections do not (\textit{pouvons} fails to prime \textit{pouvez}).

Of course, this kind of results is difficult to insert into a morpheme based approach, even a dual-route one, because there are no reason for the direct (regular) road not to operate when the target is no more the most frequent member. After a quick review of the literature we will show that much of the evidence in favour of the morpheme based account is due to the neglected role of lexical frequency effects in inflectional priming.

Nevertheless, of more interest is what these results imply for a lexeme-based approach. We argue that these results should be viewed as evidence that allomorphic (as well as non allomorphic) inflections are represented within the same lexeme, as word units whose lexical characteristics (like surface frequency) influence processing.

References
Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist


1. Introduction
One of the parameters used to determine morphological naturalness (Mayerthaler 1981, Dressler 1985, Wurzel 1984) is morphotactic transparency. In Natural Morphology, the criterion allows to specify consonantal and vocalic mutations emerging in stems and determine the defaultness of attached affixes. In this paper the function of morphotactic transparency is twofold. Firstly, the parameter describes morphophonological properties of a word form by looking at morphologically-motivated phonological alternations within a stem. Secondly, it describes the morphology of a word by specifying (allo)morphs attached to a stem. As regards naturalness, the most natural forms are those which involve no opacification, i.e. stems that undergo no phonological mutations and affixes that are biunique (allo)morphs attached by default in a given morphological context. On this basis we can distinguish four types of combinations of stems and grammatical morphemes depending on their transparency (T) or opacity (O): T stem + T suffix, T stem + O suffix, O stem + O suffix and O stem + T suffix.

2. Goal
The goal of this study is to examine how the change in the morphological composition of a word yields stem mutation. The study is based on two inflecting languages, English and Polish, out of which the first has minimal inflection, whereas the latter is strongly inflected. The choice of the right-edge is motivated by the fact that both languages rely on suffixation (Dryer 2008), although Polish also uses subtractive morphology in inflection. The right-edge context is here limited to encompass c-stems to which consonantal suffixes or zero morphemes are attached, as it is likely to yield consonantal alternations at morpheme boundaries. The data collected for English encompasses past and present (3rd pers. sg.) verbs, and plural nouns. In Polish, infinitive and past (3rd pers. masc. sg.) verbs exemplify concatenative morphology, whereas imperative verbs and genitive plural nouns are examples of truncation.

3. Data description
The most natural word forms, i.e. transparent stems and suffixes, are represented by only one context in both systems. In some Polish verbs, pasthood (3rd pers. sg. masc.) is expressed by a transparent stem and the biunique morph {-l}, e.g. znalaz+l 'he found', krad+l 'he stole' and upad+l 'he fell'. Similarly in English, T+T inflections are verbal stems to which the past morpheme {-ed} is attached, e.g. walk+ed, replace+d, clothe+d. Both languages have transparent stems followed by opaque suffixes. In Polish, the T+O forms are found in some imperative verbs and genitive plural
nouns, e.g. lącz+∅ 'connect!', troszcz+∅ 'take care!', milcz+∅ 'keep quiet!' and rzeźb+∅ 'sculptures [gen.], walk+∅ 'fights [gen.], malp+∅ 'monkeys [gen.], whereas in English such inflections are typical of present (3rd pers. sg.) verbs and some plural nouns, e.g. plant+s, talk+s, forest+s. An interesting property differentiating the two language systems is the degree to which they tolerate stem mutation. Stem opacification in English involves only consonantal alternations in some plurals, e.g. /f:v/ in wife → wives and /θ:ð/ in youth → youths. Polish allows a wide spectrum of mutations: vocalic, e.g. /ɔ:x/ in nios- → niós+l 'he carried' or próśb- → próśb+∅ 'requests [gen]', /ew:ow/ in płynę→ płynę+ɛ 'swim', consonantal: e.g. /ʃʧɛts/ in wpuszczy→ wpuś+ɛ́ 'let in!', and both, e.g. /ɔ:x/ and /dːɛ/ in zawiod- → zawieść+ɛ́ 'to disappoint'.

4. Conclusions

As regards the influence that the suffixation of biunique markers has on stems, a plausible generalization can be proposed only for English. The attachment of two transparent suffixes, i.e. the infinitival desinence {-s} and the past singular masculine exponent {-l}, results in stem alternations in Polish. The same effect is obtained when {-∅} is attached. In English, past and present formation leaves stem transparency intact, whereas plural allomorphs can opacify stems. This suggests that verbal inflection in English is typically more regular than nominal inflection and results in more transparent outputs. This regularity can be also attributed to the system of rules in Natural Morphology (Dressler 1985). English primarily relies on morphological rules, which govern morphologically-motivated allomorphy. Polish right-edge inflection, in turn, is much richer in morphonological rules, which explains the language's strongly alternating character.

An important by-product of our analysis regards the morphological complexity of the two systems. We assume that O+T inflections are more natural (less marked) than T+O due to the relevance of grammatical information carried by a suffix. Interestingly, English, being morphologically less complex, lacks any O+T inflections but has T+O forms. To the contrary, in Polish we find 2 instantiations of T+O and O+T inflections. From this fact it emerges that morphotactic non-transparency represented by O+T is just enough in a morphologically simpler language. However, Polish, which tolerates opacity and displays a particularly wide range of morphological contexts which allow for alternations within a stem, uses also a variety of more natural inflections (O+T) to neutralize the prevalence of marked T+O (and O+O) forms.
On the intraparadigmatic allomorphy of derivational affixes
(the case of Lithuanian verbs)

Jurgis Pakerys
Department of Baltic Studies, Vilnius University

It has been observed that Latin derivational verb suffixes select particular theme vowels, e.g.,
-ur- selects -i- (ös-ur-ī-re), -it- and -O- select -ā- (vīs-it-ā-re, iact-O-ā-re), -sc-
and -ess- select -ē- (cala-sc-e-re, carp-ess-e-re) (Aronoff 1994: 46). This analysis is perfectly
applicable to some Lithuanian verbal suffixes as well. So, e.g., suffix -in- selects theme
vowel -a- in the present stem and -o- in the preterit stem, cf. prs. 1 pl. maž-in-a-me ‘we
decrease’, pst. 1 pl. maž-in-o-me ‘we decreased’ (but no theme vowel is added in the
infinitive stem: maž-in-ti ‘decrease’ ← maž-as ‘small’). The same theme vowels are selected
by suffixes -e1-, -ē1-, and -o1,1, the only difference being that -e- is inserted to avoid hiatus
(and long -y- is also shortened to -i-), cf. prs. 1 pl. ger-ej-a-me, pst. 1 pl. ger-ej-o-me (inf.
ger-e-ti ‘become better’ ← ger-as ‘good’), prs. 1 pl. dal-ij-a-me, pst. 1 pl. dal-ij-o-me (inf.
dal-y-ti ‘divide’ ← dal-is ‘part’), prs. 1 pl. dovan-oj-a-me, pst. 1 pl. dovan-oj-o-me (inf.
dovan-o-ti ‘make a present’ ← dovan-a ‘present’).

There is another group of verb suffixes in Lithuanian that change their shape
depending on the morphomic stem they form. These suffixes were already briefly discussed
by Andronov (2000: 42-43) from the inflectional point of view. I accept his idea that these
suffixes can be interpreted as having zero allomorphs in some stems, but I would like to make
emphasis on the derivational analysis and the need to make distinction between non-
alternating and alternating derivational affixes.

I will start with the suffixes that have zero exponence in some of the stems. Suffix -j2-
is used to derive iterative verbs. The affix is present in the infinitive stem only and selects
theme vowel -o- in the present stem and theme vowel -e- in the preterit stem, cf. prs. 1 pl.
svaid-0-o-me, pst. 1 pl. svaid-0-e-me (inf. svaid-y-ti ‘throw repeatedly’ ← svis-ti ‘throw’,
root sviéd-, apophony ai ← ið). Suffixes -e2- and -o2- are used to derive non-productive
statives and are only present in the infinitive and the preterit stems. Suffix -e2- selects theme
vowel -t- while suffix -o2- selects theme vowel -o- in the present stem and both suffixes select
theme vowel -o- in the preterit stem, cf. prs. 1 pl. av-0-i-me, pst. 1 pl. av-ej-o-me, inf. av-e-ti
‘wear (shoes)’ (← av-ti-s ‘put on one’s shoes’), prs. 1 pl. klüp-0-o-me, pst. 1 pl.
klüp-oj-o-me, inf. klüp-o-ti ‘be on one’s knees’ (← klau-p-ti-s ‘kneel down’, root apophony ū
← au). There is also a group of derivational suffixes that are quite similar to -e2-, -ē2-, and

1 The index is added to differentiate these suffixes from -ē2-, -y2-, and -o2- discussed below.
-o₂-. They select the same theme vowels, but their allomorphs have one or two consonants present in all stems, cf.prs. 1 pl. bar-st-o-me, pst. 1 pl. bar-st-è-me, inf. bar-sty-ti ‘strew repeatedly’ (← ber-ti ‘strew’, root apophony a ← e), prs. 1 pl. skal-d-o-me, pst. 1 pl. skal-d-è-me, inf. skal-dy-ti ‘chop’ (← skel-ti ‘split’, root apophony a ← e), prs. 1 pl. mer-d-è-me, pst. 1 pl. mer-dèj-o-me, inf. mer-dè-ti ‘be in agony’ (← mir-ti ‘die’, root apophony e ← i), prs. 1 pl. mirk-s-i-me, pst. 1 pl. mirk-sèj-o-me, inf. mirk-sè-ti ‘wink, blink (repeatedly)’ (← merk-ti ‘shut one’s eye, give a wink’, root apophony i ← e), prs. 1 pl. dryb-s-o-me, pst. 1 pl. dryb-soj-o-me, inf. dryb-so-ti ‘lie lazily’ (← drib-ti ‘tumble, fall down’, root apophony y ← i).

I believe that the Lithuanian verbs listed above suggest that there is a need to make a distinction between non-alternating and alternating derivational affixes. In the latter case, the affixes have a two-fold function: on the one hand, they differentiate the derived verbs from the base words, and, on the other hand, they are responsible for the formation of distinct morphemic stems and thus play an important role in the inflectional system. Once the verbs become derivationally non-transparent (or if they are analyzed from purely inflectional point of view), the segmentation of the affixes can change and a new layer of linking elements emerges. Thus, e.g., if merdèti is analyzed as a non-derived item, one could segment a shorter affix (é- and -Ø- instead of -dè- and -d-) and could argue that the preterit stem merdèj-o- has two linking elements: the one that links the personal endings (merdèj-o-me) and the one that links the first linker (of the personal endings) to the (neo-)root (merdèj-o-me). In this case, the infinitive stem has only one linker (merdè-ti), and in the present stem, one linker has zero exponent (merd-Ø-i-me).

References


Stem dependency and irregular verbs in Tamil

Bhavani Saravanan
Stony Brook University

Tamil has a bimoraic minimality requirement. Besides this, verbs are also subject to what I term the regularity criterion that calls for all verbs to be minimally disyllabic. A monosyllabic verb, by definition, does not meet this requirement, and is invariably irregular. There are disyllabic irregular verbs also, but in many of them, the irregularity mainly manifests in the participle / past stems, like in English.

Unlike English, irregularity in monosyllabic verbs is not always seen in any one fixed stem. It has an interesting distribution, and is better explained with stem dependency (Bonami & Boye 2002). Semantically very different stems are built from the same underlying stem (Aronoff 1994). Tense stems (past, present and future) are built upon aspectual stems (participle, infinitive and imperative) in the derivation of the paradigm of a given lexeme; their form (but not function / semantic meaning) is dependent on the form of the parent stems. These stem dependencies hold within each of the three inflectional classes (consonant-final, vowel-final and approximant-final) and are also evident in irregular lexemes, which do not randomly project unpredictable inflectional realizations, but surface with conjugations associated with another inflectional class. For instance, the (partial) paradigms of two irregular homophonous approximant-final stems [vej] are compared to the partial paradigm of a regular approximant-final verb [vəraj] are given below. The first [vej] follows the inflectional pattern of a well-behaved vowel-final verb while the second [vej] projects the paradigm of a regular consonant-final verb. They are both supposed to pattern like [vəraj].

<table>
<thead>
<tr>
<th>IMP</th>
<th>FUTURE</th>
<th>PTCPL</th>
<th>PAST</th>
<th>INF</th>
<th>PRESENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>vej</td>
<td>ve-pp</td>
<td>ve-čč-</td>
<td>ve-čč-</td>
<td>ve-kkə</td>
<td>ve-kkə-r-</td>
</tr>
<tr>
<td>scold</td>
<td>ve-jj-iv-</td>
<td>ve-jj-</td>
<td>ve-jj-</td>
<td>ve-jj-ə</td>
<td>ve-jj-ə-r-</td>
</tr>
<tr>
<td>vəraj</td>
<td>vəraj-iv-</td>
<td>vəra-ja-</td>
<td>vəra-ja-</td>
<td>vəra-ja-ə</td>
<td>vəra-ja-ə-r-</td>
</tr>
</tbody>
</table>

The stem dependency tree, which gives the dependency relationships between the different tense and aspectual stems, is part of the grammar. For regular verbs, only the lexicem stem needs to be stated; all the other cells in the lexicem paradigm are filled in by the tree. For irregular lexicemes like [vej], the lexicem stem and the specific conjugation class that it illegally copies need to be stated. Once the inflectional class is known, the tree predicts the other verbal stems. Thus seemingly random conjugations of irregular verbs are also predictable if we assume that irregular lexemes merely express solidarity with an inflectional class other than the one it is expected to be assigned to.

References
Maximising stems

Andrew Spencer
University of Essex

This talk explores a number of problems which arise in a model which admits the notion ‘(morphemic) stem’. An obvious problem is deciding whether a given recurrent piece of word forms should be treated as part of a (possibly discontinuous) stem or as some kind of exponent of morphosyntactic property sets. For instance, in Romance languages verbs usually have a theme vowel (TV) which indicates inflectional (conjugation) class but which may also indicate mood (e.g. cant-a/com-e ‘sings/eats 3sg PresIndic’ vs. cant-e/com-a ‘3sg PresSubjunctive’). This problem becomes acute in languages with very regular stem formation rules which deploy exactly the same kinds of affixation process as inflection (e.g. typical Indo-European languages). More generally, this relates to a largely neglected problem: how exactly do we arrive at a segmentation into morphs of complex word forms? On any model, including word-based pattern matching, complex words need to be segmented into morphs (segmentation requirement) and the linear order of morphs needs to be represented (morphotactic requirement). However, there appears to be no way to establish a segmentation algorithmically. I discuss several aspects of this problem as they relate to stems.

Having identified recurrent partials as candidates for stems we have to distinguish between distinct stems as opposed to ‘alostems’, that is, morphophonologically distinct forms which function as variants of a single stem for the morphology (this problem is related to, but distinct from, Stump’s Indexing Autonomy Hypothesis, under which distinct morphological, indexed stems may correspond to morphophonologically defined stem types in many:many ways). For instance, is the palatalized root form piš of Russian PISAT ‘write’ a distinct stem in piš-u ‘I write’ or an allostem of the root form pis? How can we tell?

I propose a set of principles to serve as heuristics in analysing complex word forms:

**The Strictly Morphemic Stem Hypothesis (SMSH):** “All stems are morphemic” i.e. no stem serves as the (sole) exponent of a morphosyntactic property.

**The Stem Maximisation Principle:** “If a recurrent word partial fails to realize unambiguously a single morphosyntactic property set assume that that partial is (part of) a stem” This means that if a recurrent partial is associated with cells which have conflicting properties (modulo syncretism) then we treat that form as a morphemic stem.

I then argue for a Stem identification procedure, which I conceive of as a kind of word-based pattern matching:

a. State all systematic syncretisms (whole word and partial) and abstract away from them (i.e. count them as effectively a single cell).

b. **Stem maximisation principle**

c. Treat remaining recurrent partials as (non-morphemic) inflectional exponents

A syncretism on default-based models is an equivalence class defined over nodes in the inheritance hierarchy (as defined by, say, the Paradigm Function). A classical morpheme (one-to-one mapping between form and morphosyntactic property content) comes out as a trivial syncretism defining an equivalence class over a single node. (Proper, non-trivial) syncretisms seem to be more important than classical morphemes in the organization of paradigms, and in general they can be completely independent of form, so where we have conflicting analyses we
should chose the one which maximises syncretism. This principle has important repercussions for the analysis of Spanish conjugation.

The principle ‘Maximise Stems’ means that inflections will tend to be relatively transparent (‘morpheme-like’). There will be a corresponding need to define stem formation rules which integrate with the realizational morphology. I illustrate with an application to Spanish conjugation: by careful regard for exceptionless (part-word) syncretisms and by treating recurrent partials as stems unless they are uncontroverisal inflections we can say that the Imperfect Indicative of Class2/3 verbs in Spanish is defined with reference to the Conditional form: com-tamos ‘we were eating’ is inflected like cantar-tamos ‘we would sing’, with the root com-substituting for the StemINF form cantar-. (This represents a reversal of the historical development, of course). An ingenious alternative analysis of the Future/Conditional inflections in Spanish (Real Academia) segments cantaríamos as canta-ria-mos, where the segment -ría is the Conditional Tense/Mood morpheme. But this analysis loses the syncretism, and yet there is evidence that the syncretism helps define paradigm structure even in languages such as Occitan, where phonological attrition destroys any evidence for a Conditional Mood suffix.