The Ohio State University

Working Papers in Linguistics

No. 58

Edited by
Brian D. Joseph
Julia Porter Papke

The Ohio State University
Department of Linguistics

222 Oxley Hall
1712 Neil Avenue
Columbus, Ohio 43210-1298 USA

Autumn 2008

The Ohio State University
WORKING PAPERS IN LINGUISTICS

Working Papers in Linguistics is an occasional publication of the Department of Linguistics of Ohio State University containing articles by members (students and faculty) of the department. To download back issues of the OSU WPL series, please see our web page at:

http://linguistics.osu.edu/research/publications/workingpapers/

Information Concerning OSDL
OHIO STATE DISSERTATIONS IN LINGUISTICS

Since October of 1994, dissertations that have been written by students in the OSU Linguistics Department since 1992 have been distributed by the graduate student-run organization OSDL. As of September 2006, we no longer provide hard copies of dissertations. Instead, we provide them in downloadable format at the following web page:

http://linguistics.osu.edu/research/publications/dissertations/

If you have any questions about OSDL or any of the dissertations it distributes, please email: osdl@ling.ohio-state.edu

INTRODUCTION

This volume of Ohio State University Working Papers in Linguistics represents a resuscitation of the print version of the series after a multi-year hiatus. A few years ago, we experimented with producing the Working Papers in an electronic-only format (the two volumes produced in this way, OSU WPL 56 and OSU WPL 57, are available at

linguistics.osu.edu/research/publications/workingpapers/56/

and

linguistics.osu.edu/research/publications/workingpapers/57

respectively). Still, some members of the department felt that the move away from the print medium was too precipitate, and in response to increased interest in returning to print production of OSU WPL, we have been working over the past several months to put this issue together.

The papers contained herein date from various points in the past five years, as there were a few false starts on the way to reviving the Working Papers, but they nonetheless give a good representation of the kinds of research that goes on in the department. In an attempt to reflect this diversity onomastically, we have decided to title this volume simply as "Ohio State Working Papers in Linguistics 58", without a (frankly somewhat uninformative) subtitle like "Varia" that was the norm for past issues of this sort for several years. All of the papers emanated from the work of students and faculty in the Department of Linguistics, though some authors have moved on to life after the Ph.D.

We would like to thank our colleague David Odden for his technical assistance and general advice on a number of matters pertaining to the production of this issue.

B.D.J.
J.P.P.

Ohio State University Working Papers in Linguistics

No. 58

Table of Contents

ADVENTURES WITH *CAMILLE*: INVESTIGATING THE ARCHITECTURE OF THE LANGUAGE FACULTY THROUGH COMPUTATIONAL SIMULATION[1]

Peter W. Culicover
The Ohio State University

Andrzej Nowak
University of Warsaw

Wojciech Borkowski
University of Warsaw

Katherine Woznicki
The Ohio State University

A foundational issue in cognitive science is the extent to which the properties of particular mental faculties are the product of general capacities that hold for cognition in general. The debate has been especially lively in the case of language, where the particular properties appear to have no counterpart in other cognitive domains, and are therefore good candidates for being specific to the language faculty. If they are specific to language, the argument goes, it is not necessary to explain them in terms of how cognition works in general; they are presumably simply the product of evolution.[2]

On the other hand some of these properties are so specific and apparently so unrelated to functionality to that it is reasonable to question why evolution would have given rise to them. For example, in standard varieties of English it is not possible to have a gap corresponding to a displaced constituent immediately following the complementizer *that*:

---

[1] This report extends the discussion of the computational simulation of language acquisition reported on in Culicover and Nowak 2003.
[2] For a recent exchange on this general issue, see Hauser et al. 2002 and Pinker and Jackendoff 2005.

Culicover, Nowak, Borkowski & Woznicki. Adventures with CAMiLLe: Investigating the Architecture of the Language Faculty Through Computational Simulation. *OSUWPL* Volume 58, Fall 2008, pp. 1-29.

(1) Who$_i$ did you expect (*that) ___$_i$ would win.
cf. I expected (that) Fred would win.

It is not clear what evolutionary advantage would follow from a constraint that rules out **that* __$_i$, especially given that there are languages and even varieties of English that allow it.

A somewhat less categorical view is that certain properties arise from the interaction between the structures of language and the requirement that they be computed in real time by speakers, hearers and learners. For example, the hearer is faced with the task of determining the meaning of an expression on the basis of its form, and in certain cases, the complexity of the form may pose particular challenges for the computational device that determines the meaning of an expression. This is almost certainly true for well-known cases of multiple center embedding, as in (2), but it may be the basis of an explanation in other cases as well, such as Ross' (1967) island constraints.[3]

(2) The man that the criminal that the cop arrested mugged was my friend.

For the learner, the processing task is similar. On the basis of examples of form/meaning correspondences, the learner must construct general rules that say what the possible structures are, and how they are mapped into meaning. Again, it is not implausible that certain systems of realizing such mappings are more complex than others, and pose difficulty for learners or even render learning impossible.[4]

Finally, we come back to the view that the observed properties of language are not specific to language itself. Depending on the property in question, it is possible to find a range of positions under this general rubric. A representative view is that of Tomasello (2003), who claims that a substantial number of properties that theoretical linguists have posited as universals resident in the language faculty are emergent in the knowledge acquired by the learner.

Much of the debate in the literature has turned on points of logic and rhetoric. In part the likely reason for this is that it is impossible to demonstrate strictly on empirical grounds that a particular property of language is **not** specific to the language faculty. In the absence of a fully worked out alternative explanation, it is as plausible that the impossibility of a gap after the complementizer *that* in (1) is due to a specific property of the language faculty, or to the complexity of processing such structures, or to the difficulty of learning a language that treats such sentences as grammatical.

While some of the properties of language are relatively specific, and turn out not to be found in all languages, such as the one exemplified in (1), others are very general

[3] For considerable discussion of this general idea and some specific proposals, see Hawkins 1994, 2004.
[4] See Wexler and Culicover 1980.

and appear to be universal. For example, all languages appear to have nouns and verbs, all languages appear to distinguish Subject and Object, many languages can highlight constituents of a sentence by locating them in designated positions (usually clause-initial position), and all languages appear to be able to express the same range of communicative functions, such as statements, questions, requests, and so on.[5] For some if not all of these properties it is at least plausible that they are not explicitly represented in the architecture of the language faculty. Rather, they are part of the cognitive/social environment in which humans communicate. Hence they are exemplified in the linguistic experience of the learner and emerge in the learner's grammar in the course of learning.

In order to explore these issues constructively we have been developing a computational simulation of a language learner, called ***CAMiLLe*** (***C***onservative ***A***ttentive ***Mi***nimalist ***L***anguage ***Le***arner). The idea behind this simulation is to endow a learner with strictly general computational capacities for identifying patterns, expose it to data about a natural language, and see what it is able or unable to accomplish.

Assuming that the simulation itself is well-constructed, there are two types of outcomes that are useful, success and failure. If the learner is successful, we have a demonstration that a learner with a particular computational capacity is able to formulate correctly hypotheses about the grammar of the language without the benefit of specific a priori knowledge about the structures. If the learner is not successful, we have reason to believe that some a priori knowledge may be necessary in order for learning to take place.

To make the discussion more concrete, consider the rule of wh-movement, which derives wh-questions in languages like English.

(3) Who$_i$ did you call ___$_i$?
What$_i$ are you talking about ___$_i$?

If the simulation is able to learn such a rule on the basis of exemplars in which it has applied without there being specific knowledge built into the learner that such a rule is possible, this constitutes the basis for an argument that this knowledge does not have to be part of the language faculty. On the other hand, if the simulation is unable to learn wh-movement without knowing that in principle a language may have such a rule, then that consists the basis of an argument that such knowledge must be part of the language faculty.

Of course, in practice matters are typically not as straightforward as this, and the reasons for success or failure may not be of the sort that will allow us to draw firm conclusions about the architecture of the language faculty. Nonetheless, a computational simulation holds out the promise of allowing us to determine, for each putative component of the language faculty, whether it is necessary for the successful acquisition

---

[5] Everett 2005 argues that the Amazon language Pirahã lacks many of the expressive capacities of other languages.

of knowledge of language, or whether it can be dispensed with in favor of general computational principles that are not specific to language.[6]

In this paper we describe the basic architecture and capacities of our simulation, ***CaMiLLe,*** and summarize what it is able to do and what it is not able to do. Because ***CaMiLLe*** is a simulation of a **minimalist** learner, as its name suggests, it has little prior knowledge about the structure of language. On the basis of its successes and failures, we draw some tentative conclusions about the architecture of the language faculty, arguing that it must have some specific knowledge of linguistic structure beyond what we have endowed our computational simulation with, although perhaps not as much as is often claimed in the literature.[7] Moreover, on the basis of the apparent successes of our minimalist learner, we offer a hypothesis about the nature of early language development that is to some extent consonant with those who have argued against a highly structured language faculty.

1. Grammatical preliminaries

We adopt an overall perspective on grammar that addresses not only the very general and universal or quasi-universal phenomena that are found in natural language, but also the idiosyncratic and exceptional (see Culicover 1999; Jackendoff 2002; Culicover and Jackendoff 2005). Our view about the goals of syntactic theory is the following, from Culicover and Jackendoff 2005:

Simple(r) Syntax Hypothesis (SSH):

The most explanatory syntactic theory is one that imputes the minimum syntactic structure necessary to mediate between sound and meaning.

On this view,

- The job of grammar is to describe the sound-meaning correspondences.
- Some of these correspondences are unanalyzable; that is, they are individual words that correspond to primitive concepts.
- Some have linguistic structure but are simple or not entirely transparent on the meaning side (idioms) (i.e. there are no nice structure/meaning matchups).
- Some have structure and are transparent on the meaning side (i.e. there is a compositional semantics that interprets canonical phrase structure).
- Some are a combination of the above ('constructions'), ranging from quasi-idioms, double-objects, movement along a path expressions, 'syntactic nuts' (see

[6] Moreover, a simulation can be very helpful in investigating the behavior of a very complex system. Admittedly, it is sometimes possible to make analytic arguments for the necessity of some mechanisms. But it is well documented that simple rules interacting with each other may result in the emergence of unexpected properties that can be investigated only through computational simulation.

[7] Culicover and Nowak 2003 offers a detailed discussion of ***CAMiLLe***'s design and some preliminary conclusions regarding the architecture of the language faculty based on its performance. Our conclusions here are based on those of Culicover and Nowak 2003 but go beyond them in a number of respects.

above), various operator-trace binding constructions, etc. Each has some degree of predictability and generality, and some idiosyncrasies.

This approach to grammar is a 'constructionalist one', in two senses. On the one hand, it assumes that in some cases the best account of the sound/meaning correspondence is one in which meaning is not determined compositionally by the individual words. On the other hand, it assumes that the grammatical knowledge of a language learner is to some extent constructed on the basis of evidence, and is not predetermined.[8]

The evidence that a more nuanced approach to the sound/meaning correspondence is plausible is the following.

First, many words are unanalyzable correspondences between sound and meaning. Some (e.g. Hale and Keyser 2002) have argued that words with complex meanings are syntactically complex and are the product of derivations involving movement and deletion.[9] However, Culicover and Jackendoff 2005 show that the full range of lexical phenomena requires that the morphological and semantic idiosyncrasies of words be irreducible – they must be stated explicitly and individually in any characterization of grammatical knowledge, and cannot be derived from general principles.

Second, idioms have recognizable syntactic structure but unpredictable meaning, and there are vast numbers of non-idiomatic but nevertheless not strictly transparent expressions in natural languages whose meanings have to be at least in part explicitly associated with them. Some typical examples that suggest the range of possibilities are the following; they can be multiplied almost endlessly.

(4) by and large
lo and behold
beat a dead horse
make amends
cast aspersions on (*at / *to)
a flash in the pan
put up with
have a problem with

---

[8] Here we have in mind a variant of the view expressed by Quartz and Sejnowski 1997 as a 'Constructionist Manifesto': "In contrast to learning as selective induction, the central component of the constructivist model is that it does not involve a search through an a priori defined hypothesis space, and so is not an instance of model-based estimation, or parametric regression. Instead, the constructivist learner builds this hypothesis as a process of activity-dependent construction of the representations that underlie mature skills."

[9] Typical cases are words such as the verb *(to) shelve,* which means 'put on a shelf'. The issue is whether there is a syntactic representation that contains the formatives *put* and *on* that maps into this meaning, or whether the meaning is directly associated with the lexical entry of the verb.

Go Bucks!

Third, there are numerous constructional idioms that have partially transparent interpretations whose meanings are in part associated with the entire structure.

(5) a. *Way*-construction (Jackendoff 1990, Goldberg 1995):
Elmer hobbled/laughed/joked his way to the bank.
(Lit. 'Elmer went/made his way to the bank hobbling/laughing /joking.')

b. Time-*away* construction (Jackendoff 1997):
Hermione slept/drank/sewed/programmed three whole evenings away.
(Lit. 'Hermione spent three whole evenings sleeping/drinking/sewing /programming.')

c. Sound+motion construction (Levin and Rappaport Hovav 1995):
The car whizzed/rumbled/squealed past Harry.
(Lit. 'The car went past Harry, making whizzing/rumbling/squealing noises.')

d. Resultative construction
The chef cooked the pot black.
(Lit. 'The chef made the pot black by cooking in/with it.')

The constructions in (5) share the same basic syntax (not surprisingly, since they are all English VPs); what is idiosyncratic is the way in which their meanings are related to the meanings of the parts and to the structure in which they (the parts) appear.

Finally, there are the general rules of language, such as those expressed by phrase structure rules like VP → V NP, where it may be presumed that there is a corresponding rule of interpretation that composes the interpretation of the head with the interpretation of the argument to form an interpretation of the phrase.

Given this range of sound/meaning correspondences that a learner must acquire, the question naturally arises, How does the learner know where on the spectrum a given correspondence falls? What is it about a particular piece of linguistic experience that tells the learner that it is an idiom, or an expression with some idiosyncratic meaning components, or a general construction, or a fully general rule? In our view, the answer is that there is no way *a priori* for the learner to know where on the spectrum a correspondence really is. The conservative strategy is to start at the word/idiom end, and then move away from the maximally specific as the weight of the evidence warrants generalization.[10]

Our general view can thus be summarized as the following: **Construction of language produces constructions in language.** This means that as knowledge of language is constructed dynamically by a learner, what emerges are constructions that

[10] Tomasello 2003 argues that this is the way that language learners in fact proceed.

may ultimately become 'rules', but only if given enough evidence and a suitable generalization mechanism; otherwise, they remain constructions.

Our simulation of language acquisition thus explores the question of what specific prior knowledge of language the learner **requires** in order to be able to acquire the full range of grammatical phenomena found in a language. Note that we emphasize the word 'requires'. We can, if we choose, build into a learner specific knowledge about some grammatical phenomenon, and tell the learner how to identify whether a given language contains this phenomenon. It does not necessarily follow that a learner will be able to correctly identify that the language in fact contains this phenomenon.[11] But if a learner can perform the identification, this does not mean that the specific knowledge is necessary. Since the crucial question for us is what must be part of the language faculty, the way to approach the question is to begin by assuming that the learner's prior knowledge is not specific and see what kinds of failures, if any, this assumption produces.

2. CAMiLLe

Our computational simulation, ***CAMiLLe***, is conservative, in the sense that it does not generalize beyond what the evidence justifies. Hence it is different from a learner that chooses a grammar from a set of predetermined alternatives on the basis of selected 'triggering' data, as in the Principles and Parameters idealization of language acquisition.[12] Along related lines, ***CAMiLLe*** is attentive, in that every piece of data is relevant to the construction of a grammatical hypothesis, and not just particular triggering data. ***CAMiLLe*** is minimalist, in the sense that it is endowed with the minimal knowledge about linguistic structure that will allow it to form any hypotheses at all, and no more.

CAMiLLe's task is to acquire a set of form/meaning correspondences. The data that ***CAMiLLe*** is exposed to consists entirely of pairs of sentences and conceptual structure representations. We assume that the sentences are strings of words and formatives, and the meanings are expressions in a simple attribute-value language.[13] E.g.,

(6) John touch ~s the cat = `TOUCH($AGENT:MAN,$THEME:ANIMAL,$TIME:NOW)`

Relations that are typically expressed by verbs are represented as constants with an associated argument structure (e.g. `TOUCH`). Arguments are given as thematic roles with

---

[11] See, for example, Wexler and Culicover 1980, Gibson and Wexler 1994, Berwick and Niyogi 1996.

[12] See Fodor 1998.

[13] It is important to note that even the written input to ***CAMiLLe*** is a significant idealization and simplification of what is actually presented to a human learner in a real learning context. One of the most salient differences is that the written input is segmented into words, but it is not in the real input. Among other things, ***CAMiLLe*** does not have to deal with variations among speakers, false starts, and environmental sounds.

their values (like `$AGENT:MAN`). We assume that the meaning that ***CAMiLLe*** is presented with contains only primitives that are cognitively accessible to ***CAMiLLe*** at a given stage of development. For example, *John touches the cat* could have the meaning shown in (6) at an early stage, or even just `ANIMAL` at an even earlier stage. Meanings may become more sophisticated as a consequence of development of cognition and perception. E.g., later, the learner may perceive that there is John, a distinct male person, that there is a particular type of animal (a cat), that both are singular in this context, and that they participate in this relation.

(7) `TOUCH($AGENT:JOHN($TYPE:PERSON, $GENDER:MALE,$NUM:SG),`
`$THEME:CAT($TYPE:ANIMAL,$NUM:SG),$TIME:NOW)`

CAMiLLe tries to figure out how the parts of the string of words corresponds to the parts of the meaning. ***CAMiLLe*** does not know whether each word in the string is independently meaningful, or whether there are parts of the string that are idiomatic, in that the words together correspond to a single unanalyzed meaning. So at the outset, ***CaMiLLe*** hypothesizes all possible correspondences between the string of words and the meaning.

To illustrate, suppose that we expose ***CAMiLLe*** to the pair in (8).

(8) that's a bunny = `BUNNY($TYPE:ANIMAL)`

CAMiLLe will form all of the possible hypotheses to account for the meaning `BUNNY($TYPE:ANIMAL)`. In this case there are six such correspondences.

(9)
1. that's a bunny ⇔ `BUNNY($TYPE:ANIMAL)`
2. that's a ⇔ `BUNNY($TYPE:ANIMAL)`
3. a bunny ⇔ `BUNNY($TYPE:ANIMAL)`
4. bunny ⇔ `BUNNY($TYPE:ANIMAL)`
5. that's ⇔ `BUNNY($TYPE:ANIMAL)`
6. a ⇔ `BUNNY($TYPE:ANIMAL)`

In each case but the first, some of the string is taken to be meaningful and the rest is treated as noise.

Each of these rules has an equal weight (.166) when it is first created. However, there will be other sentences with the word *bunny* and the corresponding meaning `BUNNY($TYPE:ANIMAL)` (such as, *Do you want to pet the bunny?*). The more specific the rule is, the harder it is to support it further because it is more likely to be inconsistent with future experience, unless of course it is an exactly correct hypothesis about an idiom. So rule 1 will be lost unless the learner experiences many instances of *That's a bunny*, and the weight of rule 2 relative to the total number of examples exemplifying `BUNNY($TYPE:ANIMAL)` will decrease over time while the rules 3 and 4 will grow. If there are examples with *the bunny, that bunny* and *this bunny* in the strings that the learner is exposed to, and `BUNNY($TYPE:ANIMAL)` in the meanings, then the strongest correspondence will be expressed by rule 4.

The results of an experiment in which Camille is presented with the sentences in (10) containing the word *bunny* are given in (11).

(10) that's a bunny = BUNNY($TYPE:ANIMAL)
that's a nice bunny = BUNNY($TYPE:ANIMAL)
see the bunny = BUNNY($TYPE:ANIMAL)
that bunny is very soft, yes = BUNNY($TYPE:ANIMAL)
do you want a bunny ? = BUNNY($TYPE:ANIMAL)
show me the bunny = BUNNY($TYPE:ANIMAL)

(11) 1. [15] BUNNY($TYPE:ANIMAL) ⇔ bunny
2. [6] BUNNY($TYPE:ANIMAL) ⇔ a bunny
3. [2] BUNNY($TYPE:ANIMAL) ⇔ that's+1->a

Bunny appears in every sentence, the string *a bunny* appears twice, *that's a* appears twice, and *the bunny* appears twice. The first three of these are hypothesized to possibly correspond to the meaning BUNNY($TYPE:ANIMAL). As the input to the learner becomes more complex and more diverse, many such hypotheses are formed and entertained.

In our implementation of ***CAMiLLe*** it is possible to limit the number of rules that are entertained at any one time. This allows us to filter out highly implausible correspondence rules when there are more plausible alternatives available. This feature of ***CAMiLLe*** may be viewed as a variant of the idea of markedness discussed in Chomsky 1965, whereby less complex rules are favored over more complex rules, other things being equal. In our case, the measure of markedness is simply the weight of the rule that ***CAMiLLe*** acquires through experience with positive exemplars.

If there are two words that appear in identical linguistic expressions and the meanings of these two expressions are identical except for the meanings of the two words, ***CAMiLLe*** will form a category consisting of the two words. Given the input in (12), ***CAMiLLe*** forms the rules in (13).

(12) eat bunny = EAT($THEME:BUNNY)
eat doggie = EAT($THEME:DOG)
eat kitty = EAT($THEME:CAT)
eat bunny = EAT($THEME:BUNNY)
eat doggie = EAT($THEME:DOG)
eat kitty = EAT($THEME:CAT)

(13) **1. [62]** EAT($THEME:[BUNNY; CAT; DOG;])
⇔ eat+1->[bunny; doggie; kitty;]
2. [51] EAT ⇔ eat

These 'single difference' rules can be formed even when there are several differences in a string. For example, if we have *Kitty eat bunny* and *Doggie eat kitty,* then if there are enough examples, *bunny* and *kitty* can be put into the same category on the basis of their co-occurrence with *eat.*[14] It is not surprising to learn that such distributional evidence is neither necessary nor sufficient for accurate category formation; however, the implications of this observation are far from trivial, as we discuss in §4.3.

3. Templates

3.1. A minimum condition for finding minimal structure

As in the case of *eat bunny*, etc., when presented with more complex data, ***CAMiLLe*** is able to separate the constants from the variables. ***CAMiLLe*** was exposed to naturally occurring English spoken to children from the Childes database (MacWhinney 1995). Some examples of the correspondences that ***CAMiLLe*** forms are shown in 14).

(14) a. `GO($AGENT:WE)` ⇔ are<-1->we going+1->to

b. `GO($AGENT:[WE; YOU;]`
⇔ are<-1->[we; you;] are+1->going going+1->to

c. `THINK($EXPERIENCER:I)` ⇔ 1.I 2.think 3.that

d. `BE($THEME:[WHAT; WHO;]` ⇔ 1.[what; who;] 2.is 3.that 4.?

e. `BE($THEME:[HE; HERE; IT; THERE; THIS;]`
⇔ 1.[he; here; it; there; this;] 2.is 3.a

f. `[BABY; BALL; BED; BOOK; BOY; BUNNY; CAR; CHAIR; COOKIE; DUCK; HOUSE; NOSE; ONE; THAT; THIS; TRUCK;]($REF:[$DEF; $INDEF;]`
⇔ [a; that;] [?; baby; ball; bed; book; boy; bunny; car; chair; cookie; duck; house; nose; truck;]

g. `BE($PRED:[BOX; BUNNY; COLOR; DARK; FACE; FUNNY; GOOD; HOUSE; IT; LETTER; ONE; RIGHT; ROOM; TAPERECORDER; THAT; THERE; TOOTH; YOU;]`
⇔ [box; bunny; face; funny; good; house; it; letters; one; right; room; tape; teeth; there; what;]+1->.

The rules in (14a,b) show that ***CAMiLLe*** has extracted the essential correspondences of *we are going to* and *are we going to.* The notation *are<-1->we* means that *are* and *we* appear adjacent to one another in both orders. (14b) shows that *we* and *you* have a similar distribution and so form a small category with respect to these expressions. These are typical examples of what we call **templates**, that is, restricted expressions with variable slots that correspond to particular meanings.

---

[14] While it is possible to get nice results when the data is constructed by hand, as it is here, the kind of distribution illustrated in (12) does not occur in naturally occurring speech to language learners. ***CAMiLLe*** does not take into account similarity of meaning; hence even if ***CAMiLLe*** knows *eat X* is possible when *X* refers to an animal, the fact that *pig* is an animal does not allow it to hypothesize that *eat pig* is possible without actually encountering *eat pig*. This is a matter of implementation, not principle.

Example (14c) shows the emergence of another template, *I think that*. Example (14d) shows a fixed expression, *what/who is that?* and its corresponding meaning. Example (14e) illustrates the template *X is a* where *X* is a pronoun. Example (14f) is the template for *[a, that] X*, where *X* is a noun and the meaning is annotated for definiteness.[15]

This last template illustrates the fact that ***CAMiLLe*** is capable of correlating properties of a concept (e.g. definiteness of a object) with specifiers and modifiers of the head. In order for ***CAMiLLe*** to be able to do this it is critical that there be prior knowledge that such a relation may exist. The relation in question is one in which the specifier/modifer-head relation in the syntax corresponds to an attribute-head relation in the meaning. Our experiments in the early development of ***CAMiLLe*** suggested that without the knowledge that these relations exist and that there are correspondences between them, ***CAMiLLe*** cannot discover them. On this basis, we posit our first assumption about the architecture of the language faculty.

Architectural Feature 1. There are corresponding specifier/modifer-head relations in the syntax and attribute-head relations in the meaning.

Finally, in (14g) ***CAMiLLe*** has hypothesized that a noun in sentence-final position (immediately before '.') is interpreted as predicational. Hence if asked to produce an utterance with the meaning "That's a bunny", ***CAMiLLe*** would simply say "Bunny." (and if possible, point).

In general we find these results to be typical of ***CAMiLLe*'**s behavior. In the face of the very diverse input found in naturally occurring talk to children, ***CAMiLLe*** forms numerous correspondences of this type. If presented with constructed input that more systematically reveals grammatical relationships in a language, ***CAMiLLe*** is capable of extracting more sophisticated templates.

We have presented ***CAMiLLe*** with constructed input for several reasons. First, limited samples of naturally occurring speech to children may not provide sufficient examples for ***CAMiLLe*** to be able to form a reasonable hypothesis. Second, it is technically difficult to provide satisfactory meanings for large amounts of naturally occurring speech to children. Third, some relationships require morphological analyses that are not available in transcripts of naturally occurring speech. Fourth, our implementation of ***CAMiLLe*** does not provide it with the capacity to construct sufficiently general categories that can form the basis of general rules. Since these are all simply a matter of implementation and not of principle, construction of the input data allows us to explore ***CAMiLLe'***s capacities more effectively. It is critical that if

[15] Of interest is the fact that ***CAMiLLe*** includes the question mark ? in the possible strings. We attribute this to the fact that there are many questions of the form *What is that? Is that a bunny?*, etc. in the input. Owing to the way that meanings were assigned to the strings, there is a degree of error in the input that leads ***CAMiLLe*** to formulate correct hypotheses (for ***CaMiLLe*)** that appear to us to be errors.

CAMiLLe fails under these highly controlled circumstances, we are able to draw the firmest conclusions about what such a learner can and cannot do.

In general, we find that ***CAMiLLe*** is capable of learning those relations that are strictly local in the string. For example, an imperative in English typically lacks an overt subject adjacent to the verb, a question with inversion has the first auxiliary verb immediately preceding the subject instead of following it, and so on. Here we consider in somewhat greater detail what ***CAMiLLe*** does when confronted with examples of these constructions.

3.2. Imperatives

Superficially, an imperative sentence in English is of the form

(15) V …
[e.g. pet the doggie!]

Typically, the imperative lacks a subject. The form of the verb is virtually identical to the form that is used in the non-third person singular present tense, with the exception of *be.*

(16) Be quiet!
*Am/are quiet!

It is questionable whether a learner is aware of either characteristic during the earliest stage of learning. Given the overwhelming number of imperatives in speech to children, it would not be surprising if learners hypothesized that the citation form of a verb is its form in the imperative. It is plausible that at some point a learner becomes aware that imperatives differ from their declarative counterparts in that they lack a subject NP in the position where an NP might normally appear. At some later point, the learner would become aware that the form of the verb is the 'bare' form, in contrast with inflected forms in the paradigm.

We simulated the effects of assuming different sequencing of the analysis of the input data to the part of the learner, corresponding to different hypotheses about what the learner is capable of understanding about the structure of the input.

We presented ***CAMiLLe*** with a set of positive declarative and imperative sentences, with enough information so that the program could confidently identify the meanings of the noun phrases referring to actors, the verbs, and negation.

First we gave ***CAMiLLe*** examples of positive imperatives, such as

(17) be quiet! = `$IMP(BE($THEME:YOU,$PRED:QUIET))`

The results are the following rules.

(18) a. `$IMP([BUY; FIX; GIVE; GO; KEEP; LISTEN; MAKE; RECYCLE; SECURE; SELL; SING; SKI; SPY; WATCH;]`

⇔ 1.[buy; fix; give; go; keep; listen; make; recycle; secure; sell; sing; ski; spy; watch;]

b. `[BUY; FIX; GIVE; GO; KEEP; LISTEN; MAKE; RECYCLE; SECURE; SELL; SING; SKI; SPY; WATCH;]($AGENT:YOU)`
⇔ 1.[buy; fix; give; go; keep; listen; make; recycle; secure; sell; sing; ski; spy; watch;]

Representing the verbs list in square brackets as the category V, rule (18a) says that V in first position corresponds to `$IMP(V )`, where `V` is the meaning associated with V. Rule (18b) says that a verb in first position corresponds to a meaning in which the agent is `YOU`. Both of these rules are correct empirical generalizations. Neither requires that there be an empty subject represented in the input to the learner, nor does the learner posit a virtual empty subject as it computes the correspondence. In other words, ***CaMiLLe*** acquires the imperative as a construction.

Next we presented ***CAMiLLe*** with negative imperatives of the form *Don't be.* The correspondence rules are as follows.

(19) a. `[BUY; DESTROY; EXTRACT; FEAR; FIX; GO; INVITE; KILL; MARRY; READ; SELL; SING; SPY;]($AGENT:YOU)`
⇔ 2.[buy; destroy; extract; fear; fix; go; invite; kill; marry; read; sell; sing; spy;] don't don't+1->[buy; destroy; extract; fear; fix; go; invite; kill; marry; read; sell; sing; spy;]

b. `$NEG([BUY; DESTROY; EXTRACT; FEAR; FIX; GO; INVITE; KILL; MARRY; SELL; SING; SMELL; SPY;]`
⇔ 2.[buy; destroy; extract; fear; fix; go; invite; kill; marry; sell; sing; smell; spy;] don't don't+1->[buy; destroy; extract; fear; fix; go; invite; kill; marry; sell; sing; smell; spy;]

c. `$NEG` ⇔ 1.don't

d. `$IMP(*NULL*:$NEG)` ⇔ 1.don't

e. `$IMP` ⇔ 1.don't

What these rules say is that the V in second position may have the interpretation `$AGENT:YOU`, and *don't* immediately preceding second position V has the negative interpretation scoping over the interpretation of the V. Rules c-e express the correspondences of initial *don't* with `$NEG`, `$IMP($NEG`, and `$IMP`.

If we mix the positive and negative imperatives, ***CAMiLLe*** constructs all of the preceding correspondences.

3.3. Inversion

Inversion occurs in English yes-no and wh-questions, and in some less frequent constructions. ***CAMiLLe*** was presented with sets of sentences in which the meaning of the sentence with inversion contained a representation of the fact that it is a question. To simplify ***CAMiLLe***'s processing, simple subjects consisting of one or two words were used.

In the analysis of inversion in contemporary linguistic theory, the observation has been made that the auxiliary verb moves into an empty head position that is also the position occupied by the complementizer *that* in embedded sentences. The structure is that of (20), where the head of IP is Tense. In order for the auxiliary verb to move into initial position, it must therefore first move to Tense. This style of analysis is called 'head-to-head movement'.

(20) $[_{CP}$ C $[_{IP}$ DP Tense $[_{VP}$ V …]]] =>
$[_{CP}$ C $[_{IP}$ DP V+Tense $[_{VP}$ ___ …]]] =>
$[_{CP}$ V+Tense $[_{IP}$ DP ___ $[_{VP}$ …]]]

Our experiments show that ***CAMiLLe*** deals with inversion simply by correlating the initial position of the auxiliary verb with the interrogative interpretation. In the first experiment, we did not provide ***CAMiLLe*** with information about the morphological structure of the verb, and for simplicity used only the verb *is*.

(21) `$YNQ(*NULL*:BE)` ⇔ 1.is

When we introduce *do/does* into the data, ***CAMiLLe*** determines that the auxiliary in first position and the verb in third position correlate with the interrogative interpretation.

(22) `$YNQ(*NULL*:[LIKE; PLAY;])` ⇔ 1.does 3.[like; play;]

In the second experiment, we provided ***CAMiLLe*** with information about the morphology – the tense inflection is represented as a separate element in the string. In the case of inversion, the sequence *V tense* is in sentence-initial position, so V is in first position and tense is in second position. ***CAMiLLe***'s correspondence rules reflect these generalizations.

(23) `$YNQ` ⇔ 2.~tense
`$YNQ(*NULL*:HATE)` ⇔ 2.~tense 4.hate

As long as ***CAMiLLe*** pays attention to position in the string relative to the beginning of the string, templates like those in (23) will be formed. Within linguistic theory, position in the string is not linguistically significant, unless it is first or second position.[16] Certainly mention of fourth position *per se* does not appear to have linguistic relevance.[17] It is also possible to represent the template in (23) equally in terms of relative position, where *~tense* is two to the left of *hate.* But in more complex data sets, generalizations in terms of cardinal position cannot be sustained because there is too

[16] For some recent proposals regarding what constitutes second position in a sentence, see the papers in Halpern and Zwicky 1996.
[17] Any sentence will have a first position (and implicitly, a second position). But not every sentence will have a fourth position.

much variability in the position of the elements with respect to the beginning of the string, and too much intervening material of indeterminate length.

In this case, use of string position allows ***CAMiLLe*** to capture accurate generalizations about subsets of possible strings, but not about the language as a whole. In fact, the standard approach to teaching syntactic theory begins by demonstrating that there are no valid generalizations about language that mention absolute position (except perhaps first and second), because of the fact that there are phrases within the string whose length cannot be bounded. But, we suggest, such templates based on a restricted subset of the language may be correct characterizations of learners' early hypotheses about the language (see Tomasello 2003).

The failure of ***CAMiLLe*** to capture generalizations about complex data in terms of cardinal position comes as no great surprise. However, we do want ***CAMiLLe*** to be able to recognize second position, and also to recognize adjacency (one position to the left or the right). Thus we gave ***CAMiLLe*** the capacity to count, expecting that hypotheses formulated in terms of absolute or relative position based on counting would eventually disappear, and generalizations that do not involve counting would emerge. ***CAMiLLe*** did demonstrate that as long it does not have to deal with variable length phrases, it is capable of formulating relatively narrow but serviceable template correspondences in terms of position.

If the learner also has the capacity to generalize over variable length substrings (that is, phrases), then the templates may contain variables. Consider the string

(24) 1.does 3.[like; play;]

from (22). If position 2 can be an NP of arbitrary complexity, and position 3 is generalized to all verbs, then this template is adequate for a substantial body of the grammatical cases. The cases that are presumably excluded are the NPs that contain relative clauses, since these presuppose not just that the phrasal category NP has been identified, but that VP or S have been identified as possible constituents of larger phrases. If ***CAMiLLe*** could process up to the NP level of structure, then we would be able to claim with some justification that ***CAMiLLe*** is a realistic model of acquisition by an early learner, one very similar to that described by Tomasello (2003). In many respects this 'phrasal' ***CAMiLLe*** would give the impression of knowing certain constructions of English. In actual fact ***CAMiLLe*** would only have acquired templates with NP variables that give the illusion that it has formulated grammatical rules such as inversion. E.g., ***CAMiLLe*** would simply have a template for yes-no questions in which the auxiliary verb is positioned in initial position, to the left of the subject NP.

It can be argued that a complete set of such templates, that is, constructions in the sense discussed earlier, if generalized sufficiently, is sufficient to give the impression that a learner has acquired the grammar of a language. If the templates are sufficiently elaborated, it may be that constructing them is in fact extensionally equivalent to having acquired the grammar, thereby opening the question of whether what is acquired is a grammar in the more traditional sense (see Culicover 1999). But a closer examination of

CAMiLLe’s limitations, to which we turn next, shows that the templates that would be required to demonstrate knowledge of English require capacities that go beyond what ***CAMiLLe*** is presently capable of. Identification of these capacities is critical to our goal of determining what must be in the language faculty.

4. Some limitations and their significance

While there are many things that ***CAMiLLe*** does not do, either for principled or practical reasons, we highlight three here as central to our investigation into what a learner must know in order to acquire knowledge of a natural language:

- ***CAMiLLe*** does not form general phrase structure rules of the type VP → V NP.
- ***CAMiLLe*** does not identify filler-gap relations, e.g. between a fronted wh-phrase and its corresponding gap.
- ***CAMiLLe*** does not form supercategories, grouping all of the nouns into one category, all the verbs into another category, and so on.

4.1. From templates to rules

In our discussion of templates above we contrasted the situation where what is learned is a fixed string of constant forms (corresponding to some meaning), and a string of constants that contains one or more variables. An example of each is given in (25).

(25) a. gimme that
     b. gimme NP

At first glace, it might appear that making the transition from a fixed template to one with variables would be straightforward for a learner like ***CAMiLLe*** . Suppose that instead of *gimme that*, the learner hears a large number of expressions like *gimme the book, gimme a kiss, gimme a ball, gimme a red hat* and so on. If the meaning representation contains simply an element that corresponds to the head noun, then the correspondence rules will identify *gimme a* as a constant string, and the noun will be the variable. On the basis of such very systematic input ***CAMiLLe*** will hypothesize a number of plausible rules, including the template *gimme a N* The following shows the results of an experiment on input of the form *gimme a N.*

(26) `$IMP(*NULL*:GIVE)` ⇔ gimme
     `$IMP(*NULL*:GIVE)` ⇔ gimme+1->a
     `GIVE($AGENT:YOU)` ⇔ gimme
     `GIVE($RECIP:ME)` ⇔ gimme
     `GIVE($AGENT:YOU)` ⇔ gimme+1->a
     `GIVE($RECIP:ME)` ⇔ gimme+1->a
     `GIVE($THEME:[BOOK; KISS; PEAR; PENCIL;])`
     ⇔ gimme+2->[ball; kiss; pear; pencil;] a+1->[ball; kiss; pear; pencil;]

CAMiLLe in fact associates the meaning `GIVE($AGENT:YOU,$RECIP:ME)` with *gimme (a),* and picks out the word to the right of *a* as corresponding to the `$THEME`.

The next natural step would appear to be one in which sequences of the form *a book, a kiss, a pear,* and so on are recognized as units, that is, as phrases. Let us suppose for the sake of illustration that this occurs when the meaning of *a* is known. On the basis of recognizing that *a* contributes to the meaning, the phrase would be parsed into the head and the correspondence between *a* and its meaning checked off. That is, ***CAMiLLe*** would carry out the following reduction.

(27) gimme a ball =
`$IMP(GIVE($AGENT:YOU,$RECIP:ME,$THEME:BALL($REF:$INDEF)))`

⟹

gimme ball =
`$IMP(GIVE($AGENT:YOU,$RECIP:ME,$THEME:BALL`~~`($REF:$INDEF)`~~`))`

Then the more general correspondence would be formed of the form *gimme N.*

We say '***CAMiLLe*** would' because in fact ***CAMiLLe*** does not do this. The reason is instructive. On the basis of sentences of the form

(28) gimme a N

CAMiLLe can identify the overt N as corresponding to the `$THEME` of `GIVE`, as we have seen. And ***CAMiLLe*** is able to form a correspondence rule in which *a N* corresponds to `N ($REF:$INDEF)`. But there is nothing in the input tells ***CAMiLLe,*** first, to treat *a N* as a unit headed by the N, and second, to take this abstract N, call it $N^@$, as corresponding to the `$THEME` of `GIVE`. While N is concretely present in the input, $N^@$ is not. It must be created by ***CAMiLLe***, and then ***CAMiLLe*** must know what to do with it.

CAMiLLe is not helped if we provide it with concrete information about other possible complements of *gimme*, e.g. *gimme that, gimme money*, etc. In the absence of the capacity to posit headed phrases, such input simply makes ***CAMiLLe*** more confused about the combinatorial possibilities for *gimme,* since now it must deal with *gimme N* and *gimme a N..*

Similar problems arise if we ask ***CAMiLLe*** to deal with phrases consisting of more than one specifier/modifier of a head, such as *the angry dog*. Suppose that ***CAMiLLe*** knows that *angry dog* is an instance of an $N^@$. Given this, ***CAMiLLe*** is then faced with the string *the dog*$^@$*,* where the $^@$ here is our notation to indicate that this instance of *dog* is actually not original in the string but is arrived at by parsing *angry dog.* What ***CAMiLLe*** needs to know now is that having parsed *angry dog*, it must now parse *the dog*$^@$. Again, this is not something that is implicit in the computation of string/meaning correlations, and does not suffice even to form templates of the form C_i NP C_j for constant strings C_i and C_j. And it is not something that ***CAMiLLe*** will figure out on its own.

Finally, suppose that we make ***CAMiLLe*** able to deal iteratively with the output of replacing a substring with a constant. In some cases, the result is not a well-formed sentence of the language. For example, if the input is *give a book to Chris*, and we parse *a book* into *book,* the resulting string is ungrammatical.

(29) *give book to Chris

But if ***CAMiLLe*** is able to treat this on a par with original input, ***CAMiLLe*** will acquire incorrect knowledge of language. Therefore ***CAMiLLe*** must be able to distinguish between original and derived strings.

On the other hand, (29) is well-formed if the direct object is a mass noun, like *money.* So while ***CAMiLLe*** should not use (29) as a basis for deciding whether a count noun can appear as the direct object without a specifier, it should use (29) to establish and strengthen the correspondence between *give NP* and `GIVE($THEME:NP)`.

In sum, in order to arrive at the appropriate generalization, ***CAMiLLe*** must be equipped with the following three features.

Architectural Feature 2. If there is a string *M H* and *M* corresponds to a modifier of *H',* then *M H* can be replaced by $H^{@}$ in the string.
Architectural Feature 3. Process derived strings as though they are original strings.
Architectural Feature 4. Derived strings can be the basis for learning syntactic correspondences but not for learning the properties of lexical items.

Or, to put it another way, to go beyond rigid idioms and fixed templates a language learner needs to learn to parse the input. The parser manages the correspondence between sound and meaning at the point at which generalizations begin to emerge, such that some correspondences become nested within other correspondences.

There are many objections that can be raised against this observation, from different quarters. On the one hand, it might be objected that this conclusion, i.e. that ***CAMiLLe*** must be able to do parsing as well as pattern extraction and correlation, is a completely trivial one, since everyone knows that natural languages have this level of structure. Our response is to emphasize that by withholding this capacity from ***CAMiLLe***, we are able to see what ***CAMiLLe*** can do without it. Without the ability to find structure, ***CAMiLLe*** can nevertheless acquire a set of correspondences that gives the appearance of knowing something about a language. We have suggested that this may be what very early learners are doing. Whether this means that they are not actually able to parse input at an early stage, or whether there is not enough evidence to tell them that parsing is necessary, is an open question.

Another objection is that we have not made ***CAMiLLe*** sophisticated enough, in comparison, for example, to machine learning approaches that have demonstrated the possibility of discovering linguistic structure through unsupervised learning. Yuret (1998) reports "I developed an unsupervised language acquisition program that learns to identify linguistic relations in a given sentence. The only linguistically represented linguistic knowledge in the program is lexical attraction. There is no initial grammar or lexicon built in and the only input is raw text."

Yuret's program determines the cooccurrence properties of pairs of words in strings, and on this basis posits structure. Crucially, Yuret assumes that there exist syntactic relations in language, and that syntactic structure is a reflex of these relations. So the goal of his program is to discover the correlations that many be taken as evidence of syntactic relations: "Lexical attraction is the likelihood of a syntactic relation" (22).

His program finds likely dependencies within strings, and ranks them with respect to one another. The ranking in part is determined by the alternative structures that the links give rise to: if two links cross one another, the stronger one wins. Eliminating crossing links gives rise to a clean parse of the sentence where every word is an immediate constituent of a phrase, and the heads are linked to one another. The algorithm assumes right branching, since the language under investigation is English.

These characteristics of Yuret's program are in fact particular realizations of the assumption that the language learner needs to be able to parse the input. Precisely how to do this depends on the properties of the input and other assumptions built into the learner, but the core assumption is that there is a structure with certain properties that needs to be discovered. And there is a general characterization of the properties of this structure that guides the construction of the parse.

In general, we suggest that it is impossible for a learner to get structure of the sort that occurs in natural language out of unstructured input unless the learner is looking for the structure and knows at least a minimum about what its properties are. The question for the theorist, it seems to us, is not whether the learner knows that there is structure, but how much the learner knows *a priori* about the properties of the structure. Our hypothesis about ***CAMiLLe***, which remains to be tested, is that the knowledge of structure embodied in Architectural Features 1-4 is sufficient for the idioms, constructions and phrase structure of a natural language, assuming that the morphology is properly dealt with.

4.2. Unbounded dependencies and gaps

A second area in which ***CAMiLLe*** falls far short of the capacity of a human learner to acquire language involves sentences in which there is a long-distance dependency between two parts of the sentence. Before we discuss these in some detail, we will contrast them with local dependencies.

Not surprisingly, local dependencies are not a problem for ***CAMiLLe***, since in the simplest case they involve adjacency. We gave ***CAMiLLe*** a set of sentences of the form

(30) (N-<1/2/3>-<sg/pl>) <1/2/3>-<sg/pl>-V …

where the <1/2/3> is the person and <sg/pl> the number of the preceding N (which is the subject) that agrees with what is marked on the verb. Whether or not there is a subject, the morphological number corresponds to a number feature in the meaning. The results are along the lines of (31).

(31) SG ⇔ ~sg
PL ⇔ ~pl
[HE; I; YOU_SG;]($NUM:SG) ⇔ [~1; ~2; ~3;]+1->~sg
[THEY; WE; YOU_PL;]($NUM:PL) ⇔ [~1; ~2; ~3;]+1->~pl

We would expect similar results where the adjective and determiner of NP agree with the N in feature like number, gender and case, as long as there is a correspondence with some feature of the meaning.

For unbounded dependencies we have two cases.

Case 1. There is a dependent overt element that agrees in some way with an antecedent. An example is left dislocation in English.

(32) Sally$_i$, I would say that everyone thinks that she$_i$ is a great teacher.

The key property of such a sentence is that the pronoun identifies the grammatical function and thus the semantic role played by the dislocated NP, while the NP provides the identity or 'index' of the individual. The NP is said to be in a non-argument or A′ position.[18] The identity of the individual that has the semantic role identified by the pronoun cannot be determined unless the dislocated NP is linked to the pronoun.[19] The sentence also has special discourse properties (Prince 1987, 1998). We mark the discourse function here as the feature $DISC:TOPIC, as illustrated in (33) for the sentence *Sally, she eats pizza*.

(33) Sally-3-sg, she-3-sg eats pizza =
EAT($AGENT:SALLY($DISC:TOPIC),$THEME:PIZZA)

CAMiLLe's job in this case is to figure out on the basis of the morphological agreement that SALLY is the $AGENT of EAT, and that moreover an NP in this topicalized position has the feature $DISC:TOPIC. Along with the simple example of (33), ***CAMiLLe*** is also presented with sentences in which there is no left dislocation (*Sally eats pizza*) and those in which there is non-adjacent left dislocation. In order to make the task as easy for ***CAMiLLe*** as possible, we used a wide variety of examples, all of which illustrate the point that the topicalized NP, *Sally* or *pizza,* has the topic discourse function.

---

[18] Strictly speaking the NP in left dislocation has no grammatical function at all, while in a wh-question or topicalization it does have a grammatical function, although only in virtue of forming a chain with an empty position in the sentence.
[19] We are adapting here the analysis developed in Culicover and Jackendoff 2005.

Even with this very redundant information, ***CAMiLLe*** is unable to form the generalization that the left dislocated phrase followed somewhere by a pronoun corresponds to the argument in CS indicated by the syntactic function of the pronoun. The closest ***CAMiLLe*** comes is the following.

(34) a. `PIZZA($DISC:TOPIC)` ⇔ pizza -> ~3 -> ~sg ->it
b. `[EAT; WEAR;]($THEME:PIZZA)`
⇔ pizza+4->[eat; wear;] ~3+3->[eat; wear;] ~sg+2->[eat; wear;] ~3+4->it ~sg+3->it [eat; wear;]+1->it

The first rule shows that there were enough examples containing the sequence *pizza-3-sg it* for ***CAMiLLe*** to hypothesize a correspondence between this sequence and the discourse function. The second rule shows that ***CAMiLLe*** was able to see that in the sequence *pizza-3-sg <wear/eat> it* the `$THEME` role is assigned to `PIZZA`.

What ***CAMiLLe*** cannot see is that this possibility for interpreting *pizza* does not depend on fixed length expressions, but holds across arbitrarily long spans of a string. This observation takes us back to the discussion in §4.1. It appears that if ***CAMiLLe*** was able to treat arbitrarily long spans of string as though they were of fixed length, ***CAMiLLe*** would be able to deal with left dislocation. The way to make an arbitrarily long string be of fixed length is to iteratively reduce it to the heads of phrases by parsing out the adjuncts and arguments.

But, crucially, the pronoun that is to be linked to the left dislocated NP cannot simply be parsed like a normal argument, since it will be lost in the intermediate strings and not available at the end of the parse. The presence of the pronoun in the string has to be carried along in the parse. Let us work through a simple example. Suppose that the sentence is *Pizza, I would say that everyone likes it.* Assume for simplicity that V is the head of S. When the pronoun is parsed it is encoded as a feature on the verb, which is passed up through the parse. The sequence of reductions is shown in (35).

(35) Pizza, I would say that everyone likes it
Pizza, I would say that everyone likes-(it)
Pizza, I would say that likes-(it)
Pizza, I would say likes-(it)
Pizza, I would say-(it)
Pizza, I say-(it)
Pizza, say-(it)

This technique is that of passing features through a parse tree proposed originally by Harman (1963), introduced by GPSG (Gazdar et al. 1985) and implemented quite generally in HPSG (Pollard and Sag 1992). As we can see, the pronoun will be either adjacent to the left dislocated constituent, or one element away from it, at some stage of the parse. If the parsing is done in this way by ***CAMiLLe*** , then a rule such as (34), suitably generalized, will suffice.

Let us suppose that this is the correct way to characterize left dislocation. The next question is, How does ***CAMiLLe*** acquire it? Clearly there are several characteristics of this construction that could tell a learner that there is something special going on: the fronted NP is not in a position where it gets assigned a grammatical role, and hence a semantic function, while there is a pronoun in the position that identifies the grammatical role.

However, although it is straightforward for us to characterize what is going on, ***CAMiLLe*** cannot figure out the correspondence without being afforded specific knowledge about how to deal with this type of construction. The example provides evidence that the learner must have (i) the capacity to recognize that an expression lacks a grammatical function and a corresponding thematic interpretation, (ii) the capacity to recognize that this expression must have a grammatical function and a corresponding interpretation, (iii) the capacity to recognize that a proform agrees with such an expression, and (iv) the ability to link the unincorporated expression with the proform. Points (i) and (ii) are related to what has been called the θ-Criterion in GB Theory, which is, informally, that every phrase in a sentence must have a grammatical function and be interpreted; points (iii) and (iv) constitute binding, in an informal sense.

Architectural Feature 5. The θ-Criterion must hold for all expressions of a language.
Architectural Feature 6. Binding may be used to satisfy the θ-Criterion.

As in the case of the features that we have already discussed, these do not come for free but must be built into the learner.

Case 2.

Although left dislocation is a problematic case of unbounded dependency for ***CAMiLLe***, it is by no means the most problematic such case. The most familiar phenomenon of 'unbounded movement', in which the fronted constituent is bound to a gap, is far beyond ***CAMiLLe***'s capacities.

The typical case is that of a simple wh-question, such as (36).

(36) What are you looking at ___ ?

The argument of *looking at* is not in its canonical position, it is in an A′, sentence-initial position. The same considerations that led us to posit that the pronoun in left dislocation is carried through the parse leads to the conclusion that there must be a similar feature that identifies that there is a gap in the parsing of a sentence such as (36). This feature must be attached to the head of which it is an argument, and it must be carried through the parse, so that it can be bound by the fronted wh-phrase. Again, this is the approach to movement originally proposed by Harman (1963).

The problem for ***CAMiLLe*** is to identify the gap and determine that it has to be bound by the moved constituent. As before, some of the cases involve NPs that cannot be assigned a grammatical function; these must be bound to a missing NP position. One way of characterizing this relation is to say that there is a chain consisting of the moved constituent and an invisible placeholder, a trace in GB Theory. Linking of the two elements in principle accounts for the fact that the moved constituent is interpreted as though it is in the position occupied by the trace.

Experiments with the current implementation of ***CAMiLLe*** show, not surprisingly, that ***CAMiLLe*** cannot do this. ***CAMiLLe*** has no idea that there can be gaps in a string that have some syntactic reality. There are a variety of technical means for representing the trace of movement,[20] but they are equivalent in the sense that none of them is something that ***CAMiLLe*** is able to invent simply on the basis of sound/meaning pairs. The notion that there can be a missing argument, and that a dislocated NP can supply this missing argument, is something that has to be built into ***CAMiLLe***.

As is well known, the problem is actually somewhat more complex that the way that we have just characterized it, because of the fact that constituents other than argument NPs can be moved. The following examples illustrate the fact that prepositional phrases, adverbs, and adjectives can be moved.

(37) a. *Under which table* did you find the money __?
b. *How quickly* do you think they will let you know the results __?
c. *How tall* is your child __?

The point for each of these is the same as it is for NPs; these phrases are in a position where their grammatical function cannot be determined. If they can be linked to a suitable empty position in the string, their grammatical function can be determined. So, using the notation of *t* for trace, (37b) must have an analysis equivalent to the following −

(38) *How quickly*$_i$ do you think [they will let you know the results *t*]

− where the trace is in the position occupied by an adverb modifying the verb phrase of the embedded clause. Finding the gap in such cases is non-trivial, because of the fact that the adverb is not selected by the verb and thus the gap cannot be projected locally – it must be projected as a function of the unincorporated fronted adverb.

None of this is news, and it is standard in the analysis of wh-questions in which movement is involved. What is crucial to the present discussion is that the capacity to recognize the fact that the fronted constituent is not incorporated, the capacity to posit the corresponding trace (or equivalent feature) in the relevant position, and the capacity to link the two into a chain are all things that do not appear to emerge naturally from simply finding correlations between patterns of strings and patterns of meaning. It appears that

[20] The standard approaches are movement (which leaves a trace as a copy), binding a empty NP, passing a feature corresponding to a trace up through a phrase marker, and passing a feature corresponding to the selectional requirements of a head up through a phrase marker.

knowledge that movement constructions may occur is something that has to be built into ***CAMiLLe*** and something that ***CAMiLLe*** has to be seeking, in order to be able to find it.[21]

We stress that this is not a matter of a particular implementation of the movement relation, or a particular way of representing the trace. It is more fundamental than that. The following must be assumed to be a part of ***CAMiLLe***'s architecture.

Architectural Feature 7. Constituents may be in an A′ position and form a chain with an agreeing gap.

4.3. Categories

The last limitation of ***CAMiLLe*** that we discuss here is how it forms lexical categories. The questions that we are faced with are whether ***CAMiLLe*** can determine what the lexical categories are, and what the membership of each category is, simply on the basis of the properties of the sound/meaning correspondences that are exemplified for it in the input.

It generally assumed that languages have categories, such as Noun and Verb. These categories transcend semantic categories in that it is impossible to give a semantic criterion that is sufficient to identify a word as a Noun, or a Verb, etc.[22] To take Noun, as an example, some nouns refer to things that can be individuated (*book, unicorn*), some refer to substances (*water, sincerity*), some refer to places or times (*New York, tomorrow*), some refer to properties or dimensions (*sincerity, height*), and so on.

The distributional criteria that ***CAMiLLe*** has available to it are (i) context in the string, (ii) morphological form.[23] Staying with the example of Noun, in a language like English a string context might be that of following a determiner or adjective, as in

(39) <u>the</u> dog
<u>silly</u> kitty
<u>two</u> chipmunks
<u>every</u> student
<u>her</u> sincerity
<u>his</u> persistance

Another would be that the word serves as the subject of a sentence, e.g. –

---

[21] J. Feldman (p.c.) points out that it is not necessary that this particular knowledge be built in explicitly. It is conceivable that some other knowledge can be built in that will permit ***CaMiLLe*** to discover the existence of chains.

[22] This is a long-standing problem in the field. For a recent review of efforts to define syntactic categories in semantic terms, see Baker 2003. See Culicover 1999 and Croft 2001 for arguments that categories cannot be general but must be defined in terms specific to individual constructions.

[23] In practice we reduce (ii) to (i) because it simplifies the implementation.

(40) John shaves.
Mary is singing.
Dogs bark.

or as the object of a sentence –

(41) I admire sincerity.
We like dogs.

Morphological cues to the category Noun, for English at least, would be the plural marking *–s,* which works for most count nouns but not for mass nouns or proper nouns.

While distributional and morphological criteria such as these can be used to define categories, they do not define a single category Noun. In fact, the situation is very much the same as what we find when we considered semantic criteria, only worse. For a given language, like English, the only morphological criterion is singular/plural, and it only serves to distinguish the count nouns. For a language like Russian, there are several classes of nouns that are morphologically distinct; hence the endings that appear on members of one class can be entirely different from the endings that appear on members of another class. For a language like Chinese, on the other hand, there are no morphological criteria that will distinguish nouns from verbs.

These considerations suggest that morphological criteria cannot in general be used to define syntactic categories. This leaves distributional criteria. Because of the count/mass distinction, only count nouns can appear in NPs with count quantifiers, like *every* and *two*, while only mass nouns and plural count nouns can appear with mass quantifiers, like *a lot of*. The indefinite determiner appears only with count nouns. The definite determiner *the* and possessives appear with count nouns and mass nouns, but they do not typically appear with proper nouns.

(42) the dog
her sincerity
*the Mary
*my Fred

It does appear that all nouns may be preceded by an adjective, in English, even proper nouns.

(43) furry dogs
characteristic sincerity
long tall Dexter

However, this raises the question of the distributional criteria that can be used to define Adjective. The property of preceding a noun is obviously not workable, for reasons of circularity. Moreover, non-adjectives can appear in the same position, e.g., in compound noun constructions such as

(44) military training
table manners
computer software

The property of being the complement of *be* is not sufficient, because noun and verbs can also appear in this position.

(45) We are <happy/athletes/sleeping>.

And not all adjectives can appear in this position.

(46) former mayor ~ *be former
perfect idiot ~ *be perfect [in the intended sense]

Summarizing to this point, it appears that at best the phrase-internal distributional criteria can be used to define subcategories, but are not sufficient to define supercategories such as Noun and Verb. The other distributional criteria that we might appeal to are connected to grammatical function. A noun phrase typically can function as the Subject of a sentence or as the complement of a verb or preposition. Assuming that we cannot use distributional criteria to determine that something is a verb or a preposition, we can appeal only to the knowledge that a given phrase has a particular grammatical function, or that it has a particular semantic interpretation.

Distributional and morphological criteria are further suspect because by definition they cannot be cross-linguistic. The intuition that all languages have the category Noun, for example, cannot be sustained by distributional tests because there can be no tests that are valid in more than one language. As much as these tests fail even for a single language, they cannot even be envisioned as applying across languages.

Having ruled out distributional, morphological, and semantic criteria, only grammatical function is left. If, for example, N is the head of a phrase that can be Subject (or Object, etc.), and only N can be the head of such a phrase, then we might have a basis for distinguishing N from other categories. This generalization is not true, however, because phrases with heads other than N can be Subjects.[24]

(47) [For you to do that] would bother me.
[That Sandy is rich] is obvious.
[Visiting relatives] turns out to be unpleasant.

But then the question arises, how can grammatical function be determined independent of the grammar having been acquired in the first place? For example, in

[24] It is possible to stipulate that these phrases have empty N heads. It would be impossible for ***CAMiLLe*** to find this out through inspection of the evidence, but ***CAMiLLe*** could make the inference that they do if confronted with overwhelming independent evidence that only NPs can be Subjects. Or ***CAMiLLe*** could use the evidence to amend the generalization to allow for sentential and VP subjects.

order for ***CAMiLLe*** to know that a word is a noun or that a phrase is an NP in virtue of its grammatical function, ***CAMiLLe*** must be presented with the information that it is the Subject of the sentence. While it can be argued that such grammatical information must play a role in the mapping between sound and meaning (see Culicover and Jackendoff 2005), it has never been demonstrated, or even argued, that learners are presented with this type of information as part of their primary linguistic experience.

The question is, then, Can a learner such as ***CAMiLLe*** determine, on the basis of semantic and distributional information, that a phrase has a particular grammatical function in a sentence? At this point we must leave this question open, since we do not see how to define the basic grammatical functions in semantic terms. Possibly Subject and Object are bootstrapped from core (or default) cases on the basis of meaning, and subsequently become syntactically autonomous. If this was possible, then on the basis of grammatical function ***CAMiLLe*** could ultimately posit broader syntactic categories. And in principle it might be possible to account for the apparent appearance of the same categories across languages. But then we require an additional architectural feature.

Architectural Feature 7.

a. There is an a priori set of grammatical functions.

b. There is a default linking between semantic roles and grammatical functions.

If knowledge of grammatical functions, cannot be extracted from the primary linguistic experience, and we reject the assumption that they are given a priori, the conclusion would be that ***CAMiLLe*** is in principle incapable of constructing broader syntactic categories.[25] This is arguably an empirically correct conclusion, since with suitably well-defined subcategories, the learner's lack of broader syntactic categories will not be seen in its linguistic behavior. It will correctly produce and understand sentences, and it will function as predicted in experiments that ask it to generalize (e.g. "if this is a 'wug', what are two of them called?"). It will only be possible for us to see that the learner deviates from contemporary linguistic theorizing or to normal intuition by examining its internal representations. Whether or not this is an acceptable outcome, it is worth pointing out ***CAMiLLe's*** rather striking limitation of not being able to form broad syntactic categories.

5. Conclusion

[25] Crucially, we are not suggesting that ***CAMiLLe*** is incapable of forming categories, only categories that go beyond the semantic, distributional and grammatical criteria. An approach to tying syntactic categories to semantic criteria would be an acquisitional one that establishes the core syntactic categories on the basis of the restricted semantic space available to the early learner, and then uses distributional criteria to generalize them. (See for example Grimshaw 1981, Macnamara 1982, Pinker 1984, and Anward 2000 for a range of proposals.) ***CAMiLLe***'s problem appears to be that it is limited in how far it is able to generalize; this limitation may well be a principled one.

We have demonstrated through simulation of language acquisition that a language learner must be endowed with certain architectural features that are specific to language. Some of these are standardly assumed to be universal features of natural language within linguistics. While our computational simulation does not constitute proof that the standard linguistic view is correct, it provides additional motivation for it.

Other features are specific to the problem of constructing a computational learner, and are not standardly assumed to be part of the language faculty. However, in our view, it is impossible to envision the language faculty without taking into consideration what kinds of operations it must perform on real data in order to arrive at an adequate representation of the sound/meaning correspondence. It may be, as we have argued, that some of these operations have nothing to do *per se* with the content of grammatical knowledge, e.g. specifics of the structure of phrases or constraints on rules of grammar. Nevertheless they do have to do with the architecture of the language faculty. We have in mind in particular the assumption that the learner is constructing a parser with certain characteristics for the input that it is presented with, on the basis of which the sound/meaning correspondences can be hypothesized and evaluated. In our view, the characteristics of this parser are very much a part of the language faculty. In fact, we would argue that the way in which this parser and the correspondence rules together constitute the learner's grammatical knowledge; see Culicover and Nowak 2002. We envision the parser as an idealization of the device that exists in the mind of the native speaker that performs the sound/meaning mapping in real time. It is an idealization because it is not subject to memory limitations. frequency effects, lexical structure, and similar factors that determine the actual behavior of speakers in producing and comprehending language.

References

Anward, J. (2000). 'A Dynamic Model of Part-of-Speech Differentiation,' in P.M. Vogel, and B. Comrie, (eds.), *Approaches to the typology of word classes.* Berlin: Mouton de Gruyter. 3-46.

Baker, M.C. (2003). *Lexical Categories: Verbs, Nouns and Adjectives.* Cambridge: Cambridge University Press.

Berwick, R., and Niyogi, P. (1996). 'Learning from Triggers,' *Linguistic Inquiry*, 27: 605-622. .

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Culicover, P.W. (1999). *Syntactic Nuts.* Oxford: Oxford University Press.

-----, and Jackendoff, R. (2005). *Simpler Syntax.* Oxford: Oxford University Press.

-----, and Nowak, A. (2002). 'Markedness, Antisymmetry and the Complexity of Constructions,' in P. Pica, and J. Rooryk, (eds.), *Language Variation Yearbook, Vol. 2.* Amsterdam: John Benjamins. 5-30.

-----, and Nowak, A. (2003). *Dynamical Grammar.* Oxford: Oxford University Press.

Everett, D.I. (2005). ' Cultural Constraints on Grammar and Cognition in Piraha: Another Look at the Design Features of Human Language,' *Current Anthropology*, 46. .

Fodor, J.D. (1998). 'Unambiguous Triggers,' *Linguistic Inquiry*, 29 (1): 1-36. .

Gazdar, G., Klein, E., Pullum, G., and Sag, I. (1985). *Generalized Phrase Structure Grammar.* Cambridge: Harvard University Press.

Gibson, E., and Wexler, K. (1994). 'Triggers,' *Linguistic Inquiry*, 25: 407-453. .

Goldberg, A.E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Grimshaw, J. (1981). 'Form, Function, and the Language Acquisition Device,' in C. Baker, and J. McCarthy, (eds.), *The logical problem of language acquisition.* Cambridge, MA: MIT Press.

Hale, K., and Keyser, S.J. (2002). *Prolegomena to a Theory of Argument Structure.* Cambridge, MA: MIT Press.

Halpern, A., and Zwicky, A. (eds.) (1996). *Approaching Second: Second Position Clitics and Related Phenomena.* Stanford, CA: CLSI Publications.

Harman, G.H. (1963). 'Generative Grammars Without Transformational Rules: A Defense of Phrase Structure,' *Language*, 39: 597-616. .

Hauser, M., Chomsky, N., and Fitch, T. (2002). 'The Faculty of Language: What is It, Who Has It, and How Did It Evolve?' *Science*, 298: 1569-1579. .

Jackendoff, R. (1990). *Semantic Structures.* Cambridge, MA.: MIT Press.

-----. (1997). 'Twistin' the Night Away,' *Language*, 73: 534-559. .

-----. (2002). *Foundations of Language.* Oxford: Oxford University Press.

Levin, B., and Rappaport Hovav, M. (1995). *Unaccusativity: At the Syntax-Lexical Semantics Interface.* Cambridge, MA: MIT Press.

Macnamara, J. (1982). *Names for Things: A Study of Human Learning.* Cambridge, MA: MIT Press.

MacWhinney, B. (1995). *The CHILDES Project : Tools for Analyzing Talk.* Hillsdale, N.J.: L. Erlbaum.

Pinker, S. (1984). *Language Learnability and Language Development.* Cambridge, Mass: Harvard University Press.

-----, and Jackendoff, R. (2005). 'The Faculty of Language: What's Special about it?,' *Cognition, 95,* 201-236.

Pollard, C., and Sag, I. (1994). *Head-Driven Phrase Structure Grammar.* Chicago: University of Chicago Press.

-----, and Sag, Ivan. (1992). 'Anaphors in English and the Scope of Binding Theory,' *Linguistic Inquiry*, 23: 261-304. .

Prince, E.F. (1987). 'Topicalization and Left-Dislocation: A Functional Analysis,' in S.J. White, and V. Teller, (eds.), *Discourses in Reading and Linguistics.* New York;: New York Acad. of Sciences. 213-225.

Prince, E. (1998). 'On the Limits of Syntax, with Reference to Left-Dislocation and Topicalization.,' in P.W. Culicover, and L. McNally, (eds.), *The limits of syntax.Syntax and semantics. Vol. 29.* New York: Academic Press. 281-302.

Quartz, S.R., and Sejnowski, T.J. (1997). 'The Neural Basis of Cognitive Development: A Constructionist Manifesto.,' *Behavioural and Brain Sciences*, 20 (4): 537 -596.

Ross, J. R. (1967). 'Constraints on Variables in Syntax,' unpublished doctoral dissertation, MIT, Cambridge, MA.

Tomasello, M. (2003). *Constructing a Language.* Cambridge, MA: Harvard University Press.

Wexler, K., and Culicover, P.W. (1980). *Formal Principles of Language Acquisition.* Cambridge, MA: MIT Press.

Yuret, D. (1998). 'Discovery of Linguistic Relations Using Attraction,' unpublished doctoral dissertation, MIT, Cambridge, MA.

THE VOCALIZATION OF /L/ IN URBAN BLUE COLLAR COLUMBUS, OH AFRICAN AMERICAN VERNACULAR ENGLISH: A QUANTITATIVE SOCIOPHONETIC ANALYSIS[1]

David Durian
The Ohio State University

1. Introduction

The types and extent of regional phonetic and phonological variation in African American Vernacular English (AAVE), have, until quite recently, remained largely unknown and understudied, despite sociolinguists' detailed knowledge of AAVE morphosyntax (Bailey & Thomas 1998). However, with the publication of the papers collected in Thomas & Yeager-Dror (in press), and a series of other recent publications (e.g., Bailey & Thomas, 1998; Thomas, 2001; Wolfram & Thomas, 2002; Fridland & Bartlett, 2006; Labov, et al., 2006), the field has now been given access to a fairly detailed and thorough systematic account of regional variation in the vowel systems of African American (AA) speakers in a variety of locales in the United States. In addition, a good deal of the papers presented in Thomas & Yeager-Dror (to appear) also present much needed research on similarities and differences found between local and regional AAVE varieties and their corresponding European American English (EAE) counterparts.

Although these publications have provided us with a deeper understanding of vocalic variation in regional and local varieties of AAVE, several questions remain open

[1] I wish to thank Donald Winford for providing me with access to the AA data. I also wish to thank Stacey Bailey and Grant McGuire for serving as phonetic judges, as well as Robin Dodsworth for her advice concerning /l/.

David Durian. The Vocalization of /l/ in Urban Blue Collar Columbus, OH African American Vernacular English: A Quantitative Sociophonetic Analysis. *OSUWPL* Volume 58, Fall 2008, pp. 30-51.

for further exploration: 1) beyond vowel system similarities and differences, what are the phonetic similarities and differences among regional AAVE varieties for consonantal variables, such as the vocalization of /l/ and the palatalization of /s/ in consonant clusters such as /_tr/, /_t/, /_p/, and /_k/? 2) In what ways do regional AAVE varieties compare and contrast phonetically with corresponding local and regional varieties of EAE in terms of variation involving consonantal variables? At the time of publication of this paper, systematic explorations of variation involving consonantal variables such as these in AAVE have been rare in the literature, with the notable exception being Labov, et al.'s (1968) study of New York City AAVE, which investigated the vocalization of /l/. Even rarer are systematic comparisons between local EAE and AAVE varieties, with the notable exception being Fix's (2004) unpublished study of /l/ vocalization in a mixed race social network consisting of 6 AA and EA friends living in Columbus, OH.

Given the lack of previous studies, the present study attempts to begin the further discussion of the impact of race on consonantal variation via the impressionistic analysis of the vocalization of /l/ using data obtained from speakers living in Columbus, OH, a metropolis located in the heart of the North American Midland, as it is has been defined on the basis of both lexical and phonological features by Carver (1987) and Labov, et al. (2006). Columbus provides an informative context for exploring contrasts and similarities between AA and European American (EA) vocalized /l/ realization for several reasons. First, as of the 2000 census, Columbus has a population of 1.6 million residents in the Columbus Consolidated Metropolitan Statistical Area. Among the population, roughly 25% are AA and roughly 68% are EA (U.S. Census Bureau, 2000). Second, /l/ vocalization provides a salient variable for exploring comparative patterns of consonantal variation by race because previous studies of Columbus speech have found vocalization to be pervasive in the speech of both blue collar AA (Fix, 2004) and white collar EA (Dodsworth, 2005; Dodsworth, et al., 2006) community members.

However, although Fix's (2004) social network study investigated /l/-vocalization among AA Columbusites and attempted to compare the patterns with EA Columbusites, the study was limited to only 6 blue collar speakers (2 AAs and 4 EAs) living on Columbus's south side, allowing only tentative conclusions to be drawn about the patterns observed. Dodsworth (2005) and Dodsworth, et al. (2006) draw on larger speaker populations and higher token counts, but the foci of their studies was concentrated on either white collar EAs living in suburban and urban areas within the greater Columbus metropolitan area, or white collar EAs living in the Columbus suburb of Worthington, OH. Thus, a detailed comparison of EA and AA patterns of /l/ vocalization in Columbus has yet to be completed.

The following discussion attempts to address this issue by presenting the results of a pilot study investigating the vocalization of /l/ as it occurs in coda, syllabic, word final, and syllable final environments in the speech of urban blue collar AAs, and then comparing these patterns with those found previously among white collar EAs in previous studies. Specifically, the patterns are compared directly with those found among the white collar EAs included in the Dodsworth (2005) study of Worthington, OH, and secondarily with the white collar EAs analyzed in the Dodsworth, et al.(2006) study of urban and suburban Columbus, via the use of a sample population of white collar EAs sharing essentially the same background characteristics. The results of both of these studies are used because they provide detailed enough information on the distribution of vocalized /l/ variants among speakers to allow a comparative statistical analysis of the results to be conducted. In addition, the results are considered qualitatively within the context of Fix's (2004) social network analysis of blue collar Columbus AAs and EAs.

2. Previous studies of /l/ vocalization in North American AAVE and EAE

As studied here, the process of /l/-vocalization is operationally defined as the variable production of "dark" (palatalized) /l/ in syllable or word final contexts. In such instances of "final /l/", the more standard production for this variable, which I call "more articulated" or "nonvocalized" /l/—produced as [ɫ]—involves the tongue-tip making contact against the alveolar ridge, along with the tongue body being raised. The more non-standard form—"vocalized" /l/—involves realizations in which the tongue tip is not touching the roof of the mouth, so that the productions more closely resemble [w] or [ɣ] (Sproat & Fujmura, 1993). In most dialects of English that have been studied, which include Australian English (Horvath & Horvath, 2002), European American varieties of American English spoken in Philadelphia (Ash, 1982), Columbus (Dodsworth, 2005; Dodsworth, et al., 2006), Wisconsin (Carver, 1993), and the Appalachia region (Wolfram & Christian, 1976); and British English (Hardcastle & Berry, 1985; Wells, 1982), this type of vocalized /l/ functions as a back vowel or semi-vowel which may be rounded and/or labialized (Wells, 1982), and may result in a voiced glide (Ash, 1982). Early studies of, and commentaries on, AAVE (*e.g.,* Labov, et al.1968; Labov, 1970, 1972; Fasold & Wolfram, 1970; Wolfram & Fasold, 1974) indicate that /l/-vocalization is also a salient feature of North American AAVE, particularly for Northern speakers of AAVE, closely resembling EAE usage and realization patterns.

However, although it has been widely discussed in the literature as being an AAVE feature, it should be noted that only one extensive phonological study (Labov, et al., 1968) has actually been completed thus far. In this study, Labov, et al. investigated /l/-vocalization among adolescent and preadolescent members of Black and Puerto-Rican peer groups in New York City, with the occurrence of /l/-vocalization described as undergoing similar phonetic conditioning to the occurrence of /r/-vocalization in the speech of these informants. In terms of their description of the distribution of realizations of vocalized /l/ variants, the study was limited to the impressionistic phonetic analysis of vocalization occurring in word-final position, with five variants investigated (clear "l"; dark "l"; unrounded "l"; centralized "l"; and deleted "l").

Among the principal findings of this study were the following:

a) /l/ vocalization is parallel to the vocalization and centralization of /r/ in many ways, but it is less systematic (Labov, et al, 1968:114).
b) vocalization is categorical before liquids and glides (114).
c) the height of the preceding vowel is not a major constraint on the rule (116).
d) the strongest effect is inhibition of vocalization by a following vowel (116).
e) before a following consonant, rounding of the preceding vowel favors deletion; before a following vowel, vocalization is inhibited by rounding of the preceding vowel (116).

f) style has no significant effect on vocalization, but the deletion rule is favored more strongly in group style than single style (119).

Although it has been nearly 40 years since Labov, et al. conducted this study, it remains the most extensive and influential study of /l/-vocalization that has been completed in an AAVE speech community, and it is interestingly that, in the intervening years, most further reports (such as the ones cited above) of /l/-vocalization trends in AAVE speech are actually summaries of the results found in this study.

In regard to the studies of EA patterns of /l/ vocalization in North American referenced above, the most extensive to date from a sociolinguistic vantage point is Ash's (1982) study of Philadelphia. In the study, Ash investigated the occurrence of intervocalic and word-final /l/-vocalization in the speech of 49 speakers in the South Kensington neighborhood of Philadelphia. Among the principal findings of her study were that complex interactions between consonants and pauses occurring as the segment following an underlying /l/ and the frontness/height of vowels occurring as the preceding segment to /l/ are ultimately at the heart of the /l/-vocalization process. Specifically, she found following consonants to have the strongest influence on favoring vocalized variants of /l/ in her data, particularly when the /l/ is also preceded by a low or mid back vowel. When the underlying /l/ is preceded by a high or mid front vowel, then non-vocalization of /l/ appears to be favored.

In addition, Ash also briefly investigated the occurrence of /l/-vocalization in other Midland cities, including Columbus, Pittsburgh, and other smaller cities located throughout Pennsylvania, and compares these results with her Philadelphia findings. The data for this comparison were drawn from an early regional survey conducted by Labov and his associates in 1963-1973 (Labov & Wald, 1969). In general, she found that similar trends typify /l/-vocalization in the White speakers investigated in Columbus and the other Midland area cities, although it should be noted that these findings are based on tokens extracted from the speech of only 16 informants (of which 2 lived in Columbus).

Dodsworth's (2005) study of white collar EA speakers living in the Columbus suburb of Worthington provides more detailed conclusions concerning the phonological and social factors affecting its realization that flesh out the findings of Ash's (1982) within region comparison specifically for Columbus. In her study, Dodsworth conducted a VARBRUL analysis of 724 tokens obtained from 19 speakers, and found preceding and following segment to be significant linguistic factors conditioning the vocalization of /l/ among Worthingtonites. The results of her analysis are contained in table 1.

Specifically, she found that for preceding segments, back and central vowels, as well as labial consonants, most strongly favored vocalization, while front vowels and dorsal consonants most strongly disfavored vocalization. In regard to following segments, she found that dorsal and labial consonants, as well as pause, most strongly favored vocalization, while central vowels most strongly disfavored vocalization. The results generally resonate with those found by Ash (1982) for EAs living in Philadelphia.

Beyond this, Dodsworth (2005) also found that /l/-vocalization may in fact be intertwined with complex social and stylistic factors that ultimately determine how /l/ is realized, either consciously or unconsciously, by EA speakers in the greater Columbus area. Although she found that age and sex were not significant factors impacting /l/-vocalization in Worthington, a factor she did find that was significant was informants' orientation to Worthington. Speakers more strongly oriented to local Worthington identity vocalized /l/ less often than speakers who were less strongly oriented towards Worthington. Dodsworth argues based on these results that lack of vocalized /l/ use is a marker of community identity in Worthington, with the speaker's identity as a Worthington-oriented individual marked specifically by their decreased use of /l/ vocalization.

(total speakers= 21; total tokens=724, input=0.120, p <.005)

Factor (N)	**Factor Weight**	**% Vocalization**
Preceding Segment		
Central Vowel (14)***	*0.741*	*28*
Back Vowel (219)	*0.696*	*25*
Labial Consonant (114)	*0.650*	*21*
Coronal Consonant (152)	0.376	9
Dorsal Consonant (26)	0.339	7
Front Vowel (199)	0.296	6
Following Segment		
Dorsal Consonant (35)	*0.781*	*31*
Labial Consonant (187)	*0.657*	*24*
Front Vowels (28)	*0.583*	*17*
Pause (45)	*0.525*	*16*
Coronal Consonants (99)	0.446	12
Back Vowels (72)	0.396	10
Central Vowels (28)	0.162	3
Location of Residence		
"Columbus Outskirts" (185)	*0.615*	*21*
"Worthington Proper" (429)	*0.506*	*15*
Old Worthington (110)	0.293	7
Sex (Not Significant)		
Female (463)		17
Male (260)		12
Age (Not Significant)		
born c. 1990 (33)		18
born c.1975-1985 (182)		15
born c.1965-1975 (182)		15
born c. 1955-1960 (181)		16
born c. 1935-1945 (75)		14
born c. 1920-1925 (70)		12

*(Key: *=() total number of tokens of this category contained in the data set. The % number in the final column represents the % of tokens out of the () total that were vocalized. ** = Factors significantly favoring vocalization appear in italics, as indicated by any weight greater than .500.)*

Table 1: Factors conditioning /l/ vocalization in Worthington, OH (Adapted from Dodsworth, 2005:241)

At the time of the (2005) study, Dodsworth drew heavily on speakers living in either in the core area of Worthington, or speakers who lived relatively close by to this core area, given that the focus of her study was specifically on Worthington speech rather than more generally on Columbus speech. As a result, less comparative data were available from Columbusites living across the community to more robustly test her hypothesis concerning non-use of vocalized variants as a marker of Worthington identity. However, a later study, Dodsworth, et al. (2006) followed up on the findings of Dodsworth (2005) and explored amore general comparison between the vocalization patterns of both suburban and urban Columbusites. As in the Worthington study, age and sex were found to be non-significant factors, while orientation to suburban or urban identity *was* significant. Speakers who were born and raised in urban Columbus showed a significant lead over suburban born and raised speakers, with the results significant at $p<.01$.

Generally speaking, Dodsworth, et al. (2006) confirms the earlier results of Dodsworth (2005) concerning the importance of urban and non-urban affiliation on impacting the amount of vocalization shown by speakers in Columbus. However, a key element for the present study that is missing from Dodsworth, et al. (2006) is an analysis of the impact of linguistic factors such as preceding and following segment, as well as the word environment in which a vocalized token occurs. This was not completed at the time, as the study was designed to focus only on the social factors of urban affiliation, sex, and age. Given that an analysis which takes these factors into consideration among white collar EA speakers is necessary to the present study, a VARBUL analysis of baseline data drawn from a similar population of white collar Columbus EA speakers is undertaken here, in section 4.1, in order to account for these details and better facilitate a detailed comparison of /l/ vocalization trends among blue collar AAs and white collar EAs.

3. Study population and methodology

Data for African American speakers are drawn from the Sample of African-American Vernacular English in Columbus (SAAVEC), a corpus of nearly 12 hours of tape-recorded conversations among 54 working class African American speakers elicited by field worker Tamara Snow in the summer of 1992 during her study of the social networks of African Americans in Southeastern Columbus. The current study focuses on a subset of 15 of the speakers, as they contributed enough tape-recorded speech to ensure that an adequate number of /l/ tokens could be extracted for analysis. Samples consisted of conversational speech and all fieldworkers and participants were African-American and within-group members of a family and their closest neighbors (see Weldon 1994 and McGuire 2002 for more details on this study). All AA speakers were born between 1942 and 1978. The sample characteristics of the AA speakers are presented in table 2.

Data for the comparative analysis of baseline white collar EA speech data were obtained from 24 speakers, all of whom were born between 1935 and 1982. The data were drawn from speakers recorded for the Buckeye Corpus (Pitt, et al., 2007), a collection of 40 one-on-one sociolinguistic interviews conducted by EA researchers at the Ohio State University in 2000-2001 with long-time Columbus residents. The sample characteristics of the EA speakers are also provided in table 2. The EA speakers were chosen based on the following criteria:

a) Extensive background information on the speaker could be acquired from the Buckeye tape recordings themselves.[2]
b) Speakers have lived in greater Columbus community either their entire lives, or for more than 10 years.
c) Speakers who have not lived their entire lives in Columbus were born and raised somewhere in the greater Central Ohio area before moving to Columbus.
d) Speakers are from parts of Central Ohio located within, Midlands dialect region (as defined on phonological grounds per Labov, et al., 2006).

AA Speaker Data Set			
Birthdate	**Number of Speakers**	**Number of Speakers by Sex**	
		Male	**Female**
Born c. 1969-1978	7	5	2
Born c. 1942-1960	8	2	6
Total Number of Speakers	15	7	8
EA Speaker Data Set			
Birthdate	**Number of Speakers**	**Number of Speakers by Sex**	
		Male	**Female**
Born c. 1965-1978	12	6	6
Born c. 1935-1960	12	6	6
Total Number of Speakers	24	12	12

Table 2: Sample characteristics for the AA and EA speaker sample populations

For all 39 AA and EA speakers analyzed in the present study (15 AAs and 24 EAs), the occupation level of adult informants was also used to ensure talkers were representative of either blue collar AA or white collar EA speech. Information on the sex, birthdate, race, locale in which speakers were raised (if known),[3] and occupation of all 39 speakers was also obtained.

For the study of the 15 AA speakers' patterns of vocalization, 350 tokens from nearly 6 hours of recorded and transcribed audio were extracted and then phonetically analyzed impressionistically by the author and two additional judges. The analysts were: a) either trained in phonetic analysis or had sufficient familiarity with the corpus data to make accurate judgments; b) native speakers of English; and c) currently enrolled as graduate students at the Ohio State University in Columbus, OH.[4] Impressionistic analysis was used because attempts to distinguish light /l/ from fully-realized dark /l/ spectrographically using instrumental analysis in previous studies (e.g., Lehiste, 1964;

[2] At the time of writing, extensive background information for the Buckeye Corpus informants obtained via a survey administered at the time of recording has "gone missing" for the project's archives. Thus, I choose the 24 speakers from those for whom I could obtain all necessary information as a result of listening to their recorded conversations.
[3] With regard to speakers for whom we were unable to sufficiently determine this information, the location in which the informant currently lives is listed instead. These speakers include several from the SAAVEC corpus.
[4] Grant McGuire, Stacey Bailey, and David Durian. All three were non-/l/-vocalizers at the time of the analysis.

Ash, 1982) has proven unsuccesful. For the study of the 24 EA speakers' vocalization patterns, 720 tokens were extracted from nearly 24 hours of recorded and transcribed audio, and then analyzed impressionistically by the author, using the same protocols as with the AA data.[5]

As has been noted previously in the literature (e.g., Labov, et al, 1968; Ash, 1982; Dodsworth, 2005), /l/-vocalization is best analyzed as a gradient process, ranging from the presence of "non-vocalized /l/" to "fully vocalized /l/". Therefore, the three phonetic judges again rated data on a 3-point scale, in an attempt to determine whether the underlying /l/ occurring in the word type tokens analyzed was realized as either being either "/l/-ful" (possessing a clearly articulated /l/ variant), a more fully vocalized variant, or "/l/-less" (possessing little or no perceivable articulation of /l/). Represented numerically, the evaluatory ranking scale used for coding the data was operationalized as follows: 0 = speaker realized a clearly non-vocalized /l/; 1 = speaker realized an "intermediate" or "some" /l/ (their realization was more vocalized than not); and 2 = speaker realized a clearly vocalized /l/.

For the phonetic judgment task, each judge was first trained using a series of prototypes of vocalized and unvocalized /l/ variants in order to establish a sense of the acoustic range of variants they would hear. The listeners then independently heard every token two to three times each and judged whether each one sounded closer to [ɫ] or closer to another unspecified sound (i.e., unvocalized /l/). Once these independent analyses were completed, the judges' scores were tallied using Microsoft Excel. The ultimate assignment of a variant "ranking" was determined by considering the three judges' rankings of a realization, so that the most accurate determination of the realization could be made: in the case of agreement among all three judges, the assignment made by the judges was utilized without issue; in the case of agreement among only two out of three of the judges, the ranking on which the two judges agreed was used; in the case of variant judgments among all three judges, the average of the three rankings was instead utilized. In this way, a "majority opinion" among judges for each token was obtained, which helped to neutralize possible perception errors as well as the influence of any single judge in making the final determination for a given realization.

For the baseline comparative data drawn from 720 tokens spoken by the middle class EA Columbusites, the same judgment and preparation routine was used to impressionistically rate /l/ tokens. However, for this portion of the data set, I was unable to rely on additional phonetic judges to rate the tokens due to time limit restrictions over which the course of which this portion of the study could be conducted; therefore, the results for European Americans presented in section 4 represent my sole judgments. Because this is the case, it is very likely that somewhat different results might have been reported below had I been able to compare my results with those of additional raters. As well, it is quite possible that my perceptual boundaries for determining what sounds like a vocalized or nonvocalized variant of /l/ might be quite different from another persons,

[5] The tally sheets containing the impressionistic judgments of vocalization for the /l/ tokens investigated in this study are not included here to conserve space, but they can be made available to interested parties upon request.

and so additional rater judgments would help counter any bias introduced in the judgments reported due to this type of "single analyst error."

To locate /l/ tokens within the data set, TEXTANT,[6] a frequency-based collocation text analysis program developed by Don Hardy of the University of Nevada, Reno Department of English (Hardy, 2005), was used, so that all tokens within the orthographic transcripts could be located quickly and effectively. In addition, the use of this text-analysis tool provided an efficient way of generating the token type frequency analysis tables used throughout the individual feature analysis portions of this paper. During the token location and extraction process, tokens were selected systematically to minimize lexical effects, following the recommendations of Dodsworth (2005) and Ash (1982). For each speaker, attempts were made to extract 30 tokens per speaker from roughly 30 minutes of recorded conversation per speaker, so that comparable numbers of token per speaker could be located. However, in several cases, speakers did not contribute enough speech in any given conversation to ensure that 30 tokens could be collected. Thus, a variable range of 20 to 30 tokens per speaker was ultimately set to reflect this fact.

The first token of every word type located in the conversation segment containing a suitable /l/ token was extracted, except in a few cases where the sound quality of the tape was poor. In these cases, the next token of that type was extracted. In all cases, the utterance within which the word containing the /l/ token occurred was also extracted so that its occurrence within the utterance could be more accurately observed, and in the case of word-final and syllabic /l/ tokens, the phonetic segment of the following word could be obtained. Generally, only one token of any type was extracted for a single speaker, except in several cases were two tokens of a single word type were extracted so that each speaker contributed 30 tokens to the data set within the allotted 30 minute time period. Following extraction and judgment tasks, the linguistic data were also coded for place of articulation for consonantal segments, as well as height (low/mid/high) and frontness (front/back) for vocalic segments, with the coding applied to both the segment immediately preceding and immediately following /l/ for all tokens analyzed.

4. Data analysis

In the following section, the interaction of /l/-vocalization realization and the individual social factors of sex, age, and race of the informant, as well as the impact of the phonological environment in the conditioning of its realization, is discussed. The linguistic factors used to gauge the impact of phonological environment are preceding segment (the segment occur immediately before /l/), following segment (the segment occurring immediately after /l/), and word environment. Four possible /l/-vocalization word environments are represented: word-final (as in *bell* or *call*), syllable-final (as in *almost* or *also*), as the first segment in a coda consonant cluster (as in *cold* or *silk*), and in instances of syllabic /l/ (as in *little* or *couple*).

---

[6] A general-use version, on which my customized version is based, is available at http://textant.engl.unr.edu/.

For the analysis, the multivariate statistical program VARBRUL was used, so that comparisons between the results presented here and Dodsworth's (2005) study of Worthington could be most easily facilitated. In sum, three independent VARBRUL analyses are presented. The first analysis (section 4.1) is of the 720 tokens drawn from the white collar EA speakers. The second (section 4.2) is of the 350 tokens drawn from the blue collar AA speakers. The third (included in section 5) is of the entire 1070 token corpus drawn from both the AA and EA speaker populations. In total, 15 AA and 24 EA speakers' use of vocalized and non-vocalized /l/ variants are investigated, with the number of tokens investigated for each speaker ranging between 20 and 30 word token types.

In each analysis, the "vocalized /l/" category is used as the application value. As VARBRUL permits only binomial (two-category) rather than trinomial (three-category) distinctions among linguistic variants to be analyzed, tokens coded as "intermediate" and "vocalized" by the phonetic judges were cluster grouped as one category, while "non-vocalized" tokens served as the second category. This grouping choice was made because the results of a cross-tabulation analysis of differences found among AA and EA speakers based on the three-way category split, contained in table 3, reveal that the strongest difference between the groups is to be found not in their use of the fully vocalized variant, but instead, is found in their use of the intermediate variant. For the vocalized variant, speakers differed by only 1% by race, while they differed by 10% in their use of the intermediate variant (with AAs showing more robust use).

Pearson $x^2 = 1.709^{e1}$, df=2, p <.000

Type of Variant	**Race of Informant**		**Total (1070*)**
	AA (350**)	EA (720)	
Non-vocalized /l/	49%*** (170****)	60% (430)	56% (600)
Intermediate /l/	32% (116)	22% (157)	25.5% (273)
Vocalized /l/	19% (64)	18% (133)	18.5% (197)

(Key: *=total number of tokens from both data sets combined; **=total number of tokens within each data set; ***= % realization across all tokens within each data set; ****= number of tokens realized within each data set)

Table 3: Comparison of non-vocalized, intermediate, and vocalized /l/ realizations among the EA and AA speakers

This difference is significant at the p<.000 level, with a x^2 of 1.709^{e1} (and d.f.=2). Given the significance of this difference, the most sensible grouping for a binomial analysis of variance in the use of vocalized and non-vocalized /l/ variants which takes race into account is thus non-vocalized versus intermediate/vocalized rather than non-vocalized/intermediate versus vocalized.

As is discussed below, race is the only statistical significant social factor coded in the analysis conditioning vocalization. Although, as this analysis also reveals, additional data collection is needed before we can say definitively that the other social factors are indeed insignificant (as the analysis here suggests). As the analysis also reveals in regard to linguistic factors, preceding segments also play a strong role in conditioning /l/ vocalization across populations in Columbus generally, regardless of a given speaker's racial background.

4.1. The distribution of variant /l/ realizations in the EA data set

The first VARBRUL analysis conducted was of the baseline data set concerning patterns of vocalization among the 24 white collar EAs. For this analysis, 720 tokens were used. As the results in table 4 reveal (*p<.02*), sex and age are not significant factors conditioning variation in the EA data set, a finding that resonates with Dodsworth's (2005) Worthington study. As in Worthington, EA females in Columbus show a mild lead (5%) in the use of vocalized variants over men, while differences based on the age of the speaker are virtually non-existent.

Turning to the linguistic factors, we see that the only significant factor impacting variation is preceding segment. In an initial VARBRUL run, with preceding and following segments separated by vowel height (low/mid/high), vowel frontness (front/back), and place of articulation coded for consonants (coronal, dorsal, labial), no social or linguistic factors emerged as significant. However, when the vocalic segment groups were recoded into a simple front versus back contrast (and the height contrast.

(total speakers= 24, total tokens=720, input=0.398, p <.02)

Factor (N)	Factor Weight	% Vocalization
Preceding Segment		
Back Vowel (308) ***	*0.578*	*47*
Labial Consonant (47)	0.483	35
Front Vowel (283)	0.445	36
Coronal Consonant (47)	0.381	30
Dorsal Consonant (5)	0.231	20
Following Segment (Not Significant)		
Dorsal Consonant (12)	0.617	50
Labial Consonant (187)	0.578	46
Coronal Consonant (305)	0.502	40
Front Vowels (99)	0.480	40
Back Vowels (72)	0.427	35
Pause (45)	0.299	24
Word Environment (Not Significant)		
Word Final (414)	0.530	43
Syllabic (112)	0.487	33
Coda (112)	0.475	39
Syllable Final (82)	0.401	38
Sex (Not Significant)		
Female (360)	0.534	43
Male (360)	0.466	38
Age (Not Significant)		
Younger (born c. 1973-1985) (360)	0.510	41
Older (born c.1950-1965) (360)	0.490	39

*(Key: *=() total number of tokens of this category contained in the data set. The % number in the final column represents the % of tokens out of the () total that were vocalized. ** = Factors significantly favoring vocalization appear in italics, as indicated by any weight greater than .500.)*

Table 4: Factors conditioning /l/ vocalization in the EA data set

removed[7]), as in Dodsworth's (2005) Worthington study, preceding segment emerged as a significant factor.

As in the Worthington data, back vowels most strongly favor vocalization when they precede /l/. In contrast to Worthington, consonants, as well as front vowels, all have a strong impact on disfavoring vocalization, a pattern that concurs with Ash's (1982) study of King of Prussia. In particular, Ash found mid/high front vowels to have a strong impact on disfavoring vocalization. In terms of following environment, although non-significant, the trends shown by the influence of the segments on favoring or disfavoring vocalization do generally conform to those found by Dodsworth (2005) in the Worthingon data. However, the differences among segment types do not appear to be as robustly differentiated here, which appears to explain why following segment is not a significant factor impacting variation in the present study population.

4.2. The distribution of variant /l/ realizations in the AA data set

The second VARBRUL analysis, of the 350 AA tokens, is presented in table 5. The results are significant at $p<.008$. As these results reveal, as with EA speakers, sex and age are not significant factors conditioning variation in the AA data set. Differences based on the age of the speaker are again virtually non-existent, although among AA speakers, it is males rather than females who show a mild lead (5%) in use of vocalized variants, in contrast to the EA data set.

Turning to the linguistic factors, we see that, in marked contrast to the EA data, all three factors coded are significant: preceding segment, following segment, and word environment. Also in contrast to the EA data, preceding segment emerged as a significant factor with the data coded to account for both height *and* frontness of the vocalic segments, rather than simply vowel frontness (although the segments for following environment required the same regrouping). In regard to word environment, syllable final and coda /l/ show a sharp contrast in favoring vocalization when compared to word final and syllabic /l/, with the syllabic environment most strongly disfavoring vocalization. For preceding segment, back vowels, regardless of height, most strongly favor vocalization when they precede /l/, essentially the same trend as in the EA data (although note that in the AA data high front vowels specifically have the most robust impact). While in contrast to the EA data, only high and mid front vowels, as well as coronal consonants disfavor vocalization. For following segment, dorsal consonants, front vowels, and pause most strongly favor vocalization, also in contrast to the EA data, where we saw dorsal, labial, and coronal consonants functioning as the strongest favoring segment types. However, both groups show a similar trend for following back vowels strongly disfavoring vocalization.

---

[7] The central vowel /^/ is also clustered with the back vowels. For Columbus speakers, this is the most sensible choice, given that the nuclei of /aU/, /oU/, and /u/ are all undergoing fronting in AA and EA speech (see Thomas, [1989]/1993; Durian, et al., in press).

(total speakers= 15, total tokens=350, input=0.520, p <.008)		
Factor (N)	**Factor Weight**	**% Vocalization**
Preceding Segment		
High Back Vowel (20)***	*0.807*	*80*
Low Front Vowel (4)	*0.790*	*75*
Low Back Vowel (62)	*0.759*	*77*
Mid Back Vowel (66)	*0.663*	*70*
Labial Consonant (19)	*0.540*	*47*
Dorsal Consonant (5)	*0.526*	*40*
High Front Vowel (75)	0.412	43
Coronal Consonant (50)	0.261	26
Mid Front Vowel (49)	0.180	22
Following Segment		
Dorsal Consonant (22)	*0.789*	*73*
Front Vowels (27)	*0.685*	*63*
Labial Consonant (62)	*0.570*	*51*
Pause (83)	*0.548*	*51*
Coronal Consonant (122)	0.365	51
Back Vowels (34)	0.362	32
Word Environment		
Syllable Final (16)	*0.848*	*94*
Coda (80)	*0.691*	*61*
Word Final (214)	0.430	47
Syllabic (39)	0.308	36
Sex (Not Significant)		
Female (175)	0.438	49
Male (175)	0.562	54
Age (Not Significant)		
Younger (born c. 1969-1976) (168)	0.456	51
Older (born c.1940-1965) (182)	0.533	52

*(Key: *=() total number of tokens of this category contained in the data set. The % number in the final column represents the % of tokens out of the () total that were vocalized. ** = Factors significantly favoring vocalization appear in italics, as indicated by any weight greater than .500.)*

Table 5: Factors conditioning /l/ vocalization in the AA data set

Interestingly, the finding of pause as a factor strongly favoring vocalization concurs with Ash's (1982) and Dodsworth's (2005) previous findings for EA speakers, although the EA speakers drawn from the Columbus population differ somewhat surprisingly from that pattern, with pause showing the strongest tendency towards disfavoring vocalized variants in the data analyzed here. This difference, as well as others between the groups in terms of the relative influence of other favoring preceding and following segment groups suggest that speech norms among EA and AA speakers in Columbus for /l/ vocalization may not be fully shared. However, given the strong influence of following dorsal consonants and preceding back vowels on conditioning vocalized variant, the differences between groups may not be as strong as they might appear at first blush, and are not so great as to rule out the conclusion that at least some overall similarities in speech norms for vocalization exists between the groups. This possibility is investigated in more detail in section 5.

5. Discussion of the results

Although the AA and EA populations compared in section 4 are quite disparate in terms of socioeconomic class, the comparative analysis reveals some interesting linguistic facts about general speech norms within the greater Columbus community. First, if the AA and EA speakers here are representative of their respective segments of the larger Columbus population, the comparison provides additional evidence confirming Dodsworth's (2005) argument that the white collar EA Worthington informants' decreased use of vocalized variants of /l/ marks their status linguistically as members of a separate social community, since Worthingtonites vocalize /l/ significantly less than other communities in Columbus.

Second, it confirms the findings of Ash's (1982) brief comparative study of Philadelphia speech with other Midland's area cities (including Columbus), that /l/-vocalization appears to be a regionally diagnostic feature of Midland speech based on its

(total speakers= 39, total tokens=1070, input=0.434, p <.009)

Factor (N)	**Factor Weight**	**% Vocalization**
Preceding Segment		
Back Vowel (456)***	*0.627*	*56*
Labial Consonant (96)	0.463	38
Front Vowel (411)	0.419	36
Dorsal Consonant (10)	0.343	30
Coronal Consonant (97)	0.303	28
Following Segment		
Dorsal Consonant (34)	*0.722*	*65*
Labial Consonant (249)	*0.557*	*48*
Front Vowels (126)	*0.531*	*45*
Coronal Consonant (427)	0.480	43
Pause (128)	0.437	41
Back Vowels (106)	0.410	34
Race		
AA (350)	*0.589*	*51*
EA (720)	0.456	40
Word Environment (Not Significant)		
Coda (192)	0.529	48
Word Final (629)	0.498	45
Syllabic (151)	0.491	34
Syllable Final (98)	0.467	47
Sex (Not Significant)		
Female (535)	0.516	45
Male (535)	0.484	43
Age (Not Significant)		
Younger (born c. 1969-1985) (528)	0.507	44
Older (born c.1940-1965) (542)	0.493	43

*(Key: *=() total number of tokens of this category contained in the data set. The % number in the final column represents the % of tokens out of the () total that were vocalized. ** = Factors significantly favoring vocalization appear in italics, as indicated by any weight greater than .500.)*

Table 6: Factors conditioning /l/ vocalization in the combined AA and EA data set

occurrence in a number of major Midland cities. This can be seen by the fact that all three groups of speakers referenced in this paper variably realize vocalized and nonvocalized variants of /l/ to at least some extent. Furthermore, it suggests that speech norms regarding phonetic conditioning of vocalized variants is generally the same among speaker groups, regardless of race, although as mentioned in section 4.2, at first glance, the results suggest some subtle differences exist between AAs and EAs.

However, as we will now see, when the data is examined more closely, these differences are not actually as strong as they initially appear. This is revealed a the side-by-side comparison of the % realization of vocalized tokens uttered by AA and EA speakers when their co-variance with both the preceding and following phonetic segment is also considered. For these comparisons, the simpler front/back vowel contrast (without further disambiguation of the segment categories by height) is used, as this was the only grouping found to be statistically significant for preceding environment in the VARBRUL analysis of the EA data set in section 4.1. The same grouping contrast is used in following environment comparison to maintain parity between preceding and following environment segment type groupings. It should be noted that the preceding and following segment groupings contained in the analysis below emerged as statistically significant factors (*p<.009*), along with race, conditioning variation in the data when the EA and AA data sets were combined, and all 1070 tokens analyzed, in a third independent VARBRUL analysis (see table 6 for results).

Figure 1 plots a side-by-side comparison of % vocalization exhibited by AA and EA speakers as it co-varies with the type of phonetic segment that precedes an underlying /l/. Here we see that, as previously discussed, back vowels show the strongest influence on conditioning vocalized /l/ realizations among both AA and EA speakers, with 74% of the vocalized tokens occurring after back vowels among AA speakers and 47% among EA speakers. As also discussed, coronal consonants show a strong influence on

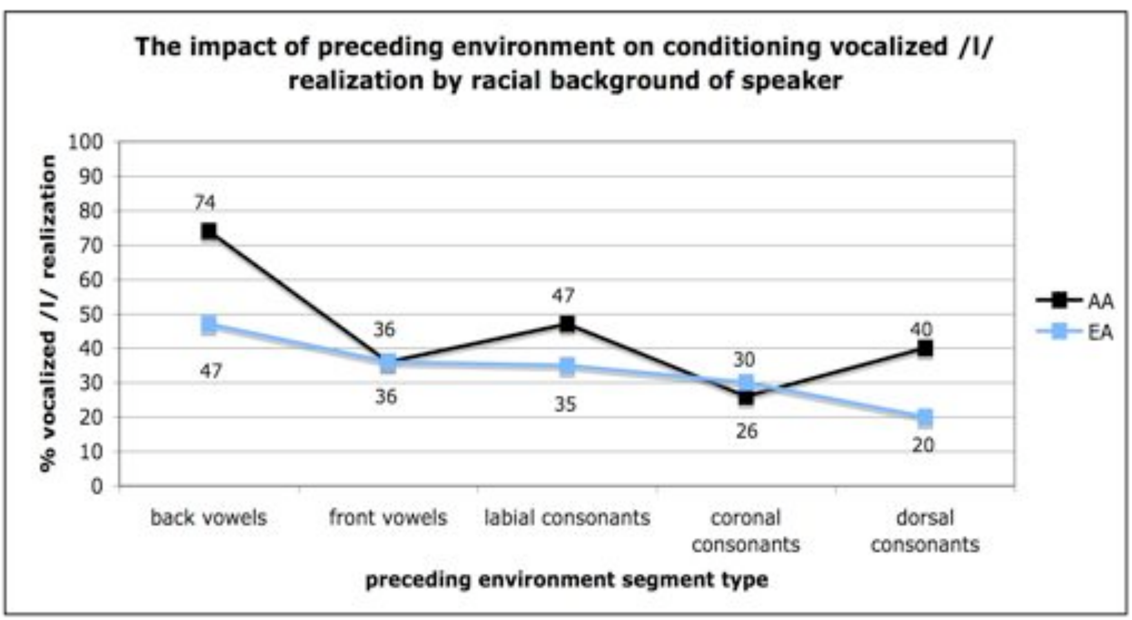


Figure 1: A side-by-side comparison of the impact of preceding environment in the AA and EA data sets

disfavoring vocalization, with only 30% of the vocalized tokens occurring after coronal consonants among AA speakers and 26% among EA speakers. However, AAs and EAs differ in that coronal consonants show a stronger impact in the speech of the AA speakers than the EA speakers, as the strongest segment found to impact EA speech is actually dorsal consonants.

Other differences between the populations include fairly weak differentiation in the impact of front vowels, labial consonants, and coronal consonants on conditioning vocalization among EA speakers, with only a 6% overall difference found between front vowels and coronal consonants, and only a 5% difference between labial and coronal consonants. In contrast, among AA speakers, weak differentiation is found instead between labial consonants and dorsal consonants, as well as between dorsal consonants and front vowels, although the over all difference of 11% between labial consonants and front vowels makes for a strong difference between those two groups, as revealed by the results of the VARBRUL analysis in section 4.2, with labial consonants slightly favoring vocalization and front vowels generally disfavoring vocalization.

The differences between the speaker groups in figure 1 suggest that some differences may exist regarding phonetic conditioning and speech norms among AAs and EAs for /l/ vocalization. However, it should be noted that the higher frequency of vocalized variants among EAs when a front vowel precedes the /l/ may in fact be the byproduct of the stronger ratio of tokens taken from words containing word final and syllabic /l/ tokens than in the AA data set, which has led to the possible skewing of the results for front vowels. This seems possible considering that front vowels conditioned significantly fewer vocalized variants in both the AA data and in Dodsworth's (2005) white collar EA Worthington data in section 2 than among the EA speakers analyzed here.

Also, considering the general cline of variance shown by labial, dorsal, and coronal consonants among both AA and EA speakers, and the similar cline shown by the Worthington data, it also seems plausible that a hierarchy of disfavoring vocalization may exist among these segment types when consonants precede /l/. Given the differences in overall frequency of vocalization among the Columbus EAs, the Worthington EAs, and the Columbus AAs, the data suggests dorsals typically lead over coronals, and coronals typically lead over labials, in disfavoring vocalization. However, once the overall vocalization shown by speakers reaches beyond 50%, as in the difference between the Columbus AAs and Columbus EAs, coronals appear to over take dorsals in showing a stronger disfavoring effect, hence the difference between the groups if one views the % realization differences between the speaker groups as indicating a kind of rank ordering of the environments involved in disfavoring tendencies towards vocalization.

This change in ordering makes sense if one considers that coronals differ from dorsal consonants and the vocalized variants [w] or [ɣ] not only in terms of articulation involving the use (in the case of coronals) or non-use (in the case of dorsals and the vocalized variants) of the tongue blade, but also phonologically in terms of the feature [+front] versus [+back]. In English, coronal consonants differ phonologically from the vocalized variants [w] or [ɣ], in that the coronal consonants as a group tend to be [+front], whereas the vocalized variants and dorsal consonants are [+back]. In the case of less pervasive vocalization trends among speakers—for instance, the white collar EA Worthingtonites—the difference in the feature [+front] versus [+back] may not function as a strong enough contrast on conditioning vocalization of the following /l/ by consonantal segments. However, once frequency of vocalization increases to the levels found among the Columbus AA speakers, the contrast may become significant enough that the impact of the segment groups is modified, perhaps as a result of analogy, modeled on the contrastive relationship of preceding back vowels to front vowels in their impact on conditioning vocalized /l/ variants, or by analogy modeled on a similar consonantal contrast of [+front] to [+back] consonants in the following segment environment.

As shown in figure 2, a similar kind of hierarchical relationship based on the [+front] versus [+back] contrast appears to exist for consonantal segments when they follow a vocalized /l/. Although in this case, the hierarchy of influence is the direction of favoring vocalization, with dorsals leading labials, and labials leading coronals. Unlike preceding segments, here, the ordering of the relationship is the same for both AA and EA speakers, although AA speakers show higher % realization numbers as vocalization is more pervasive in the speech of AAs. AAs show 23% more vocalization than EAs when a dorsal consonant follows, 6 % more when a labial follows, and 11% more when a coronal consonant follows. Phonologically, a hierarchical relationship is plausible if one again considers that dorsal consonants share with the vocalized variants [w] or [ɤ] the feature [+back], whereas coronal consonants are maximally contrastive with the vocalized variants and dorsal consonants on the feature [+front] in English.

One additional difference between the groups in figure 2 that bears noting and discussion is the difference in impact of a following pause on conditioning variation among the populations. Here, we see a 27% difference between the groups, with AAs vocalizing /l/ 63% of the time when a pause follows, in contrast to EAs, who vocalized /l/ only 24%. As with differences for the impact of preceding front vowels discussed above, we argue that this difference may also be the byproduct of skewing, based on the

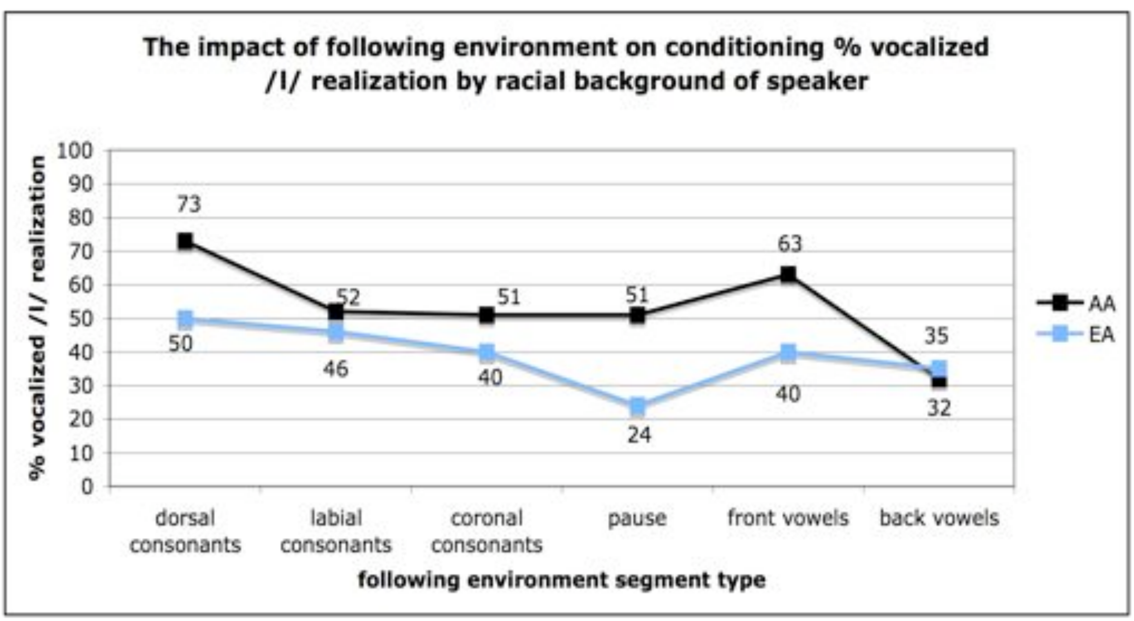


Figure 2: A side by side comparison of the impact of following environment in the AA and EA data sets

difference in the ratio of pre-pause tokens analyzed in the AA set versus the EA data set. A greater number of the EA tokens were drawn from continuous speech, from words with word final /l/ that were immediately followed by another word than in the AA corpus. Therefore, pre-pause tokens are likely underrepresented in the EA data, and given the evidence suggested by the Worthington EA data, it is quite likely a higher percentage of vocalized tokens would be found to occur before pause in the Columbus EA data if more tokens had been used in the analysis, as pause had a factor weight of .524 in the Worthington EA VARBRUL analysis, whereas it had only a weight of only .299 in the Columbus EA analysis.

Based on the combined evidence contained in figures 1 and 2 for explaining why the differences between AA and EA speakers are less than they may appear on the surface, we therefore suggest that speech norms for phonetic conditioning are in fact quite similar for AAs and EAs in Columbus, with the greatest difference between the groups being frequency of realization rather than difference due to phonetic or phonological conditioning differences at the segmental level.

However, one linguistic difference that does appear to be quite robust between blue collar AAs and white collar EAs in Columbus is the impact of word type on conditioning vocalized variants. Although the difference between AAs and EAs for word type was non significant in the VARBRUL analysis, a cross-tabulation of the combined 1070 token data set reveals that this difference is significant at $p < .05$ ($x^2 = 8.008$; d.f.=3) when the covariance of this factor and race are considered. As shown in figure 3, AAs show a significant lead in the realization of vocalized variants in syllable final and coda position. For syllable final position, AAs realize vocalized variants of /l/ 93%, whereas the EAs vocalized only 38%. For the coda position, AAs vocalized 61% of the /l/ tokens,

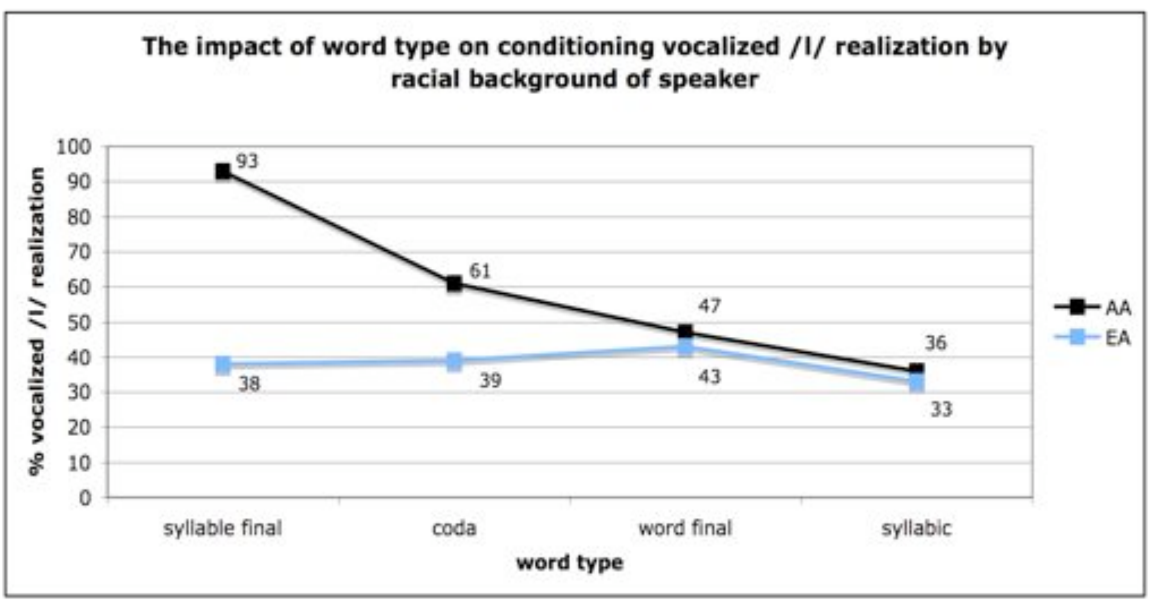


Figure 3: A side-by-side comparison of the impact of following environment in the AA and EA data sets

while EAs vocalized only 39%. Also interesting to note is that the differences between AAs and EAs are essentially non-existent in word final and syllabic position, with AAs leading EAs by only 4% word finally and only 3% in syllabic /l/ position.

Given the robust differences for syllable-final and coda-final position, we argue that these environments may somehow be marked as salient for marking the racial identity of individual speakers in Columbus. In Columbus AAVE, speakers tend to realize word-final consonant clusters as reduced consonant clusters, a process that has been found to operate in a number of blue collar AA speaker communities throughout the United States in previous studies (e.g., Labov, et al, 1968; Wolfram, 1969; Bailey & Thomas, 1998). We suggest that because this is the case, EA speakers may be aware of this tendency in AA speech, and view it as a stereotypical and stigmatized feature of blue collar Columbus AAVE. Furthermore, we suggest that they may view coda /l/ vocalization as being somehow similar, given that the end result perceptually of /l/ vocalization in coda position could sound to EA speakers like a reduced consonant cluster. Thus, EAs, particularly white collar EAs, may in fact be aware enough of vocalization in the coda /l/ environment as to avoid vocalizing it, so as to avoid use of what they perceive as a stigmatized feature of AA speech.

That is, unless an EA speaker is looking to signal more solidarity with AA speakers. In Fix's (2004) study of the social network of a group of six blue collar AA and EA speakers, 2 of whom were AA, 2 of whom were EA but with racially-integrated networks (i.e., those with AA friends), and 2 of whom were EAs without racially integrated networks, she found strong differences between the racially integrated EAs and the non racially integrated EAs in regard to the percentage of vocalization shown by the EAs for tokens occurring in these positions, particularly the coda environment. In her results, the racially integrated EAs vocalized coda /l/ 69%, while the non-integrated EAs vocalized it only 44%. Fix posited that the increased use of vocalization in the coda environment among the racially-integrated EAs is a result of influence from the AA peers, and that their increased use may also be a sign that they are trying to signal increased solidarity with the AA speakers through its use. The data presented here supports Fix's argument, as the percentage of realization shown by the integrated EAs is quite close to that of the AAs in the present study, while the percentage of realization of the non-integrated EAs is fairly close to that of the EAs in the present study.

Hence, a third conclusion presented by the data here is that the results suggest that vocalization, particularly in the coda and syllable final environments, may be marker of blue collar AA linguistic identity in a similar fashion to Dodsworth's (2005) Worthington, OH speakers. Except in this case, AAs may be marking a distinctive cultural identity through the *increased* use of vocalized /l/ variants, rather than *decreased* use, as in Worthington. Given that Labov, et al's (1968) found that vocalization is also pervasive in the speech of blue collar New York City AAVE speakers, its increased use among Columbus AAs may also serve as a kind of "dual status" marker, both of distinctive local AA identity in Columbus, and as a way of indexing a more "national" AA identity. However, it should be noted that given the small N of Fix's (2004) study, as well as the lack of data drawn from *blue collar* EAs in the present study, more data need to be obtained and analyzed in a later study to confirm or disconfirm this conclusion.

Turning to a more general comparison of the AA results with those found in Labov et al.'s (1968) New York City study, when we compare the trends noted in the Columbus AAVE data with the New York City data, the findings presented here confirm three of the main patterns that were noted there. Specifically, both studies noted the pronounced influence of following consonants on conditioning vocalization, particularly when a preceding back vowel is present in the same word among AAs. As well, both studies report that preceding consonants and high front vowels have on the non-vocalization of /l/. In addition, both studies note that the strong influence of coda and syllable final environments on conditioning vocalized variants among AA speakers.

Two additional findings of Labov, et al. (1968) which are confirmed by the Columbus data, but have been qualitatively confirmed by the present study rather than quantitatively confirmed, are the following: a) that vocalization is categorical before liquids and glides (114), and b) that the height of the preceding vowel is not a major constraint on the rule (116). The confirmatory findings in the Columbus data concerning this first point are based on the observations of the three judges who rated the AA data. All three noted this phenomenon when listening to the data for coding purposes. Findings supporting Labov's second point are made available when we consider the AA data as it was presented in the VARBRUL analysis presented in section 4.2 versus how it was presented in the VARBRUL analysis contained earlier in section 5. Here, we see that vowel frontness is by far more important in influencing the vocalization of /l/ than vowel height, as the distributional patterns for vowel segments are more strongly stratified by this feature than by the feature of height.

6. Conclusion

As the results of this pilot study have demonstrated, our understanding of the patterns of variation involving the vocalization of /l/ among blue collar AA speakers, as well as the comparative similarity and difference of these patterns from those found among white collar EA speakers in Columbus, has begun to become rarified. /l/ vocalization has been shown to be a general feature of Columbus speech, utilized to some degree by all AA and EA speakers, with blue collar AA speakers showing the strongest tendencies towards vocalizing, particularly in coda and syllable final position. Furthermore, the results suggest vocalized /l/ may be a marker of AA racial and linguistic identity in Columbus, with this distinctive cultural identity signaled by AA speakers' increased use of vocalized variants.

However, as the results also reveal, additional data collection within both the AA and EA speech communities in Columbus is required, so a more detailed exploration of the conclusion regarding vocalization as a marker of AA identity can be explored. Specifically, data from blue collar EAs in Columbus needs to be obtained so that a more robust comparative study of EA and blue collar AA differences and similarities involving vocalization of /l/ can be completed. As well, further analysis of /l/ in the speech of the AA speakers analyzed needs to be undertaken so that the patterns of variation noted here can be more confidently assessed. Specifically, instrumental analysis of this data is essential, making use of comparative "normative data" recordings obtained from speakers of Columbus AAVE under laboratory conditions, so that a more detailed understanding of /l/-vocalization within this population can be acquired. A second area that needs to be addressed in a future study is the analysis of additional tokenized /l/ data extracted from the untranscribed portions of the AA data corpus. At the present time, only half of this data has been transcribed, and therefore, it is quite possible that nearly twice as many /l/ tokens could be extracted from the data in a future study.

References

Ash, Sharon. 1982. *The vocalization of /l/ in Philadelphia*. Doctoral dissertation, University of Philadelphia, Pennsylvania, PA.

Bailey, Guy, and Erik R. Thomas. 1998. Some aspects of African-American Vernacular English phonology. In Mufwene, Salikoko S, John Rickford, John Baugh, and Guy Bailey, *African American English*. London: Routeledge. pp. 85-109.

Carver, Craig. 1987. The midlands: The Upper South and Lower North dialects. *American regional dialects*. Ann Arbor, MI: University of Michigan Press. pp. 161-204.

Carver, Craig. 1993. Regional speech variation in Wisconsin. In Timothy C. Frazer (ed.), *"Heartland" English: Variation and transition in the American Midwest*. Tuscaloosa, AL: University of Alabama Press. pp. 199-214

Dodsworth, Robin. 2005. Attribute networking: A technique for modeling social perceptions. *The Journal of Sociolinguistics* 9.2:225-253.

Dodsworth, Robin, Bartek Plichta, and David Durian. 2006. An acoustic study of Columbus /l/ vocalization. *Paper presented at NWAV 35, Columbus, OH.*

Durian, David, Robin Dodsworth, and Jennifer Schumacher. In press. Convergence in urban working class Columbus, OH AAVE and EAE vowel systems? In Thomas, Erik R., and Malcah Yeager-Dror (Eds.), *Vowel phonology and ethnicity*. Publication of the American Dialect Society 93. Durham: Duke University Press.

Fasold, Ralph W., and Walt Wolfram. 1970. Some linguistic features of Negro dialect. In Ralph W. Fasold and Roger W. Shuy (eds.) *Teaching Standard English in the inner city*. Washington, DC: Center for Applied Linguistics. pp. 41-86.

Fix, Sonya. 2004. /l/ vocalization and racial integration of social networks: Sociolinguistic variation among whites in a Columbus Ohio community. Poster presented at NWAV(E) 33, Ann Arbor, MI.

Fridland, Valerie, and Kathy Bartlett. 2006. The social and linguistic conditioning of back vowelfronting across ethnic groups in Memphis, Tennessee. *English Language and Linguistics* 10.1:1-22.

Hardcastle, William, and William Barry. 1985. Articulatory and perceptual factors in /l/vocalizations in English. *Journal of the International Phonetic Association 15*:3-17

Hardy, Donald E. 2005. Technology and stylistics: The Web connection. In Caldas-Coulthtard, Carmen Rosa, and Michael Toolan (Eds.), *The writer's craft, the culture's technoloogy.* Ney York: Rodopi. pp. 195-208.

Horvath, Barbara M., and Robert J. Horvath. 2002. The geolinguistics of /l/ vocalization in Australia and New Zealand. *Journal of Sociolinguistics* 6:319-346.

Labov, William. 1970. *The study of non-standard English*. Champaign, IL: National Council of Teachers of English.

Labov, William. 1972. *Language in the inner city*. Philadelphia: University Pennsylvania Press.

Labov, William, and Benji Wald. 1969. Some general principles of chain shifting. Paper given at the Winter Meeting of the LSA, 1969.

Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, phonology, and sound change.* Berlin: Mouton de Gruyter.

Labov, William, Paul Cohen, Clarence Robins, and John Lewis. 1968. *A study of the Non-Standard English of Negro and Puerto Rican speakers in New York City. Volume I: Phonological and grammatical analysis.* Cooperative Research Report 3288. Philadelphia, PA: U.S. Regional Survey.

Lehiste, Ilse. 1964. *Acoustical characteristics of selected English consonants*. Bloomington, IN: Indiana University Press.

McGuire, Grant. 2002. The realization of interdental fricatives in Columbus, OH, AAVE. Paper presented at the Montreal-Ottawa-Toronto Phonology Workshop, Toronto.

Pitt, Mark.A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume, and Eric Fosler-Lussier. 2007. *Buckeye Corpus of Conversational Speech* (2nd release). Columbus, OH: Department of Psychology, Ohio State University. Available online: http://www.buckeyecorpus.osu.edu/

Sproat, Richard, and Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21:291-312.

Thomas, Erik R. [1989]/1993. Vowel changes in Columbus, OH. *Journal of English Linguistics* 22:205-215.

Thomas, Erik R. 2001. *An acoustic analysis of vowel variation in New World English*. Publications of the American Dialect Society, 85. Durham, NC: Duke University Press.

Thomas, Erik R., and Malcah Yeager-Dror (Eds.). In press. *Vowel phonology and ethnicity*. Publication of the American Dialect Society 93. Durham: Duke University Press. U.S. Census Bureau. 2000. *United States Census 2000*. Available online: http://www.census.gov

Weldon, Tracey. 1994. Variability in negation in African-American Vernacular English. *Language Variation and Change* 6.3:359-397.

Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Wolfram, Walter. 1969. *A linguistic description of Detroit Negro speech.* Washington, DC: Center for Applied Linguistics.

Wolfram, Walt, and Donna Christian. 1976. *Appalachian speech*. Arlington, VA: Center for Applied Linguistics.

Wolfram, Walt, and Ralph W. Fasold. 1974. *The study of social dialects in American English*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Wolfram, Walt, and Erik R. Thomas. 2002. *The development of African American English*. Oxford, Blackwell.

KNOWLEDGE- AND LABOR-LIGHT MORPHOLOGICAL ANALYSIS[1]

Jirka Hana
The Ohio State University

Abstract

We describe a knowledge and labor-light system for morphological analysis of fusional languages, exemplified by analysis of Czech. Our approach takes the middle road between completely unsupervised systems on the one hand and systems with extensive manually-created resources on the other. For the majority of languages and applications neither of these extreme approaches seems warranted. The knowledge-free approach lacks precision and the knowledge-intensive approach is usually too costly. We show that a system using a little knowledge can be effective. This is done by creating an open, flexible, fast, portable system for morphological analysis. Time needed for adjusting the system to a new language constitutes a fraction of the time needed for systems with extensive manually created resources: days instead of years. We tested this for Russian, Portuguese and Catalan.

[1]This paper differs only slightly from a paper presented at the Department of Linguistics of The Ohio State University in 2004 and finished in 2005. The morphological analyzer described in this paper was developed as part of a joint project with Anna Feldman and Chris Brew aimed at developing a portable resource-light tagger. I thank Chris Brew, Anna Feldman, Jan Hajič, Brian Joseph and Detmar Meurers for providing valuable feedback and to Steve Evans for helping me in my struggle with the English language.

The work described in this paper was partially supported by NSF CAREER Award 0347799.

Jirka Hana. Knowledge- and labor-light morphological analysis. *OSUWPL*, Volume 58, Fall 2008, pp. 52–84.

1 Introduction

This paper describes a knowledge- and labor-light system for morphological analysis of Slavic languages, namely Czech and Russian. Our approach takes the middle road between completely unsupervised systems à la (Goldsmith 2001) on the one hand and systems with extensive manually-created resources à la (Hajic 2004) on the other. These approaches are scientifically interesting and there are cases when they are also practically justifiable (e.g., the former for analyzing understudied languages and the latter for applications requiring very high precision). However we believe that for the majority of languages and majority of purposes neither of these extreme approaches seem warranted. The knowledge-free approach still lacks precision and the knowledge-intensive approach is usually too costly. We show that a system that uses a little knowledge can be effective. We exploit the 80:20 rule: The part of the work that is easy to do *and* that matters most is done manually or semi-automatically and the rest is done automatically.

Czech this way? We use Czech to test our hypotheses. We do not suggest that morphological analysis of Czech should be designed exactly in the way we do. An excellent high precision system using manual resources[2] already exists (Hajic 2004). The main reason for working with Czech is that we can easily evaluate our system on the Prague Dependency Treebank – a large morphologically annotated corpus (http://ufal.mff.cuni.cz/pdt).

However, no manual resources, including those of (Hajic 2004), can cover arbitrary text – there is an unbounded universe of names (people, products, companies, musical groups, ...) technical terms, neologisms, quotes from other languages; typos, ... We suggest that for languages such as Czech and Russian, morphological analysis should rely on extensive manual resources backed up by a system similar to ours. Less dense languages (e.g., Sorbian, Romany, Czech used in chat-rooms or in any other specialized settings, etc.) can use less of the expensive manual resources and more of the automatic or semiautomatic resources.

The system. For our work, we developed an open, flexible, fast and portable system for morphological analysis. It uses a sequence of analyzing modules. Modules can be reordered, added or removed from the system. And although we provide a basic set of analyzing modules, it is possible to add other modules for specific purposes without modifying the rest of the system. The modules we provide are re-usable for both resource-light and resource-intensive approaches, although the latter option is not explored in detail here.

Nouns only. In the rest of the paper we focus exclusively on nouns. We have several reasons for this:

1. they are hard for the unsupervised systems, because their endings are highly homonymous (at least in Slavic languages);

[2]We use the term *manual resources* to refer to manually-created resources, *automatic resources* to automatically created resources (with possibly some minor manual input) and *semi-automatic resources* to automatic resources manually corrected (fully or partially).

Lemma freq decile	Number of tokens	Corpus coverage (%)	Cumulative coverage (%)	Lemmas not in `tr2` (%)
10	164 643	74.1	74	0.2
9	22 515	10.1	84	6.7
8	11 041	5.0	89	22
7	6 741	3.0	92	36
6	4 728	2.1	94	48
5	3 179	1.4	96	61
4	2 365	1.1	97	65
3	2 364	1.1	98	70
2	2 364	1.1	99	75
1	2 364	1.1	100	77

Note: Each decile contains 2364 or 2365 noun lemmas.

Table 1: Corpus coverage by lemma frequency

2. they are the class where the manually-created resources approach fails the most – they are the most open class of all (consider proper names);

3. for practical reasons, we have to limit the scope of our work.

Written language. Finally, it is necessary to stress that we are concerned only with analysis of a written text, not speech.

2 Motivation – Lexical statistics of Czech

To motivate our approach, we provide some statistics about Czech nouns, assuming that nouns in other Slavic languages behave similarly. The statistics are based on the `t1` and `tr2` corpora (§A.1). The `tr1` corpus contains 222 304 noun tokens (out of 619 984 all tokes) corresponding to 42 212 distinct forms (87 321) and 23 643 lemmas (43 056).[3]

Table 1 and Figure 1 break lemmas into deciles by their frequency and compare their corpus coverage. Similarly as the Zipf's law (Zipf 1935; Zipf 1949), they make two things apparent:

- It is quite easy to get a decent coverage of a text with a small number of high frequency lemmas. The 2.4K lemmas in the 10th decile cover 3/4 of noun tokens in

[3]The lemmas in `tr1` (and in the whole PDT), distinguish not only between homonyms but often also between words related by polysemy. For example, there are at least four different lemmas for the word *strana*: strana-1 'side (in space)', strana-2 'political party' strana-3 '(contracting) party, (on somebody's) side, ..', strana-4 'page'. All four have the same morphological properties – it is a feminine noun, paradigm *žena*. While this statistics treats them as four distinct entities, our Guesser and automatically acquired lexicons do not distinguish between them. However, the statistics are still valid, because only relatively few lemmas have such distinction.

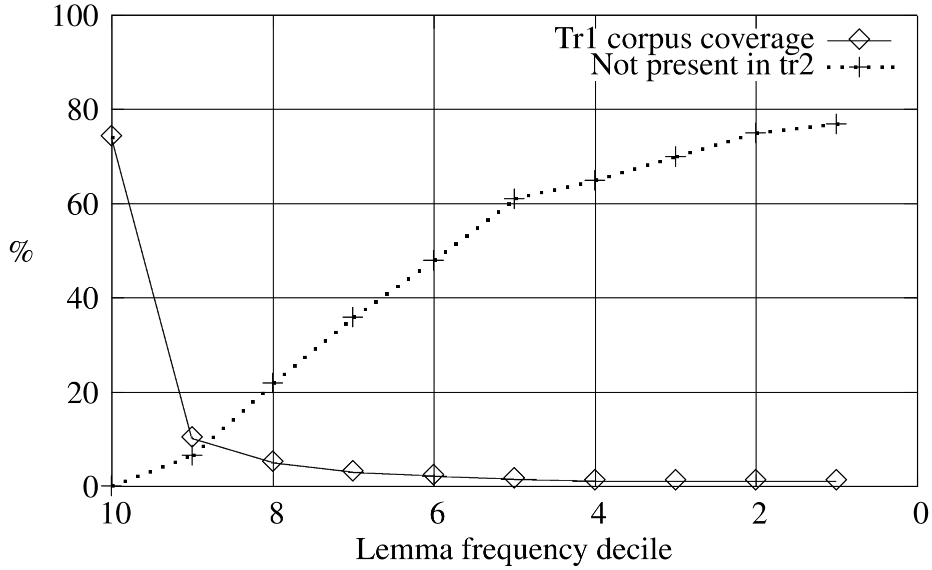


Figure 1: Lemma characteristics by frequency

the corpus, 7.1K lemmas in the top three deciles cover nearly 90% of all noun tokens. That means that even in labor-light systems, it is not necessary to go the way of completely automatically acquired morphology.

- It is very hard, practically impossible, to get a perfect coverage of a running text even with very large lexicons.
  - First, the lemmas in each of the lower deciles add relatively much smaller coverage.
  - Second, infrequent lemmas also tend to be text specific. 77% of the lemmas in the lowest decile of the `tr1` corpus did not occur in the `tr2` corpus – even though the corpora are very similar (they both consists from texts from the same newspapers and magazines). Even when we take the first half of the lemmas (decile 1-5), 70% of the lemmas are text specific!

These facts justify our approach – to provide manually a small amount of information that makes the most difference and let the system learn the rest. This makes it possible to keep the amount of necessary labor close to that of the unsupervised system with quality not much worse than that of the expensive system with manual resources.

3 A Morphological Analyzer of Czech

In this section, we introduce both the general framework for doing and training resource-light morphological analyzes and its instantiation on Czech. Application to other languages is discussed in §4.

In this section, we discuss the analyzer in general (§3.1); the strategy of using it (§3.2); how morphological paradigms are seen by a linguist and how by our system (§3.3); automatic creation of morphological resources – a large lexicon (§3.4) and a list of abbreviations (§3.5). Finally, we evaluate the whole system in §3.6 and suggest several possible enhancements in §3.7.

3.1 Morphological analyzer

Morphological analysis is a function that assigns a set of lemmas (base forms), each with a set of tags, to a form:

(1) MA: form $\rightarrow$ set(lemma $\times$ set(tag))

ženou	→	{ (*žena* 'woman',	{noun fem sing inst }),	
		(*hnát* 'hurry',	{verb pres pl 3rd }) }	
ženy	→	{ (*žena* 'woman',	{noun fem sing gen,	
			noun fem pl nom,	
			noun fem pl acc,	
			noun fem pl voc }) }	

Our goal was to design an open, fast, portable and easily configurable morphological analyzer. It is a modular system that queries its analyzing modules in a particular order. Any module can be loaded several times with different parameters (say, different lexicons). A module receives information about the word, its potential prefixes and its context (currently just the preceding word with its analysis, and the following word). The module returns zero or more analyses. An analysis must contain information about a lemma and a tag. Depending on the mode the morphological analyzer is run in, it can also contain additional information, like a paradigm name, ending length, etc.

3.2 General Strategy

We focus our work and knowledge on creating a limited amount of resources that make the most difference and that are easy to create. The rest is done automatically. The system uses a mix of modules with various level of precision and invested effort. The modules are run in a cascading way. Modules that make less errors and overgenerate less are run before modules that make more errors and overgenerate more. Modules on the subsequent level are used for analysis only if the modules from the previous level did not succeed (although this is configurable).

The system contains three types of modules (in addition there are specialized modules for handling numbers, abbreviations, symbols, etc.):

1. Simple word lists – each word form is accompanied by information about its lemma and tags.

2. Guesser – analyzes words using only information about paradigms.

   On the plus side, (1) the Guesser has a high recall and (2) is very labor-light – it is enough to specify the paradigms. However, the disadvantages are that (1) it has a low precision (overgenerates a lot) and (2) it is quite slow – there are too many things to check and perform on too many analyses.

3. Lexicons – analyzes words using a lexicon and a list of paradigms.

   Lexicon-based analysis has just the opposite properties of the Guesser. It requires a lexicon, which is usually very costly to produce. However, (1) only analyses that match the stem in the lexicon and its paradigm are considered; (2) it is very fast, because stem changes, etc. can be computed before hand and be simply listed in the lexicon. The problem of the costly lexicon is partly addressed in §3.4.

Traditional labor-intensive systems use information about paradigms together with a large lexicon, possibly backed up by a guesser (e.g., Hajic 2004; Mikheev & Liubushkina 1995). Word lists are usually used for languages with simple inflectional morphology like English. It might seem obvious that for Czech, a language with 7 cases, 2 numbers and 4 genders, form lists are out of the question. However, in practice only few lemmas occur in a larger number of forms. Table 2 summarizes the distribution of lemma occurrences in the `tr1` corpus in terms of the number of encountered forms. It can be seen that 64% of the lemmas occur only in one form.

	Lemmas		
Nr of forms	Count	Percentage	Cumulative percentage
1	15 192	64.26	64
2	4 155	17.57	82
3	1 807	7.64	89
4	948	4.01	93
5-9	1 523	6.44	100
10-17	18	0.08	100
Total	23 643	100	

Table 2: Noun lemma distribution by the number of forms in the corpus

Entering a lexicon entry is very costly. While it is usually easy (for a native speaker) to assign a lemma to one of the major paradigm groups, it takes considerably more time to select the exact paradigm variant differing only in 1 or 2 forms (in fact, this may be even idiolect-dependent). For example, it is easy to see that *atom* 'atom' does not decline according to the neuter paradigm *město* 'town' but it takes more time to decide to which of the hard masculine inanimate paradigms it belongs (See Table 3). On the other hand, entering possible analyses for individual word forms is usually very straightforward.

Therefore, our system uses a list of manually entered analyses for the most common forms, an automatically acquired lexicon for less common words and finally, the ending-based guesser as a safety net covering the rest.

		hard masculine inanimate paradigms				
	atom ‘atom’	hrad ‘castle’	ostrov ‘island’	rybník ‘pond’	zámek ‘chateau’	domeček ‘small house’
S1	atom-0	hrad-0				
S2	atom-u	hrad-u	ostrov-u/a	rybník-u/a		
S3	atom-u	hrad-u				
S4	atom-0	hrad-0				
S5	atom-e	hrad-e			zámk-u	domečk-u
S6	atom-u	hrad-ě/u		rybníc-e/ík-u	zámk-u	
S7	atom-em	hrad-em				
P1	atom-y	hrad-y				
P2	atom-ů	hrad-ů				
P3	atom-ům	hrad-ům				
P4	atom-y	hrad-y				
P5	atom-y	hrad-y				
P6	atom-ech	hrad-ech			zámc-ích	domečc-ích/čk-ách
P7	atom-y	hrad-y				

Table 3: Forms of *atom* ‘atom’ and the hard masculine inanimate paradigms

Note that the process of providing the form list is not completely manual – a native speaker selects the correct analyses from those suggested by the ending-based guesser. Analyses of closed-class words can be entered by a non-native speaker on the basis of a basic grammar book. Finally, there is the possibility to manually process the automatically acquired lexicon: a native speaker removes the most obvious errors for the most frequent lexical entries. They remove errors that are easy to identify and that have the highest impact on the results of the system. We did not use this possibility when building the analyzer for Czech, but we did use when annotating development corpora for Portuguese and Russian.

3.3 Czech paradigms

3.3.1 Czech paradigms seen by a linguist

Simply put, in a fusional language like English or Czech, a paradigm is a set of endings with their tags, e.g., *0* – noun singular, *s* – noun plural. The endings are added to stems producing word forms characterized by those tags, e.g., *cat* – noun singular, *cats* – noun plural. However, life is not easy, and the concatenation is often accompanied by various more or less complicated phonological/graphemic processes affecting the stem, the ending or both, e.g., *potato-**e**s, countr**i**-es, kni**v**-es*, etc.

As a more complex illustration, consider several examples of Czech nouns belonging to the *žena* ‘woman’ paradigm, a relatively ‘well-behaved’ paradigm of feminine nouns, in Table 4.[4] Without going too deeply into linguistics, we can see several complications:

[4] Abbreviations of morphological categories, e.g., `S1` – singular nominative, are based on Hajic’s (2004)

	woman	owl	draft	goat	iceberg	vapor	fly
S1	žen-a	sov-a	skic-a	koz-a	kr-a	pár-a	mouch-a
S2	žen-y	sov-y	skic-**i**	koz-y	kr-y	pár-y	mouch-y
S3	žen-ě	sov-ě	skic-**e**	koz-**e**	k**ř-e**	pá**ř**-e	mou**š-e**
S4	žen-u	sov-u	skic-u	koz-u	kr-u	pár-u	mouch-u
S5	žen-o	sov-o	skic-o	koz-o	kr-o	pár-o	mouch-o
S6	žen-ě	sov-ě	skic-**e**	koz-**e**	k**ř-e**	pá**ř**-e	mou**š-e**
S7	žen-ou	sov-ou	skic-ou	koz-ou	kr-ou	pár-ou	mouch-ou
P1	žen-y	sov-y	skic-**i**	koz-y	kr-y	pár-y	mouch-y
P2	žen-0	sov-0	skic-0	koz-0	k**e**r-0	p**a**r-0	m**u**ch-0
P3	žen-ám	sov-ám	skic-ám	koz-ám	kr-ám	pár-ám	mouch-ám
P4	žen-y	sov-y	skic-**i**	koz-y	kr-y	pár-y	mouch-y
P5	žen-y	sov-y	skic-**i**	koz-y	kr-y	pár-y	mouch-y
P6	žen-ách	sov-ách	skic-ách	koz-ách	kr-ách	pár-ách	mouch-ách
P7	žen-ami	sov-ami	skic-ami	koz-ami	kr-ami	pár-ami	mouch-ami

Table 4: Examples of the *žena* paradigm nouns

1. Ending variation: *žen-ě*, *sov-ě* vs. *burz-e*, *kř-e*, *pář-e*; *žen-y* vs. *skic-i*.

   The dative and local sg. ending is *-ě* after alveolar stops (*d, t, n*) and labials (*b, p, m, v, f*). It is *-e* otherwise.

   Czech spelling rules require the ending *-y* to be spelled as *-i* after certain consonants, in this case: *c, č, ď, ň, š*. The pronunciation is the same ([ɪ]).

2. Palatalization of the stem final consonant: *kr-a – kř-e*, *mouch-a – mouš-e*.

   The *-ě/e* ending affects the preceding consonant: *ch* [x] → *š*, *g/h* → *z*, *k* → *c*, *r* → *ř*.

3. Epenthesis: *kr-a – ker*.

   Sometimes, there is an epenthesis in genitive plural. This usually happens when the noun ends with particular consonants. There are certain tendencies, but in the end it is just a property of the lexeme; cf. *občank-a – občanek* 'she-citizen, id-card' vs. *bank-a – bank* 'bank' (both end with *nk*, but one epenthesises and the other not). Some nouns allow both possibilities, e.g., *jacht-a – jachet/jacht* 'yacht'

4. Stem internal vowel shortening: *pár-a – par*.

   Often the vowels *á, í, ou* shorten into *a, i/ě, u* in gen. pl. and sometimes also in dat., loc. and ins. pl. If the vowel is followed by multiple consonants in nom. sg, the shortening usually does not happen. In many cases there are both short and long variants (*pár-a – par – pár-ám/par-ám*, *pár-ách/par-ách*, *pár-ami/par-ami* 'vapor'), usually stylistically different.

tagset, the tagset we use, and are discussed in §A.3.

form	lemma	gloss		category
měst-a	město	town	NS2	noun neut sing gen
			NP1 (5)	noun neut pl nom (voc)
			NP4	noun neut pl acc
tém-a	téma	theme	NS1 (5)	noun neut sing nom (voc)
			NS4	noun neut sing acc
žen-a	žena	woman	FS1	noun fem sing nom
pán-a	pán	man	MS2	noun masc anim sing gen
			MS4	noun masc anim sing acc
ostrov-a	ostrov	island	IS2	noun masc inanim sing gen
předsed-a	předseda	president	MS1	noun masc anim sing nom
vidě-l-a	vidět	see	VpFS	verb past fem sing
			VpNP	verb past neut pl
vidě-n-a			VsFS	verb passive fem sing
			VsNP	verb passive neut pl
vid-a			VeMS	verb transgressive masc sing
dv-a	dv-a	two	CYS1	numeral masc sing nom
			CYS4	numeral masc sing acc

Table 5: Homonymy of the *a* ending.

It would be possible to discuss in a similar manner all the Czech (noun) paradigms. Depending on how you count, there are roughly 13 basic paradigms – 4 neuter, 3 feminine and 6 masculine; plus there are nouns with adjectival declension (another 2 paradigms). In addition, there are many subparadigms and subsubparadigms, all of which involves a great amount of irregularity and variation on the one hand and a great amount of homonymy on the other (see Table 5). For a more detailed discussion, see for example (Karlík *et al.* 1996; Fronek 1999).

3.3.2 Czech paradigms seen by an engineer

There are two different ways to address phonological/graphemic variations and complex paradigm systems when designing a morphological analyzer:

- A linguistic approach. Such a system employs a phonological component accompanying the simple concatenative process of attaching an ending. This implies a smaller set of paradigms and morphemes. Two-level morphology (Koskenniemi 1983; Koskenniemi 1984) is an example of such a system and (Skoumalová 1997) is an example for Czech. The problem is that implementing morphology of a language in such a system requires a lot of linguistic work and expertise. For many languages, the linguistic knowledge is not precise enough. Moreover, it is usually not straightforward to translate even a precisely formulated linguistic description of a morphology into the representation recognized by such system.

  In Czech, the forms of the noun *kra* 'iceberg$_{\texttt{FS1}}$', *kře* 'iceberg$_{\texttt{FS36}}$', *ker* 'iceberg$_{\texttt{FP2}}$'

etc. (see Table 4) would be analyzed as involving the stem *kr*, the endings *-a*, *-ě* and *-0* and phonological/graphemic alternations. Forms of the noun *žena* ‘woman$_{\texttt{FS1}}$’ (*ženě* ‘FS36’, *žen* ‘FP2’, etc.) would belong to the same paradigm as *kra*.

- An engineering approach. Such a system does not have a phonological component, or the component is very rudimentary. Phonological changes and irregularities are factored into endings and a higher number of paradigms. This implies that the terms *stem* and *ending* have slightly different meanings than they traditionally do. A stem is the part of the word that does not change within its paradigm, and the ending is the part of the word that follows such a stem.

  Examples of such an approach are (Hajic 2004) for Czech and (Mikheev & Liubushkina 1995) for Russian. The previous version of our system (Hana *et al.* 2004) also belongs to this category. The advantages of such a system are its high speed, simple implementation and straightforward morphology specification. The problems are a very high number of paradigms (several hundreds in the case of Czech) and impossibility to capture even the simplest and most regular phonological changes and so predict the behavior of new lexemes.

  For example, the English noun paradigm above (*0 – s*) would be captured as several other paradigms including, *0 – s*, *0 – es*, *y – ies*, *f – ves*.

  In Czech, the forms of the noun *kra* ‘iceberg$_{\texttt{FS1}}$’ would be analyzed as involving the stem *k* followed by the endings *-ra*, *-ře* and *-er*. Forms of the nouns *žena* ‘woman$_{\texttt{FS1}}$’ and *kra* would belong to two different paradigms.

Our current system is a compromise between these two approaches. It allows some basic phonological alternations (changes of a stem-tail[5] and a simple epenthesis), but in many cases our *endings* and *stems* are still different from the linguistically motivated ones. Therefore, many of the paradigms are still technical.

Currently, our system is capable of capturing all of the processes described in §3.3 except the stem internal vowel shortening:

1. Ending variation: A paradigm can have several subparadigms. There are three paradigms corresponding to the linguistic paradigm *žena* (see Table 4): `NFzena`, subparadigm `NFkoza` and its subparadigm `NFskica`.
   - A subparadigm specifies only endings that are different from the main paradigm. `NFkoza` is like `NFzena` but has *-e* in `S3` and `S6`; `NFskica` is like `NFkoza` but has *-i* in `S2`, `P1`, `P4` and `P5`.
   - Each paradigm restricts the possible tails of stems that can decline according to it. For example, `NFskica` requires the stems to end in *c, č, ž, š* or *j*.
2. Palatalization: A paradigm can specify a simple replacement rule for changing stem-tails. For example, the paradigm `NFkoza` says that stem-final *ch* changes to *š* in `S3` and `S6`.

[5]We use the term *tail* to refer to a final sequence of characters of a string. We reserve the word *ending* to refer to those tails that are morphemes (in the traditional linguistic sense or in our technical sense).

lemma	gloss	paradigm	$stem_1$	$stem_2$	$stem_3$
žena	woman	NFzena	žen	=1	—
sova	owl	NFzena	sov	=1	—
chodba	corridor	NFzena	chodb	chodeb	—
skica	draft	NFskica	skic	=1	=1
koza	goat	NFkoza	koz	=1	=1
kra	iceberg	NFkoza	kr	ker	kř
pára	vapor	NFkoza	pár	par	pář
moucha	fly	NFkoza	mouch	much	mouš
váha	weight	NFkoza	váh	=1	váz

Table 6: Examples of lexical entries for some nouns of the *žena* paradigm

3. Epenthesis: An ending can be marked as allowing epenthesis. All the three paradigms allow epenthesis in `P2`.

The current paradigm module cannot capture stem vowel changes. Therefore, the Guesser analyzes such forms incorrectly. It still provides the correct tags but not the correct lemma. For example, *par* is analyzed as a form of the incorrect lemma *para* instead of the correct *pára*; the tag `NNFP2-----A----` is correct.

Our system specifies 64 noun paradigms (still not exploiting all the possibilities) and 14 common paradigms for adjectives and verbs. The choices on what to cover involve a balance between precision, coverage and effort. More work would be somewhat beneficial but our goal is to stop before the return on effort becomes too low.

3.3.3 Paradigms and Lexicons

A lexicon entry contains information about the lemma, its paradigm and stem or stems. The Lexicon-based Analyzer does not use the information about stem changes that Guesser uses, but instead refers to the stems listed directly in the lexicon entry. This not only speeds up the processing but also makes it possible to capture phonological changes or irregularities that the Guesser is currently unable to handle, including the stem vowel changes mentioned above. Table 6 lists several lexicon entries, for most of them the full declensions can be found in Table 4. $Stem_2$ is used in genitive plural (`P2`) for all paradigms. This stem expresses epenthesis (*chodb* → *chodeb*) and stem vowel shortening (*pár* → *par*). Entries belonging to the `NFskica` or `NFkoza` paradigms can specify a third stem used in dative and locative singular (`S3`, `S6`). This stem expresses palatalization (*mouch* → *mouš*).

3.4 Lexicon acquisition

The morphological analyzer supports a module or modules employing a lexicon containing information about lemmas, stems and paradigms. There is always the possibility to provide this information manually. That, however, is very costly. In this section we describe how to acquire a lexicon approximation from a large raw corpus.

This approach differs from the work by Mikheev (1997) or Hlaváčová (2001). Mikheev's (1997) algorithm attempts to acquire a lexicon that would cover forms not covered by a large manually created lexicon. Similarly, Hlaváčová (2001) describes a guesser that acquires rules for analyzing unknown words on the basis of a large set known words (it associates tails, usually endings, often preceded by a final part of a stem, with tags). In other words, in both cases it is assumed that a manually created lexicon covers most of the text and the automatically created lexicon or rules are used only as a backup. In our case, it is the main lexicon that is acquired automatically (note that our form lists are significantly smaller than the lexicons used in Mikheev (1997) or Hlaváčová (2001)).

3.4.1 General idea

The general idea is very simple. The ending-based Guesser module overgenerates. Part of the ambiguity is usually real but most of it is spurious. We can use a large corpus to weed the spurious analyses out of the real ones. In such a corpus, open-class lemmas are likely to occur in more than one form. Therefore, if a lemma-stem-paradigm candidate suggested by the Guesser occurs in other forms in other parts of the corpus, it increases the likelihood that the candidate is real and vice versa.

To make it more concrete: if we encounter the word *talking* in an English corpus, using the information about paradigms, we assume that it is either the *-ing* form of the lemma *talk* or that it is a monomorphemic word (such as *sibling*). Based on this single form we cannot really say more. However, if we also encounter the forms *talk*, *talks* and *talked*, the former analysis seems more probable; and therefore, it seems reasonable to include the lemma *talk* as a verb into the lexicon. If we encountered also *talkings*, *talkinged* and *talkinging*, we would include both lemmas *talk* and *talking* as verbs.

3.4.2 Examples & Problems

We can use our Morphological Analyzer to analyze all the words in the corpus and then create all the possible hypothetical lexical entries consistent with these analyses. After that, we would like to run some filtering that would drop most of the bad entries and leave a small number of entries that would include the good ones. In this subsection, we discuss some of the problems associated with such a filtering.

Let's consider for example the lemma *podpora* 'support'. It is a feminine noun belonging to (a variant of) the *žena* paradigm. The `raw` corpus contains 8138 tokens of this lemma in 9 forms – see Table 7.[6] There are 192 (!) ways to assign a lemma and a paradigm to various subsets of these forms (see Table 8). Most of them sound very funny to a native speaker; only a minority sounds funny to an average learner of Czech; none sounded funny to our Guesser. In this case, we are lucky that we got nearly all the forms of the paradigm, only the vocative singular form is missing, which is not very surprising.

We could select the hypothetical entry that has the highest number of forms. While it would be the correct choice in this case, this strategy would not work in all cases. Con-

[6] We ignore all colloquial forms.

forms	possible case	occurences
podpor-a	S1	810
podpor-y	S2, P1, P4, P5	1633
podpoř-e	S3, S6	782
podpor-u	S4	4128
podpor-o	S5	0
podpor-ou	S7	625
podpor-0	P2	123
podpor-ám	P3	11
podpor-ách	P6	20
podpor-ami	P7	6
podporaa	typo	1

Table 7: Forms of the lemma *podpora* in the `raw` corpus.

# of covered forms	# of entries
9	1
6	2
5	2
4	8
3	7
2	3
1	169

Table 8: Candidate entries for *podpora* forms.

sider for example the noun *bezvědomí* 'unconsciousness' and the adjective *bezvědomý* 'unconscious'. Ignoring negation, *bezvědomí* has 4 theoretical forms, but one of them accounts for 70% of the categories, moreover those much more frequent ones (cf. *pondělí* 'Monday', which declines the same way, in Table 16). *Bezvědomý* has potentially more than 20 forms (cf. *mladý* in Table 17). The problem is that the common form of the former is also a form of the latter. So if we considered a simple majority of forms, the nouns similar to *bezvědomí* would usually lose. We could instead compare the realized percentages of the theoretical number of forms. However, this unnaturally penalizes paradigms with the following properties:

- Paradigms with distinct rare forms. There are many rare categories that are not realized even for a common lemma. For example, vocative is extremely rarely found in a written text. However, for certain paradigms, the form is very easy to find because it is simply the same as a form of a frequent category (e.g., *bezvědomí* 'unconsciousness$_{\texttt{S5/1/2/3/4/...}}$', vs. *pane* 'Mister$_{\texttt{S5}}$' (only `S5`)).
- Paradigms with large number of distinct forms in general. One form is enough to see 25% of forms of a word like *bezvědomí*, while 5 forms are necessary for the same percentage of a word like *bezvědomý*.
- Paradigms with alternative forms: The paradigm *hrad* has only one nominative plural, while the paradigm *pán* has two (e.g., *páni* / *pánové* 'gentlemen$_{\texttt{P1}}$'). Should we count those alternative forms as one or as two? What if some (but not all) work also for a different category?

A different problem is presented by "stolen" forms. Consider the word *atom* 'atom', an inanimate noun of the *hrad* paradigm. The `raw` corpus contains 161 tokens of this lemma in 7 forms – see Table 9. Seeing those 7 forms is not enough to decide whether the words belongs to an animate or inanimate paradigm. There are 5 paradigms each covering all 7 forms; see Table 10 listing two of them. If the `raw` corpus contained only those forms, we could simply keep all 5 hypotheses and still be happy to drop the other 122 hypotheses covering smaller number of forms. The problem is that the corpus also contains 208 tokens of the adjective *atomové* 'atomic$_{\texttt{FS2/FP1/..}}$' that however also fit nom. pl. of the animate paradigm *pán*. Therefore the incorrect paradigm *pán* seems to cover more forms than the correct *hrad* paradigm.[7]

For a native speaker of Czech, it is hard to resist mentioning some of the other non-existing lexical entries our algorithm found at various levels of development:

- Neuter noun *bylo* (paradigm *město*; forms *bylo*$_{S14}$, *byla*$_{S2/P14}$, *byl*$_{P2}$, *byly*$_{P7}$). In fact these are past participle forms of the verb *být* 'to be': *bylo*$_{NS}$, *byla*$_{FS/NP}$, *byl*$_{MS}$, *byly*$_{FP/IP}$. The word lists providing analyses for the most frequent word forms fix this particular problem.
- Neuter noun *architektuře* 'baby architect?' (paradigm *kuře* 'chicken'; form *architektuře*$_{S14}$). In fact, it is a form of the feminine noun *architektura* 'architecture' (*architektuře*$_{S36}$).

[7] It does not help that the corpus also contains the name *Atoma* which looks like animate gen/acc. sg.

forms	possible case	occurrences	
atom-0	S1, S4	48	36%
atom-u	S2, S3, S6	28	21%
atom-e	S5	0	0%
atom-em	S7	1	0%
atom-y	P1, P4, P5, P7	22	17%
atom-ů	P2	30	23%
atom-ům	P3	1	0%
atom-ech	P6	1	0%
Total		132	100%

Table 9: Forms of the lemma *atom* in the `raw` corpus.

	masculine inanimate	*atom* in `raw`	masculine animate	*atom* in `raw`
S1	hrad-0	+	pán-0	+
S2	hrad-u	+	pán-a	
S3	hrad-u	+	pán-u/ovi	+/–
S4	hrad-0	+	pán-a	
S5	hrad-e		pan-e	
S6	hrad-ě/u	–/+	pán-u	+
S7	hrad-em	+	pán-em	+
P1	hrad-y	+	pán-i/ové	–/(+)
P2	hrad-ů	+	pán-ů	+
P3	hrad-ům	+	pán-ům	+
P4	hrad-y	+	pán-y	+
P5	hrad-y	+	pán-i	
P6	hrad-ech	+	pán-ech	+
P7	hrad-y	+	pán-y	+
Total		7		7 (8)

Table 10: Fit of the forms of *atom* to the *hrad* and *pán* paradigms.

- Masculine animate noun *papír* (paradigm *pán*; forms: *papír*$_{\mathtt{S1}}$, ***papírové***$_{\mathtt{P15}}$, *papíru*$_{\mathtt{S6}}$, ...). In fact, these are forms of the nonanimate noun *papír* 'paper' (*papír*$_{\mathtt{S14}}$, *papíru*$_{\mathtt{S236}}$, ...), and the adjective *papírový* 'made from paper' (*papírové*$_{\mathtt{FP145/IP14/MP4}}$)

  The set of endings of animate and inanimate declensions are very similar. One of the distinctions is that only animate paradigms can contain the *-ové* ending (P15). However, many adjectives are derived from inanimate nouns by suffix *-ov-* that can be in certain forms followed by *-é*.[8] The simple higher-number-of-forms wins approach would produce systematic errors.

3.4.3 The algorithm

The algorithm has 4 steps:

1. **Morphological analysis of a raw corpus.**

   For this we can use any morphological analysis that provides information not only about lemmas and tags, but also about the paradigms used. We used our MA system configured to provide the necessary information.

2. **Creating all possible hypothetical lexical entries.**

   Every entry has to contain information about its lemma, paradigm and set of forms that occurred in the corpus.

3. **Filtering out bad entries.**

   The general idea is that the entry that covers the highest number of forms wins. However, taking into account the problems mentioned above, we allow several refinements:

   - Certain forms can be excluded from the counting. Used for endings that cause systematic errors. See the example with *papírové* at the end of the previous section.
   - Certain entries are not dropped even when competing entries cover more forms. Used for paradigms with very low number of distinct forms: *stavení* or *jarní*.
   - An entry covering less frequent forms (e.g., instrumental or vocative) need not be considered if it does not cover frequent forms as well (e.g., nominative).
   - Size of the winning crust can be specified, in relative or absolute terms. A crust of, say, 15% means that not only the entries with the highest number of forms, but also entries with the number of forms 15% smaller are kept. This decreases the precision of the lexicon but increases the recall (i.e., leads to a higher ambiguity and a lower error rate).
   - Minimal number of tokens and/or forms for an entry can be specified. This allows limiting the algorithm to entries with statistically reliable number of forms/tokens.

[8] See Table 10 for an example of animate and inanimate paradigms and Table 17 for adjectival paradigms.

4. **Creating a lexicon.**

   This step is quite uninteresting – it is necessary to create appropriate lexical entries for items that survived all the filtering. For that we need information about the lemma and paradigm which we have and about stem(s) which we can easily derive.

3.5 All caps abbreviation acquisition

The purpose of this subsection is to show that simple methods that do not rely on language-dependent work can be effective. Our goal is to acquire a list of usual all-caps abbreviations (e.g., *BMW*, *ČR* 'Czech Republic', *OSN* 'United Nations', *USA*, *OECD*) from a large unannotated corpus. This is different from the goals of many papers dealing with finding definitions of highly specific abbreviations in scientific tests (e.g., Chang *et al.* 2002; Larkey *et al.* 2000; Yeates 1999; Yeates *et al.* 2000). Here, we are interested in frequent abbreviations and we do not attempt to find their meaning (since they are common, they are usually not defined in the text as the papers cited above require).

3.5.1 A naïve approach and its problems

The solution might look very straightforward: It seems to be enough to extract simply all words that occur only in all caps. However, there are at least three problems with this approach (all frequencies are relative to the `raw` corpus, see §A.1):

1. For various reasons (typographical conventions, errors, etc.) all-caps abbreviations are sometimes written in lower case. For example, *ČTK* 'Czech News Agency' (about 550 occurrences, or 18% are in lower case, abbreviation of article sources are often written in lower case; similarly *AP*, *ITAR*, etc.), *ECU* 'European Currency Unit' (20, 3%), *DIK* a certain kind of investor (30, 13%), etc.

2. Many of the abbreviations are homographs (ignoring case) with normal words. For example, *JE* 'nuclear plant' vs. *je* 'is' (310,000 or 99.8%), *OF* 'a political movement' vs. *of* 'English preposition in institution names, etc.' (1300, 63%), *ME* 'European Cup' vs. *me* 'English or French pronoun in song names, etc.' (about 260 occurrences, or 8% are in lower case); *NATO* vs. *nato* 'with that' (600, 12%), etc. Other examples are *VŠE* 'School of Economics' vs. *vše* 'all', *DNA* vs. *dna* 'bottom', etc.

3. Many non-abbreviations are often written in caps (in titles, for emphasis, typographical conventions, etc.). For example, *PRAHA* 'Prague' (about 7400 occurrences, or 17% are in caps; especially in position when marking the source of a message; similarly other towns), *Jiří* 'a male's name' (1708, 13%; caps are used especially when the name stands as a name for a paragraph – a paragraph about that person; similarly other personal names), *DNES* 'Today – a newspaper' (300, 1%; for some reason that newspaper name is often written in caps), *EKOFLÓRA* 'company name' (3, 100%; for some reason that company writes its name in caps), etc.

And of course these cases are not independent: a non-abbreviation from 2) can be in a title and therefore written in all caps, an abbreviation from 2) can be spelled with a lower case, etc.

3.5.2 Using several heuristics

We decided to identify abbreviations using several non-strict heuristics. These heuristics were found after *quick* inspections of the results obtained by the above simple criterion and of the results of roughly 5 refinements (recall that we intentionally do not use annotated corpora). All-caps abbreviations tend:

1. to occur in all-caps.

2. to be relatively short. Most abbreviations have 2 (*ČR* 'Czech Republic', *OH* 'Olympic Games', *EU*), 3 (*BMW*, *ODS* 'Civic Democratic Party', *ČNB* 'Czech National Bank') or 4 (*OECD*, *ČSSD* 'Czech Socially Democratic Party') characters. Long abbreviations are possible (*UNPROFOR*, *UNICEF*) but very uncommon; moreover, they often decline like normal words (abl. sg. *UNPROFORu* or *UNPROFOR*).

3. to occur not only in all-caps contexts. However, some of them are often accompanied by other abbreviations (*RB – RB OSN* 'Security Council of the United Nations', *ÚV – ÚV KSČ* 'Central Committee of the Communist Party of Czechoslovakia').

4. to contain consonant clusters that would be phonologically impossible under normal pronunciation. Of course, (1) many abbreviations consist of a usual sequence of graphemes (*OS* 'Operating System', *ODA* 'Civic Democratic Alliance'); (2) a text can contain many foreign words that contain 'weird' sequences of consonants (*ss* or *tt* in *Massachusetts*, or *ss* in *Gross*). However, a word that is all-caps and consists exclusively of non-syllabic consonants is nearly 100% an all-cap abbreviation (or a typo).

Naturally, these criteria give very unreliable results when applied to low-frequency tokens.

3.5.3 The algorithm

The algorithm considers all words that ever occur in all caps and collects some basic statistics about them – their frequency, the frequency of their non-all-caps variants, frequency of occurrence in all-caps contexts, their length and whether they have the all-consonant property.

The algorithm has a set of rules each corresponding to one of the above tendencies. Each rule increases or decreases a word's score for various degree of compliance with the tendency. The words are punished for a low frequency or going against the (1)–(3) tendencies; and rewarded for having the all-consonant property. Words with a score above a specified threshold are considered to be all-caps abbreviations. In addition, we ignored common roman numerals (arbitrarily I–XXX).

3.5.4 Evaluation

Training data. The all-caps acquistion algorithm was trained on the non-annotated `raw` corpus described in A.1.

Testing data. We tested the algorithm on the `te` corpus. The corpus is annotated with lemmas, and most abbreviation lemmas are marked by a *:B* code. For example, *BMW* is assigned the lemma *BMW-1_:B_;K* (the company) or *BMW-2_:B_;R* (the car)[9]; *atd* 'etc' is assigned *atd-1_:B*. Some abbreviations are also marked with '8' in the variant slot of their tags.

However, some abbreviations are not marked in either way.[10] For example, *USA* is assigned the lemma *USA_;G* and tag `NNIPX-----A----`. Therefore we decided to identify all-caps abbreviations as words that have all-caps lemmas, are at least two characters long and are not annotated as roman numerals. We manually checked those that do not contain *:B* code in the lemma or '8' in the tag. Out of the 1375 all-caps words in the corpus, we identified 769 as all-caps abbreviations.

We ran the algorithm in three ways:

1. In the first evaluation (the baseline), the algorithm simply checked whether an all-caps word it was among the abbreviations learned from the `raw` corpus or not.

2. In the second evaluation, the algorithm checked whether a word was among the abbreviations learned from the `raw` corpus or not.

3. In the third evaluation, we inspected and corrected the 50 most frequent all-caps words that were accepted as abbreviations and the 50 most frequent all-caps words that we refuted as abbreviations. There were no errors in the accepted list, but there were 6 errors in the refuted list. The errors exemplify the problems of the method and the rules we selected:

   (a) *UNPROFOR* – refuted, because too long (Rule 2). This may be okay, because this is not a classical abbreviation but an acronym, it is also often declined (*UNPROFORu*, *UNPROFOR-u*, *UNPROFORU*).

   (b) *ÚV* 'Central Committee' – refuted because mostly followed by an all-caps word (Rule 3), e.g., *ÚV KSč* 'Central Committee of the Communist Party of Czechoslovakia'.

   (c) *OF* 'Civic Forum' – refuted because often occurs in lower case as English *of* (Rule 1); however nearly none *OF* occurred in all-caps environment.

   (d) *KAN* 'Club of Active Non-partisans' – refuted because often occurs in lower-case as a lower-case abbreviation in names of water works and sewage companies.

[9] *;K* stands for a company, *;R* stands for product. I omit the lemma comments that are in parens and are in Czech. See (Hajic 2004) for more details.

[10] According to (Hajic 2004), eventually all abbreviations should be marked by the '8' in their variant slot.

(e) *RB* 'Security Council' – refuted because mostly followed by all caps *OSN* 'U.N.' (Rule 3).

(f) *OV* 'County Committee' – the same story as *ÚV* in (b).

Experiment	Recall	Precision	F-measure
1 – baseline	100.0%	55.1	71.1
2 – unsupervised	97.3%	91.2	94.1
3 – High frequency hypotheses manually corrected	98.1%	91.4	94.6

Table 11: Evaluation of the abbreviation learner

3.5.5 Possible Enhancements

There are many ways to improve the suggested methods for acquiring a list of abbreviations (and non-abbreviations). It would be worth focusing on the abbreviations that have the same form (ignoring case) as normal words (§3.5.1 item 2). We believe that a more subtle implementation of Rule 2 (§3.5.2) would help – currently only the case of the immediately preceding and following words is considered. Similarly, if a token of an abbreviation candidate occurs several times in a few adjacent sentences or paragraphs, it would be good to compare those tokens: do they all use all caps? (then it is probably an abbreviation), do only some of them use all caps? (then it is probably a normal word in a title or emphasis). It is also possible to use a relatively small annotated tuning corpus to tune the criteria and their parameters.

3.6 Evaluation of the whole system

We evaluated our Morphological Analyzer against the `te` corpus manipulating two parameters:

- Whether a lexicon automatically acquired from the `raw` corpus is used.
- Size of a word list capturing analyses of the most frequent word forms (top forms list, or TFL). The lists were created on the basis of the `raw` corpus.

The results are summarized in Table 12. It is worth repeating that we are concerned only about nouns. The TFLs help without a question – they lower both error-rate (they help with irregular words that are not covered by our paradigms) and ambiguity. The automatic lexicon lowers ambiguity (by pruning incorrect lexical entries), but also increases error-rate (by pruning correct lexical entries). Without TFL, ambiguity decreases by 40% and error rate increases by 38%. With 10K-TFL, ambiguity decreases by 32% and error rate increases by 25%. Depending on what the results will be used for, it may or may not make sense to use an automatic lexicon. The quality of the results is worse than the quality of

Lexicon	–	–	–	+	+	+	Hajič[11]
Top forms list	0K	5K	10K	0K	5K	10K	
Error rate	3.6	2.9	2.7	5.8	3.9	**3.6**	1.3
Ambiguity tag/w	19.6	13.1	11.5	11.7	8.5	**7.8**	3.8
Speed w/s[12]	3000	3500	4800	4500	6500	8200	

Table 12: Evaluation of the Czech morphological analyzer (on nouns)

(Hajic 2004), a system with a large manually created lexicon: Our recall error is roughly three times as large and precision error twice as large.

As mentioned before, the Guesser is relatively slow, therefore using a TFL and/or lexicon increases the speed of analysis.

3.7 Possible enhancements

Currently, the main effort is focused on improving lexicon acquisition: (i) considering frequencies and contexts of word forms when eliminating incorrect hypotheses; (ii) replacing sequential application of heuristics with their weighted parallel combination; (iii) using information about common derivation patterns to extend the algorithm over several lemmas related by derivation and eliminating some of the systematic errors mentioned above. We are also exploring the possibilities of combining our approach with various machine learning techniques. Finally, we are in the process of improving our tools used by native (or informed) speakers to provide the limited amount of information needed by the analyzer in fast and effective way.

4 Application to Russian

To test the portability of our approach to other languages, we created a similar morphological analyzer for Russian (Hana *et al.* 2004), Portuguese (Hana *et al.* 2006) and Catalan (Feldman *et al.* 2006). Here, as an example, we discuss the modification of the system for Russian.

4.1 Russian versus Czech

A detailed comparative analysis of Czech and Russian is beyond the scope of this paper. However, we would like to mention a number of the most important facts. Both languages are Slavic (Czech is West Slavonic, Russian is East Slavonic). Both have extensive inflectional morphology whose role is important in determining the grammatical functions of phrases. In both languages, the main verb agrees in person and number with the subject;

[11] 300K lexicon (Hajič, p.c.)

[12] Running on Sun Java RE 1.5.0.01 with HotSpot, MS Windows XP on Pentium Celeron 2.6 GHz, 750MB RAM. The time need to initialize the system (load and compile lexicons, paradigms etc.) is not included.

adjectives agree in gender, number and case with nouns. Both languages are free constituent order languages. The word order in a sentence is determined mainly by discourse. It turns out that the word order in Czech and Russian is very similar. For instance, old information mostly precedes new information. The "neutral" order in the two languages is Subject-Verb-Object. Here is a parallel Czech-Russian example from our development corpus:[13]

(2) a. [Czech]

Byl	jasný,	studený	dubnový	den
was$_{Masc.Past}$	bright$_{Masc.Sg.Nom}$	cold$_{Masc.Sg.Nom}$	April$_{Masc.Sg.Nom}$	day$_{Masc.Sg.Nom}$

i	hodiny	odbíjely	třináctou.
and	clocks$_{Fem.Pl.Nom}$	stroke$_{Fem.Pl.Past}$	thirteenth$_{Fem.Sg.Acc}$

b. [Russian]

Byl	jasnyj,	xolodnyj	aprel'skij	den'
was$_{Masc.Past}$	bright$_{Masc.Sg.Nom}$	cold$_{Masc.Sg.Nom}$	April$_{Masc.Sg.Nom}$	day$_{Masc.Sg.Nom}$

i	časy	probili	trinadtsat'.
and	clocks$_{Pl.Nom}$	stroke$_{Pl.Past}$	thirteen$_{Acc}$

'It was a bright cold day in April, and the clocks were striking thirteen.' [from Orwell's '1984']

Of course, not all utterances are so similar. However, most of the differences are on syntactic levels and in the level of usage. (Hana *et al.* 2004) and (Hana & Feldman 2004), discuss some ways how to address some of those differences.

On the level of morphology, the languages are very close. The order and function of morphemes are nearly identical. Obviously, there are some differences. For example, Russian does not have vocative; Russian marks reflexivity by a verb suffix, while Czech by a reflexive clitic; Russian verb negation is marked by a separate particle, while Czech verb negation is marked by a prefix; Russian adjectives and participles do not distinguish gender in plural; etc. Naturally, the morphemes have different shapes (and are written in different scripts), but even from this point of view, they are also often similar.

4.2 Data

Tag system. We adapted the Czech tag system (see §A.2). It has about 900 tags which is significantly less than the Czech tagset with about 4300 tags. There are two reasons for this: (i) a theoretical one – some Russian categories have fewer values (case; many Czech morphemes distinguish various levels of colloquiality); (ii) the Czech tag system is very elaborate and specially designed to serve multiple needs, while our tagset is designed solely to capture the core of Russian morphology and demonstrate the portability of our techniques for morphological processing. See (Hana *et al.* 2004) for more details.

[13] All Russian examples in this paper are transcribed in the Roman alphabet. Our system is able to analyze Russian texts in several Cyrillic encodings and various transcriptions.

Testing data. For evaluation purposes, we selected and morphologically annotated (by hand) a small portion from the Russian translation of Orwell's *1984*. This corpus contains 4011 tokens and 1858 distinct forms.

Development data. For development testing, we used another part of *1984*. Since we want to work with minimal language resources, the development corpus is intentionally small – 1788 tokens. We used it to test our hypotheses and tune the parameters of our tools.

Raw corpus. For lexicon acquisition (cf. §3.4), we used a large raw corpus – Uppsala Russian Corpus.[14] The corpus contains about 1M tokens (roughly 35 times smaller than the Czech `raw` corpus).

Future data. In the near future, we would like to increase the size of the testing corpus to roughly 10K tokens. These tokens will come from newspaper texts. We plan to include some newspaper texts into the development data as well, however, we still want to keep the size of the development data very small, probably somewhere around 3K tokens.

4.3 Evaluation

We evaluated our system against the 4K tokens of the testing corpus manipulating two parameters:

- Whether a lexicon automatically acquired from the `raw` corpus is used.
- Whether the longest ending filtering (LEF, see (Hana *et al.* 2004)[15]) is used.

For practical reasons, we did not use the top-frequency lists, although we believe they would help significantly. We plan to employ them in a near future.

The results are summarized in Table 13. Again, we are concerned only with nouns; the results on all tokens are shown only for reader's information and comparison with our previous work (Hana & Feldman 2004; Hana *et al.* 2004). The results are not directly comparable with the results for Czech (§3.6) because of the different nature of the testing corpora (newspapers/magazines for Czech vs. fiction for Russian).

In comparison with (Hana *et al.* 2004), we significantly decreased the recall error (9.6% → 6.0%, i.e., 38% relative reduction; on all tokens with both lexicon and LEF), but at the same time ambiguity increased, though about half the relative size (3.1% → 4.0; i.e.,

[14] The corpus is freely available from Uppsala University at http://www.slaviska.uu.se/ryska/corpus.html

[15] This is a simple heuristic to decrease the number of analyses. The heuristic assumes the correct ending is usually one of the longest candidate endings. A similar approach was used by Mikheev (1997). In English, it would mean that if a word is analyzed either as having a zero ending or an *-ing* ending, we would consider only the latter; obviously, in the vast majority of cases that would be the correct analysis. In addition, we specify that few long but very rare endings should not be included in the maximum length calculation. To stay within the labor-light paradigm, we capture only the few most common systematic errors the LEF does.

	Lexicon	–	+	–	+
	LEF	no	no	yes	yes
All	recall error	**3.6**	5.6	5.7	6.0
	ambiguity (tag/w)	12.9	4.6	9.5	**4.0**
Nouns	recall error	**1.7**	5.3	1.7	5.3
	ambiguity (tag/w)	24.5	6.9	18.1	**5.9**

Table 13: Evaluation of the Russian morphological analyzer

23% relative increase). We believe that for most applications this is an improvement. For example, the tagging error of the (Hana *et al.* 2004) tagger ran on the results of this analyzer decreased from 26.5% to 23% on all tokens, i.e., 13% relative reduction. (Feldman 2006) reports further improvement of the tagger – her error rate is 18.6%.

5 Conclusion

We have shown that a morphological system with a small amount of manually created resources can be successful.

Time needed for adjusting the system to a new (inflectional) language constitutes a fraction of the time needed for systems with extensive manually created resources: days instead of years. Two things are required – a reference grammar (for information about paradigms and closed class words) and a large amount of text (for learning a lexicon; e.g., newspapers from the internet). It is also advisable to have access to a native speaker. First, because reference grammars are often too vague, and second, because a quick glance at results (i.e., at an automatically acquired lexicon or at an analyzed text) can provide feedback leading to a significant increase of accuracy. Also, as Tables 2 and 12 show, providing (manually or semi-automatically) correct analyses for the most frequent words helps a lot. However, all of these require only limited linguistic knowledge.

The quality of the results is worse than that of systems with manual resources (roughly tripling recall error and doubling precision error). However, we believe that the approach still has a large space for improvement and that eventually the results will be very similar. Some of such enhancements were mentioned in §3.7.

In the near future, we plan to compare effectiveness (time and price) of our approach and the standard resource-intensive approach when annotating a medium-size corpus (e.g., 100K tokens). The resource-intensive system has lower ambiguity and error rate and therefore an annotator can work faster (less things to select from, less things to add). On the other hand, creation of such a system is very time consuming.

A Data

A.1 Corpora

During our work on Czech , we used several corpora for various purposes:

- a large raw corpus (`raw`) to train and tune our tools;
- annotated corpora (`te`) to test our tools;
- a small annotated corpus (`tu`) a to tune our tools;
- annotated corpora (`tr`, `tr1`, `tr2`) to report some statistics about Czech texts.

All of these corpora are either part of the Prague Dependency Treebank 1.0 (PDT, Böhmová *et al.* 2001) or are part of the PDT distribution. Let's discuss them in more detail:

- `Raw` consists of all the texts labeled as *Raw texts* in the PDT distribution.[16] The texts come from a Czech daily newspaper Lidové Noviny from the years 1991-1995. It contains over 39M tokens or nearly 2.4M sentences.
- `Te` consists of all the annotated texts labeled as evaluation data in PDT.[17] It consists of about 125K tokens or 8K sentences. The texts come from two daily newspapers, a business weekly and a popular scientific magazine.
- `Tr` consists of all the annotated texts labeled as training data in PDT. It consist of about 1.5M tokens or 95K sentences. The texts come from the same sources as the `Te` texts. To allow evaluation of how particular statistics transfers from one corpus to another, we split the corpus into two parts, each with about 620K tokens.[18] These smaller corpora are referred to as `tr1` and `tr2`. It is worth noting that we analyzed the `tr` corpus for the sake of this paper, and only after finishing our work on the tools. That means we did not use a source that would not be available for some other Slavic languages. The results are reported in §2.
- `Tu`. We need some annotated data to tune the parameters of our modules. The data should (1) be as close as possible to the data that would be obtained by morphologically annotating a large corpus; but (2) they should be also labor cheap. We decided to manually annotate a very small amount of data reflecting frequency of words in a corpus.

  From the `raw` corpus, we extracted word forms[19] and their frequencies. We split these words into groups by their frequency percentiles. From each of these groups

[16] See http://ufal.mff.cuni.cz/pdt/Corpora/Raw_Texts/index.html for more details.

[17] See http://ufal.mff.cuni.cz/pdt/Corpora/PDT_1.0/Doc/PDT10_data.html for more details.

[18] The remaining tokens are not used in this paper. PDT is organized by sources and date of publication. To prevent differences between the two corpora caused by such organization, we split the corpus into 40 pieces and put all the odd pieces into `tr1` and all the even pieces `tr2`.

[19] We ignore capitalization, forms differing only in capitalization are considered to be the same forms. It would be hard and probably unnecessary to address this issue properly. We also exclude all forms containing digits.

we randomly selected 10 noun forms. To favor more frequent words, i.e., words with a greater impact, the groups are between the following percentiles 0, 50, 75, 90, 95, 99, 100.[20] We morphologically annotated these forms. This is done semi-automatically – we select (in few cases add) the right analyses from the output of our Guesser. Naturally, we do not consider the context of those forms – so for example *hrad* 'castle' would be annotated as both nominative and accusative.

The number of occurrences of each form in the tuning corpus is the same as its frequency in the `raw` corpus.

A.2 Tagset

We used the Czech tag system employed in PDT (Hajic 2000). Every tag is represented as a string of 15 symbols each corresponding to one morphological category. We refer to the positions in such a string as slots. Two slots are not used (13,14); the slot 2 (detailed POS) uniquely determines the slot 1 (POS). For example, the word *vidělo* 'saw$_{\texttt{neut.sg}}$' is assigned the tag `VpNS---XR-AA---` because it is a verb (`V`), past participle (`p`), neuter (`N`), singular (`S`), does not distinguish case (`-`), possessive gender (`-`), possessive number (`-`), can be any person (`X`), is past tense (`R`), not gradable (`-`), affirmative (`A`), active voice (`A`), and is the basic stylistic variant (the final hyphen).

No.	Description	Abbr.	No. of values
1	POS	p	12
2	SubPOS – detailed POS	s	75
3	Gender	g	11
4	Number	n	6
5	Case	c	9
6	Possessor's Gender	f	5
7	Possessor's Number	m	3
8	Person	e	5
9	Tense	t	5
10	Degree of comparison	d	4
11	Negation	a	3
12	Voice	v	3
13	Unused		1
14	Unused		1
15	Variant, Style	i	10

Table 14: Overview of the Czech positional tagset

The tagset uses about 4300 tags. Thus, it is much larger than the Penn Treebank tagset, which uses only 36 non-punctuation tags (Marcus *et al.* 1993). It is also larger than

[20] We ignore the words with only one token (i.e., most of the forms below the median) because it would be hard to sort out spelling errors, etc.

case	
1	nominative
2	genitive
3	dative
4	accusative
5	vocative
6	locative
7	instrumental

number	
S	singular
P	plural
gender	
M	masculine animate
I	masculine inanimate
F	feminine
N	neuter

Table 15: Explanation of glosses.

the tagset we developed on its basis for Russian, a language similar to Czech – about 900 tags (see also Hana *et al.* 2004:§4.2).

A.3 Morphological Glosses

The morphological glosses in this paper are based on the tagset above. The noun glosses have the structure gender-number-case, for possible values see Table 15. For example, `FS2` stands for feminine singular genitive. When not relevant or obvious from the context we leave some of the slots out, e.g., `S2` – singular genitive. If a word is ambiguous we separate glosses by slashes. If case is ambiguous we simply list all the relevant case numbers, e.g., `S14` singular nominative or accusative.

B Czech

The Czech language is one of the West Slavic languages. It is spoken by 10+ million speakers mostly in Czechia. In this section, we discuss properties of morphology and syntax of the language relevant to our work. For a more detailed discussion, see for example (Karlík *et al.* 1996; Fronek 1999; Petr 1987).[21]

B.1 Morphology

Like other Slavic languages, Czech is a richly inflected language. The morphology is important in determining the grammatical functions of phrases. The inflectional morphemes are highly ambiguous (see Table 5). There are three genders: neuter, feminine and masculine. The masculine gender further distinguishes the subcategory of animacy. Sometimes, it is assumed that there are four genders: neuter, feminine, masc. animate and masc. inanimate; we follow that practice. In addition to singular and plural, some dual number forms survive in body part nouns and modifiers agreeing with them.[22] There are seven cases:

[21] Alas, there is no recent detailed Czech grammar in English. The dictionary (Fronek 1999) provides a basic overview. A short overview is also in an appendix of (Hana 2007).

[22] In colloquial Czech, there is no dual. The colloquial plural forms are the same as the official dual forms. For example, official: *velkýma rukama* 'big$_{\texttt{FD7}}$ hands$_{\texttt{FD7}}$' vs. *velkými lžícemi* 'big$_{\texttt{FP7}}$ spoons$_{\texttt{FP7}}$' (there is

	N Monday	F song	F fly	M Jirka	M brother	I castle
S1	pondělí	píseň	moucha	Jirka	bratr	hrad
S2	pondělí	písně	mouchy	Jirky	bratra	hradu
S3	pondělí	písni	mouše	Jirkovi	bratru/ovi	hradu
S4	pondělí	píseň	mouchu	Jirku	bratra	hrad
S5	pondělí	písni	moucho	Jirko	bratře	hrade
S6	pondělí	písni	mouše	Jirkovi	bratru/ovi	hradu
S7	pondělím	písní	mouchou	Jirkou	bratrem	hradem
P1	pondělí	písně	mouchy	Jirkové	bratři/ové	hrady
P2	pondělí	písní	much	Jirků	bratrů	hradů
P3	pondělí	písním	mouchám	Jirkům	bratrům	hradům
P4	pondělí	písně	mouchy	Jirky	bratry	hrady
P5	pondělí	písně	mouchy	Jirkové	bratři	hrady
P6	pondělích	písních	mouchách	Jircích	bratřích	hradech
P7	pondělími	písněmi	mouchami	Jirky	bratry	hrady

Table 16: Examples of declined nouns.

nominative (1), genitive (2), dative (3), accusative (4), vocative (5), locative (6), instrumental (7). It is a common practice to refer to the cases by numbers. Only nouns, only in singular, and only about half of the paradigms have a special form for vocative, otherwise the vocative form is the same as nominative.

There is a significant difference in morphology and lexicon between the official and colloquial levels of Czech. The official variant is a semi-artificial language. Sometimes it is claimed, with some exaggeration, that the official Czech is the first foreign language Czechs learn. Since we analyze written texts where the official language is predominant we largely ignore the colloquial language and its forms here.

Nouns. Traditionally, there are 13 basic noun paradigms – 4 neuter, 3 feminine, 4 animate and 2 inanimate; plus there are nouns with adjectival declension (another 2 paradigms). In addition, there are many subparadigms and subsubparadigms. All of this involves a great amount of irregularity and variation. As an illustration, Table 16 shows the declension of a few nouns. For discussion on noun paradigms see §3.3.

Adjectives. Adjectives follow two paradigms: *hard* and *soft*. Both of them are highly ambiguous, filling the 60 (4 genders $\times$ 2 numbers $\times$ 7 cases + 4) non-negated first grade categories with only 12, resp. 8 forms. See Table 17.[23]

no 'hands$_{\texttt{FP7}}$' or spoons$_{\texttt{FD7}}$); colloquial: *velkýma rukama* 'big$_{\texttt{FP7}}$ hands$_{\texttt{FP7}}$' vs. *velkýma lžícema* 'big$_{\texttt{FP7}}$ spoons$_{\texttt{FP7}}$'.

[23] In colloquial Czech, the hard declension is slightly different: in endings *ý* → *ej*, *é/í* → *ý*. Moreover, neuter in plural uses feminine forms, which in turn can be the same as masculine forms. There is no dual, and the inst. pl. has the same ending as the official dual.

<table>
<tr><th></th><th>M</th><th>I</th><th>N</th><th>F</th><th>M</th><th>I</th><th>N</th><th>F</th></tr>
<tr><td>S1</td><td colspan="2">mladý</td><td>mladé</td><td>mladá</td><td colspan="4">jarní</td></tr>
<tr><td>S2</td><td colspan="3">mladého</td><td>mladé</td><td colspan="3">jarního</td><td>jarní</td></tr>
<tr><td>S3</td><td colspan="3">mladému</td><td>mladé</td><td colspan="3">jarnímu</td><td>jarní</td></tr>
<tr><td>S4</td><td>mladého</td><td>mladý</td><td>mladé</td><td>mladou</td><td>jarního</td><td colspan="3">jarní</td></tr>
<tr><td>S5</td><td colspan="2">mladý</td><td>mladé</td><td>mladá</td><td colspan="4">jarní</td></tr>
<tr><td>S6</td><td colspan="3">mladém</td><td>mladé</td><td colspan="3">jarním</td><td>jarní</td></tr>
<tr><td>S7</td><td colspan="3">mladým</td><td>mladou</td><td colspan="3">jarním</td><td>jarní</td></tr>
<tr><td>P1</td><td>mladí</td><td>mladé</td><td>mladá</td><td>mladé</td><td colspan="4">jarní</td></tr>
<tr><td>P2</td><td colspan="4">mladých</td><td colspan="4">jarních</td></tr>
<tr><td>P3</td><td colspan="4">mladým</td><td colspan="4">jarním</td></tr>
<tr><td>P4</td><td colspan="2">mladé</td><td>mladá</td><td>mladé</td><td colspan="4">jarní</td></tr>
<tr><td>P5</td><td>mladí</td><td>mladý</td><td>mladá</td><td>mladé</td><td colspan="4">jarní</td></tr>
<tr><td>P6</td><td colspan="4">mladých</td><td colspan="4">jarních</td></tr>
<tr><td>P7</td><td colspan="4">mladými</td><td colspan="4">jarními</td></tr>
<tr><td>D7</td><td colspan="4">mladýma</td><td colspan="4">jarníma</td></tr>
</table>

Table 17: Adjectival paradigms.

Negation and comparison forms are expressed morphologically. Negation by the prefix *ne-*, comparative by the suffix *-(e)jší-* and superlative by adding the prefix *nej-* to the comparative. The comparative and superlative forms are declined as soft adjectives.

Pronouns. Pronouns have either a noun or adjectival declension.

Numerals. Only *jeden* '1', *dva* '2', *tři* '3', and *čtyři* '4' fully decline, all of them distinguishing case and *jeden* and *dva* also gender. The inflection of the other cardinal numerals is limited to distinguishing oblique and non-oblique forms. Numerals expressing hundreds and thousands have in certain categories a choice between an undeclined numeral form or a declined noun form (*sto dvaceti*, *sta dvaceti* '120.genitive'). Ordinal complex numerals have all parts in the ordinal form[24] and fully declining (*dvacátý pátý* '25th'). Two-digit numerals may have an inverted one-word form (*pětadvacátý* '25th', lit: five-and-twentieth).

Verbs. Verbs distinguish three tenses. Only the forms of the present tense are marked inflectionally, distinguishing number and person. Inflection is also used to mark infinitive, past participles, passive participles and imperatives.

As in all Slavic languages, verbs also distinguish aspect – perfective and imperfective. Aspect is usually marked by prefixes, sometimes suffixes or by suppletion. Change of aspect is usually accompanied by a change, often subtle, in lexical meaning. For exam-

[24] Again, this is the case of the official language, complex numerals in colloquial Czech usually have only their tens and units in ordinal forms.

ple, *psát* ‘write$_{\texttt{imp}}$’, *napsat* ‘write$_{\texttt{perf}}$’, *dopsat* ‘finish writing$_{\texttt{perf}}$’, *sepsat* ‘write up$_{\texttt{perf}}$’, *sepisovat* ‘write up$_{\texttt{imp}}$’, etc.

Five main conjugational types are recognized. They are discriminated on the basis of the third person singular endings: (1) *-e*; (2) *-n-e*; (3) *-j-e*; (4) *-í*; (5) *-á*. Each class has several, quite similar, paradigms (6, 3, 2, 3, 1; 15 in total).

Certain categories are expressed analytically; various forms of the verb *být* serve as the auxiliary:

- past tense: present tense aux + past participle; auxiliary is omitted in 3rd person. E.g., *psal jsem* ‘I wrote/was writing$_{\texttt{masc}}$’[25]
- future tense: future aux + infinitive. E.g., *budu psát* ‘I will write’
- passive: present tense aux + pass. participle. E.g., *jsem obdivován* ‘I am adored$_{\texttt{masc}}$’
- conditional: conditional aux + past participle. E.g., *psala bych* ‘I would write$_{\texttt{fem}}$’
- past conditional: conditional aux + aux in past participle + past participle. E.g., *byla bych psala* ‘I would have written$_{\texttt{fem}}$’

B.2 Syntax

Word order. Czech, like most other Slavic languages, has an exceptionally free word order. Unlike English, word order in Czech is used to express topic-focus structure (cf. Sgall *et al.* 1986) and definiteness. Thus for example, the words in sentence (3) can be rearranged in all 24 possible ways. Each of the sentences has a different topic-focus structure, but all of them are grammatically correct. Prototypically, old information precedes new information.

(3) Včera Petr viděl Marii.
yesterday Peter$_1$ saw Mary$_4$
‘Yesterday, Peter saw Mary.’

More precisely, Czech word order is very free as regards the possibility of moving entire phrases; virtually any scrambling is possible. However, scrambling resulting in discontinuous phrases is much less common. It is limited to certain syntactic constructions, to many sentences involving a contrastive theme and to many sentences involving clitics (see, for example, Hana 2007.)

Agreement. Finite verbs agree with their subjects in number and person, participles in gender and number. Attributes agree with the nouns they modify in case, number and gender.

[25]The different position of the auxiliaries in these examples is due to the fact that some are clitics and some not.

Numeral expressions. Numerals expressions with *jeden* '1', *dva* '2', *tři* '3', *čtyři*, '4', *oba* 'both' behave in a "normal" way: a numeral agrees with its noun in case; *jeden*, *dva* and *oba* also in gender. However, numerals *pět* '5' and above in nominative or accusative positions are followed by nouns in genitive plural. The other cases behave the usual way.

Negation. Sentence negation in Czech is formed by the prefix *ne-* attached to the verb. As in the other Slavic languages, multiple negation is the rule, and negative subject or object pronouns, adjectival pronouns and adverbs combine with negative verbs.

(4) Nikdy nikomu nic neslibuj.
never to-nobody$_3$ nothing$_4$ not-promise$_{\texttt{imper}}$
'Never promise anything to anybody.'

References

BÖHMOVÁ, ALENA, JAN HAJIC, EVA HAJIČOVÁ, & BARBORA HLADKÁ. 2001. The Prague Dependency Treebank: Three-Level Annotation Scenario. In *Treebanks: Building and Using Syntactically Annotated Corpora*, ed. by Anne Abeillé. Kluwer Academic Publishers.

CHANG, JEFFREY T., HINRICH SCHÜTZE, & RUSS B. ALTMAN. 2002. Creating an Online Dictionary of Abbreviations from MEDLINE. *JAMIA* .

FELDMAN, ANNA. 2006. *Portable Language Technology: A Resource-light Approach to Morpho-syntactic Tagging*. The Ohio State University dissertation.

——, JIRKA HANA, & CHRIS BREW. 2006. A cross-language approach to rapid creation of new morpho-syntactically annotated resources. In *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006). Genoa, Italy*.

FRONEK, JOSEF. 1999. *English-Czech/Czech-English Dictionary*. Praha: Leda. Contains an overview of Czech grammar.

GOLDSMITH, JOHN. 2001. Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics* 27.153–198.

HAJIC, JAN. 2000. Morphological Tagging: Data vs. Dictionaries. In *Proceedings of ANLP-NAACL Conference*, 94–101, Seattle, Washington, USA.

——. 2004. *Disambiguation of Rich Inflection: Computational Morphology of Czech*. Praha: Karolinum, Charles University Press. In press.

HANA, JIRI. 2007. *Czech Clitics in Higher Order Grammar*. The Ohio State University dissertation.

——, & ANNA FELDMAN. 2004. Portable language technology: The case of Czech and Russian. In *Proceedings from the Midwest Computational Linguistics Colloquium, June 25-26, 2004*, Bloomington, Indiana.

——, ——, & CHRIS BREW. 2004. A resource-light approach to Russian morphology: Tagging Russian using Czech resources. In *Proceedings of EMNLP 2004*, ed. by Dekang Lin & Dekai Wu, 222–229, Barcelona, Spain. Association for Computational Linguistics.

HANA, JIRKA, ANNA FELDMAN, LUIZ AMARAL, & CHRIS BREW. 2006. Tagging Portuguese with a Spanish tagger using cognates. In *Proceedings of the Workshop on Cross-language Knowledge Induction, 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006), Trento, Italy*.

HLAVÁČOVÁ, JAROSLAVA. 2001. Morphological Guesser or Czech Words. In *Text, Speech and Dialogue*, ed. by V. Matoušek, Lecture Notes in Computer Science, 70–75. Berlin: Springer-Verlag.

KARLÍK, PETR, MAREK NEKULA, & Z. RUSÍNOVÁ. 1996. *Příruční mluvnice češtiny [Concise Grammar of Czech]*. Praha: Nakladatelství Lidové Noviny.

KOSKENNIEMI, KIMMO. 1983. Two-level model for morphological analysis. In *IJCAI-83*, 683–685, Karlsruhe, Germany.

——. 1984. A general computational model for word-form recognition and production. In *COLING-84*, 178–181, Stanford University, California, USA. Association for Computational Linguistics.

LARKEY, LEAH S., PAUL OGILVIE, M. ANDREW PRICE, & BRENDEN TAMILIO. 2000. Acrophile: an automated acronym extractor and server. In *DL '00: Proceedings of the fifth ACM conference on Digital libraries*, 205–214. ACM Press.

MARCUS, MITCHELL, BEATRICE SANTORINI, & MARY ANN MARCINKIEWICZ. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19.313–330.

MIKHEEV, ANDREI. 1997. Automatic rule induction for unknown word guessing. *Computational Linguistics* 23.405–423.

——, & LIUBOV LIUBUSHKINA. 1995. Russian Morphology: An Engineering Approach. *Natural Language Engineering* 3.235–260.

PETR, JAN. 1987. *Mluvnice češtiny [Czech Grammar]*. Praha: Academia.

SGALL, PETR, EVA HAJIČOVÁ, & JARMILA PANEVOVÁ. 1986. *The Meaning of the Sentence and Its Semantic and Pragmatic Aspects*. Prague, Czech Republic/Dordrecht, Netherlands: Academia/Reidel Publishing Company.

SKOUMALOVÁ, HANA. 1997. A Czech morphological lexicon. In *Proceedings of the Third Meeting of the ACL Special Interest Group in Computational Phonology, Madrid*, 41–47. ACL.

YEATES, STUART. 1999. Automatic extraction of acronyms from text. In *New Zealand Computer Science Research Students' Conference*, 117–124.

——, DAVID BAINBRIDGE, & IAN H. WITTEN. 2000. Using compression to identify acronyms in text. In *DCC '00: Proceedings of the Conference on Data Compression*, p. 582. IEEE Computer Society.

ZIPF, GEORGE K. 1935. *The Psychobiology of Language*. Houghton-Mifflin.

—— 1949. *Human Behavior and the Principle of Least-Effort*. Addison-Wesley.

MORPHOLOGICAL COMPLEXITY OUTSIDE OF UNIVERSAL GRAMMAR[1]

Jirka Hana & Peter W. Culicover
The Ohio State University

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]We thank Chris Brew, Beth Hume, Brian Joseph, John Nerbonne and 3 anonymous reviewers from the Cognitive Science Journal for valuable feedback on various versions of this paper. We also thank to Mary Beckman and Shari Speer.

Jirka Hana & Peter W. Culicover. Morphological Complexity Outside of Universal Grammar. *OSUWPL*, Volume 58, Spring 2008, pp. 85–109.

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 assume 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 assume 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**.[2] 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 variant of *s* added to the end of the stem.[3]

(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

[2]The 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.

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

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

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 & 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 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 *et al.* (1985); Hawkins & 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.[4]

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 1988:310 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 & 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 & 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 several 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 develop an alternative measure that we believe addresses all of these issues.

[4]For 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.

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 explanation 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. 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) (Dutch)
- word-final distortions often go undetected, cf. (Marslen-Wilson & Welsh 1978; Cole 1973; Cole & Jakimik 1978; Cole & Jakimik 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 & 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 *et al*. 1993; Marslen-Wilson 1993) show, changes involving non-adjacent segments are generally more disruptive to word recognition than changes involving adjacent segments.

3.2 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. The results in (Hay *et al*. 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.[5]

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 with the same type of objects, while *eat* and *drink* occur with a different type of objects). Similarly, words belonging to the same morphological category often share certain semantic features (e.g., referring to multiple entities). Note however, that the opposite implication is not true: two words sharing some semantic properties, and occurring in similar contexts do not necessary have to be inflectionally related (cf. *walk* and *run*).

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

[5] The 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*.

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.

3.3 Computational acquisition of paradigms

Several algorithms exploit the fact that forms of the same lemma[6] 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 iteratively trained probabilistic grammar generating the word forms. Similarly Baroni *et al.* (2002) successfully 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 & 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 & Wicentowski (2000) acquire morphology of English irregular verbs by comparing the distributions of their forms with regular verbs, assuming they are distributed equally.[7] They also note that forms of the same lemma

[6]The term lemma is used with several different meanings. In our usage, every set of forms belonging to the same *inflectional* paradigm is assigned a lemma, a particular form chosen by convention (e.g., nominative singular for nouns, infinitive for verbs) to represent that set. The terms *citation form*, *canonical form* are used with the same meaning. For example, the forms *break*, *breaks*, *broke*, *broken*, *breaking* have the same lemma *break*. Note that in this usage, only forms related by inflection share the same lemma, thus for example, the noun *songs* and the verb *sings* do not have the same lemmas.

[7]Obviously, 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,

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

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 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. For now, 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.

proper names and uncountable substances would probably belong to different distributions.

The algorithm in Yarowsky & 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.

[8] A 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.

[9] Of 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.

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 to language in context. If they have limited access to context, their language development is very different, as Peters & 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 $\mathrm{prdgm}(w)$ is a set of words belonging to the same paradigm as the word w, then we can define PSI as:

(6) $\mathrm{PSI}(L) = \mathrm{avg}\{\mathrm{ipd}(w)/\mathrm{epd}(w) \mid w \in L\}$

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

(7) $\mathrm{epd}(w) = \mathrm{avg}\{\mathrm{ed}(w,u) \mid u \in L\}$

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

(8) $\mathrm{ipd}(w) = \mathrm{avg}\{\mathrm{ed}(w,u) \mid u \in \mathrm{prdgm}(w)\}$

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 1983).

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 :: String, String -> Integer
   | [], []      = 0
   | u,  []      = length u                        // DELETE u
   | [], v       = length v                        // INSERT v
   | u:us, v:vs = min [                           // the minimum of
      (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: Edit Distance Algorithm of Model 0 (Levenshtein)

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 DELETEing 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 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 1983). 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, while 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 (9), differing only in the position of the affix.

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

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

[10] A 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.

	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 1: Sample prefixing and suffixing languages

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 2: 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 3: Comparing prefixed and suffixed words in Model 0

```
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 [                 // one of:
       ed(us, vs),                             //     SUBSTITUTE
       ed(us, v:vs),                           //     DELETE
       ed(u:us, vs) ]                          //     INSERT
```

Figure 2: Edit Distance Algorithm of Model 1

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, $\mathrm{w_p} \in L_P$ and $\mathrm{w_s}$ the analogous word $\in L_S$.[11] Obviously, it is more probable that $\mathrm{ipd}(w_p) \geq \mathrm{ipd}(w_s)$ than not. Asymptotically, for infinite languages, the $\mathrm{epd}(w_p) = \mathrm{epd}(w_s)$. Therefore, for such languages $\mathrm{PSI}(L_P) > \mathrm{PSI}(L_S)$. 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 (6). 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 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

[11] If 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.

[12] Note 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.

worst sequences for L_P contain 6 DELETEs and 6 INSERTs; the cost is 12.

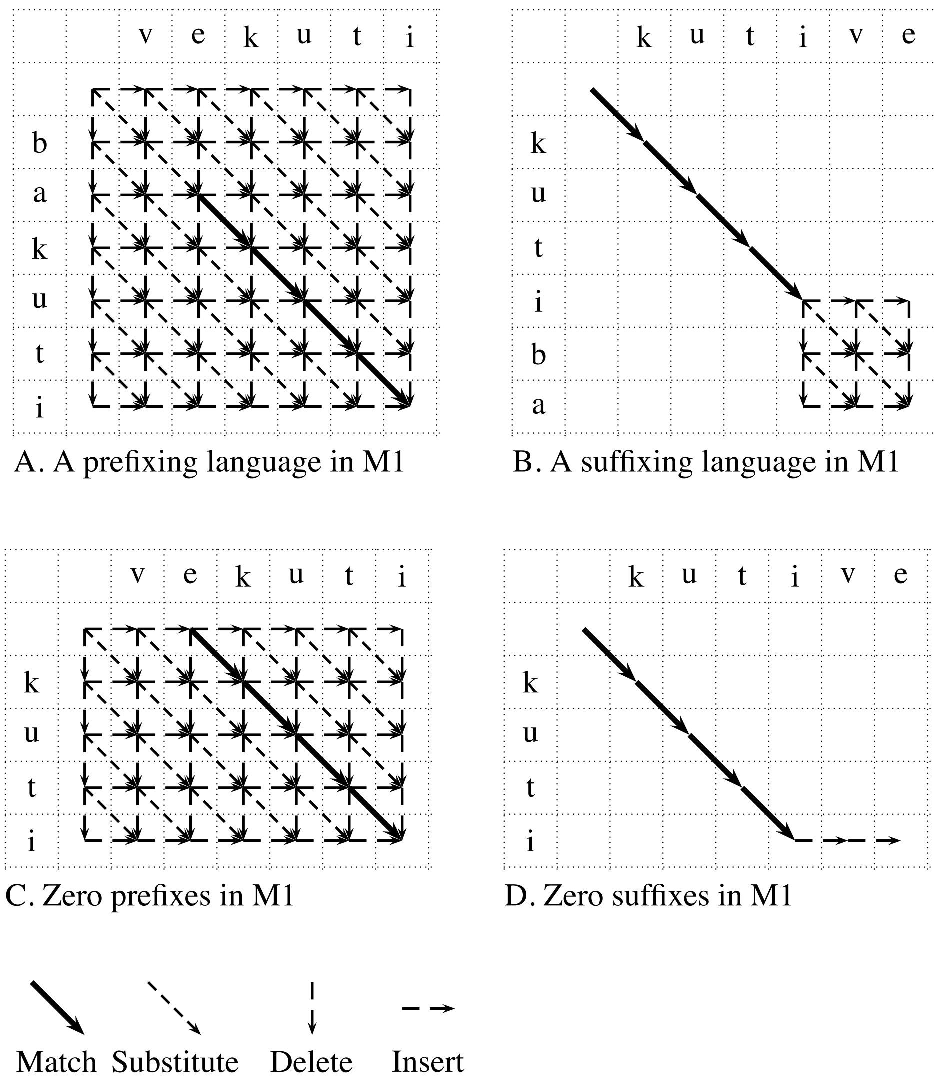


Figure 3: Comparing words in in Model 1

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 – 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]

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

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

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 4: Experiment: Parameters

mean	1.29
standard deviation	0.17
Q1	1.16
median	1.27
Q3	1.33

Table 5: Experiment: Results

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:

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

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

We generated 100 such pairs of languages with the parameters summarized in Table 4, calculating statistics for sufPref. 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 §B.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 §B.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* [bɛntʃ] / *benches* [bɛntʃɪs]. 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. Several studies (Johnson & Jusczyk 2001; Saffran *et al.* 1996) show that the degree of syllable predictability is one of the cues used in word segmentation. 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 syllables[15] 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 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.

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

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) (see §3.3).

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 Morphology acquisition by 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 accord 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.[16] In a similar experiment Gasser & 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).

B Templatic morphology, Metathesis

B.1 Templatic morphology

In templatic 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.

(11)	lomed	'learn$_{\texttt{masc}}$'	shatak	'be-quiet$_{\texttt{pres.masc}}$'
	lamad	'learnt$_{\texttt{masc.sg.3rd}}$'	shatak	'was-quiet$_{\texttt{masc.sg.3rd}}$'
	limed	'taught$_{\texttt{masc.sg.3rd}}$'	shitek	'made-sb-to-be-quiet$_{\texttt{masc.sg.3rd}}$'
	lumad	'was-taught$_{\texttt{masc.sg.3rd}}$'	shutak	'was-made-to-be-quiet$_{\texttt{masc.sg.3rd}}$'

Phonological alternations are possible – e.g., stops alternating with fricatives. 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 perceptionally in different "dimensions" and therefore clearly separable, it would not be more complicated than morphology using continuous affixes

[16] The 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%).

or suprasegmentals. Fowler (1983) convincingly argues on phonetic grounds for such assumption – consonants are modulated on the basic vowel contour of syllables.

Ravid's (2003) study also suggests that template morphology is not more difficult to acquire than a concatenative one. She finds that in case of forms alternatively produced by template and concatenative processes, children tend to acquire the template option first. She 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.

B.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 phase[17] by Churchward (1940), and in many cases these are distinguished by metathesis (examples due to Hoeksema & Janda (1988:228)):[18]

	Complete phase	Incomplete phase		
	aírɛ	ai̯ɛ́r	‘fish	
(12)	púrɛ	pu̯ɛ́r	‘rule, decide’	(Rotuman)
	tíkɔ	ti̯ɔ́k	‘flesh’	
	sɛ́ma	sɛ̯́ám	‘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 & 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 metathesis can be usually analyzed as two separate metathesis, each motivated by an independent phonological constraint.

Processing Metathesis. Mielke & 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)

[17] According to Hoeksema & 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.

[18] In 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.

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, MARCO. 2000. *Distributional cues in morpheme discovery: A computational model and empirical evidence*. UCLA dissertation.

——, JOHANNES MATIASEK, & HARALD TROST. 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*, 48–57.

BECKER, THOMAS. 2000. Metathesis. In *Morphology: A Handbook on Inflection and Word Formation*, ed. by Geert Booij, Christian Lehmann, & Joachim Mugdan, 576–581. Berlin: Mouton de Gruyter.

CHOMSKY, NOAM. 1973. Conditions on Transformations. In *A Festschrift for Morris Halle*, ed. by S. Anderson & P. Kiparsky. New York: Holt, Rinehart and Winston.

CHURCHWARD, C. MAXWELL. 1940. *Rotuman Grammar and Dictionary*. Sydney: Australasia Medical Publishing Co. Reprint New York:AMS Press, 1978.

COLE, RONALD ALLAN. 1973. Listening for mispronunciations: A measure of what we hear during speech. *Perception and Psychophysics* 153–156.

——, & JAKIMIK. 1980. How are syllables used to recognize words? *Journal of the Acoustical Society of America* 965–970.

——, & J. JAKIMIK. 1978. Understanding speech: How words are heard. In *Information Processing Strategies*, ed. by G. Underwood. London: Academic Press.

CONNINE, C. M., D. G. BLASKO, & D. TITONE. 1993. Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language* 193–210.

COOPER, W., & J. PACCIA-COOPER. 1980. *Syntax and Speech*. Cambridge, MA: Harvard University Press.

CULICOVER, PETER W., & ANDRZEJ NOWAK. 2002. Markedness, antisymmetry and complexity of constructions. In *Variation Yearbook. Vol. 2 (2002)*, ed. by Pierre Pica & Johann Rooryk, 5–30. Amsterdam: John Benjamins.

CUTLER, A., J.A. HAWKINS, & G. GILLIGAN. 1985. The suffixing preference: a processing explanation. *Linguistics* 23.723–758.

FOWLER, CAROL 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, MICHAEL. 1994. Acquiring receptive morphology: a connectionist model. In *Annual Meeting of the Association for Computational Linguistics, 32*, 279–286.

——, & CHAN-DO LEE. 1991. A short-term memory architecture for the learning of morphophonemic rules. In *Advances in Neural Information Processing Systems 3*, ed. by R. P. Lippmann, J. E. Moody, & D. S. Touretzky, 605–611. San Mateo, CA: Morgan Kaufmann.

GIVÓN, TALMY. 1979. *On Understanding Grammar*. New York: Academic Press.

GREENBERG, JOSEPH H. 1957. *Essays in Linguistics*. Chicago: University of Chicago Press.

HALL, CHRISTOPHER J. 1988. Integrating Diachronic and Processing Principles in Explaining the Suffixing Preference. In *Explaining Language Universals*, ed. by J. A. Hawkins, chapter 12.

HARRIS, ZELLIG SABBATAI. 1955. From phoneme to morpheme. *Language* 31.190–222.

HAWKINS, JOHN A., & GARY GILLIGAN. 1988. Prefixing and suffixing universals in relation to basic word order. *Lingua* 74.219–259.

HAY, JENNIFER, JANET PIERREHUMBERT, & MARY BECKMAN. 2003. Speech Perception, Well-formedness and the Statistics of the Lexicon. In *Papers in Laboratory Phonology VI*, ed. by John Local, Richard Ogden, & Ros Temple. CUP.

HOEKSEMA, JACK, & RICHARD D. JANDA. 1988. Implications of Process-morphology for Categorial Grammar. In *Categorial Grammars and natural Language Structrues*, ed. by Richard T. Oehrle, Emmon Bach, & Deirdre Wheeler, 199–247. Academic Press.

JACQUEMIN, CHRISTIAN. 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*, 156–165. ACM.

JANDA, RICHARD D. 1984. Why Morphological Metathesis Rules Are Rare: On the Possibility of Historical Explanation in Linguistics. In *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, ed. by Claudia Brugman & Monica Macaulay, 87–103, Berkeley, CA. Berkeley Linguistics Society.

JOHNSON, ELIZABETH K., & PETER W. JUSCZYK. 2001. Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language* 44.548–567.

LEVENSHTEIN. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory* 10.707–710.

MARSLEN-WILSON, WILLIAM D. 1993. Issues of process and representation in lexical access. In *Cognitive Models of Language Processes: The Second Sperlonga Meeting*, ed. by G. Altmann & R. Shillcock, Hove. Lawrence Erlbaum Associates.

——, & LORRAINE KOMISARJEVSKY TYLER. 1980. The temporal structure of spoken language understanding. *Cognition* 1–71.

——, & A. WELSH. 1978. Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology* 29–63.

MCCARTHY, JOHN. 2000. The prosody of phase in Rotuman. *Natural language and linguistic theory* 18.147–197.

MIELKE, JEFF, & ELIZABETH HUME. 2001. Consequences of Word recognition for Metathesis. In *Surface Syllable Structure and Segment Sequencing*, ed. by Elizabeth Hume, N. Smith, & J. van de Weijer. Leiden: HIL.

NOOTEBOOM, SIEB. 1981. Lexical retrieval from fragments of spoken words: beginnings versus endings. *Journal of Phonetics* 9.407–424.

PETERS, ANN M., & LISE MENN. 1993. False starts and filler syllables: ways to learn grammatical morphemes. *Language* 69.742–777.

PLUNKETT, K., & V. MARCHMAN. 1991. U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition* 38.73–193.

PRASADA, S., & S. PINKER. 1993. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8.1–56.

RAVID, DORIT. 2003. A developmental perspective on root perception in Hebrew and Palestinian Arabic. In *The processing and acquisition of root-based morphology.*, ed. by Y. Shimron. Amsterdam: Benjamins.

RUMELHART, D., & J. MCCLELLAND. 1986. On learning the past tense of english verbs. In *Parallel Distributed Processing*, ed. by D. Rumelhart & J. McClelland, volume II. Cambridge: MIT Press.

SAFFRAN, J.R., R.N. ASLIN, & E.L. NEWPORT. 1996. *Science* 274.1926–1928.

SANKOFF, DAVID, & JOSEPH B. KRUSKAL (eds.) 1983. *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*. The David Hume Series. CSLI. Republished in 1999.

SAPIR, EDWARD. 1921. *Language, an introduction to the study of speech*. New York: Harcourt.

SPENCER, ANDREW, & ARNOLD M. ZWICKY (eds.) 1998. *The Handbook of Morphology*. Oxford, UK and Malden, MA: Blackwell.

WEXLER, KENNETH, & PETER W. CULICOVER. 1980. *Formal principles of language acquisition*. Cambridge, Mass: MIT Press.

YAROWSKY, DAVID, & RICHARD WICENTOWSKI. 2000. Minimally supervised morphological analysis by multimodal alignment. In *Proceedings of the 38th Meeting of the Association for Computational Linguistics*, 207–216.

DISCOURSE CONSTRAINTS ON EXTRAPOSITION FROM DEFINITE NP SUBJECTS IN ENGLISH

Laurie A. Maynell
The Ohio State University

Abstract

This paper examines English restrictive relative clauses that are extraposed from definite NP subjects, and their relationship to the discourse context in which they may be uttered. In contrast to previous work on this topic (Huck & Na, 1990, 1992), I demonstrate that extraposed relative clauses need not contain information that is given with respect to discourse context. Rather, extraposed clauses may contain either discourse-given information or discourse-new information. What is critical for extraposition of relative clauses from definite NP subjects is how informative the relative clauses are with respect to the Question Under Discussion as defined by Roberts (1996). In that sense, these extraposed relatives must provide new information with respect to a localized portion of the discourse content, and not with respect to the discourse as a whole. Thus, the acceptability of this particular structure depends not on a syntactic configuration but on local information structure in a discourse.

Laurie A. Maynell. Discourse Constraints on Extraposition from Definite NP Subjects in English. *OSUWPL* Volume 58, Fall 2008, pp. 110-137.

1. Introduction.

The term *extraposition* refers to a range of syntactic structures, including *it*-extraposition, (of sentential subjects or infinitival clauses), and extraposition from NP subjects or objects, usually of prepositional phrases or relative clauses. The main characteristic of extraposition is that in a sentence, a constituent appears to the right of its canonical (or noncanonical alternative) location. In the following example pairs, the constituent given in italics is subject to extraposition. The first of each pair shows a sentence in its unextraposed form, the second in its extraposed form.

(1) a. *That he admitted to killing terrorists* really surprised me.
b. It surprised me *that he admitted to killing terrorists*.

(2) a. *To curb government spending* would be a good idea.
b. It would be a good idea *to curb government spending*.

(3) a. A law *that would require factories to reduce their emission of air pollutants by 70% over the next 3 years* was enacted during the previous administration.

b. A law was enacted during the previous administration *that would require factories to reduce their emission of air pollutants by 70% over the next 3 years*.

(4) a. A clown *in pink overalls* wandered into the dining hall.
b. A clown wandered into the dining hall *in pink overalls*.

(5) a. The countess greeted any man *who had a fortune* courteously.
b. The countess greeted any man courteously *who had a fortune*.

(6) a. Carson showed a book *with a tattered cover* to the audience.
b. Carson showed a book to the audience *with a tattered cover*.

Example (1) shows “it” extraposition, in which a sentential subject is located at the right edge of the sentence and the subject position is filled with the empty *it*. (2) is similar to (1) except that the extraposed constituent is an infinitival clause. (3) and (4) show extraposition of a relative clause and of a PP, respectively, from NP subjects. (5) and (6) show extraposition of a relative clause and a PP from NP objects.

This paper focuses on extraposition of relative clauses[1] from definite NP subjects, structures that have generally been viewed as either ungrammatical or unacceptable

[1] Extraposition of PPs from definite NP subjects is similarly problematic. Compare

(i) The cocktail waitress from Miami entered the dining room.
(ii) ?? The cocktail waitress entered the dining room from Miami.

(iii) The cocktail waitress in a red dress spilled a drink.
(iv) ?? The cocktail waitress spilled a drink in a red dress.

(Guéron, 1980; Guéron & May, 1984; Wittenberg, 1987; Culicover & Rochemont, 1990; Huck and Na, 1990, 1992). Compare (7b) with (8b).

(7) a. A cocktail waitress *who was wearing a blond wig* entered the dining room.
b. A cocktail waitress entered the dining room *who was wearing a blond wig*.

(8) a. The cocktail waitress *who was wearing a blond wig* entered the dining room.
?? b. The cocktail waitress entered the dining room *who was wearing a blond wig*.

The structures in (7b) and (8b) appear to be similar. Previous work on the syntax of these structures has focused on ruling out sentences such as (8b) by placing constraints on the structural configurations of extraposition at logical form (Guéron, 1980, Guéron & May, 1984, Baltin, 1981, 1983, 1984). Another study (Wittenberg, 1987) accounts for the difficulty in interpretation of sentences such as (8b) semantically using rules of interpretation in a discourse representation theory. Others (Huck & Na, 1990, 1992) have attempted to account for the acceptability of extraposition from definite NPs in terms of the discourse structure in which they may be felicitously uttered. Similar to the last of these studies, the present one is also an examination of the structure of discourse in which sentences with extraposed relatives may be felicitously uttered. In trying to explain the acceptability of extraposition of relative clauses from definite NPs, Huck & Na (1990) claim that the information contained in extraposed relatives must be given with respect to the discourse, in order to match the given status of the definite NP subject that the relative clause modifies. In contrast, I argue that information in an extraposed relative clause may be either given with respect to the discourse or the hearer, or it may be hearer-new and therefore also discourse-new (Prince, 1981), as long as it is the answer to the question under discussion (Roberts, 1996).

The rest of the paper is organized as follows. Section 2 sketches one syntactic account of extraposition that incorporates the definiteness restriction demonstrated above with example (8b). Section 3 covers the pragmatic account of extraposition proposed by Huck and Na (1990). In Section 4, an alternative pragmatic account of the definiteness

---

(ii) is only acceptable on the reading that the cocktail waitress traveled straight from Miami into the dining room, which is not equivalent to the unextraposed counterpart in (i). (iv) is acceptable only if interpreted such that the waitress spilled the drink into a red dress, whether it was worn by someone or not (or, less plausibly, such that the drink was hiding or otherwise contained inside a red dress when she spilled it). This reading is not equivalent to (i), in which the cocktail waitress is actually wearing a red dress.

Constructing appropriate examples of extraposition from NP of prepositional phrases is trickier than examples with extraposed relative clauses, because the PP can sometimes be construed with the verb phrase.

(v) The cocktail waitress in a red dress entered the room.
(vi) The cocktail waitress entered the room in a red dress.

(vi) could be the answer to the question “How did she enter?”

restriction is provided, using five separate discourse examples. Section 5 provides a summary and discussion.

2. A syntactic account of extraposition of relative clauses from definite NPs.

As an example, the structural account of extraposition given in Guéron & May, 1984, (henceforth GM) will be illustrated in this section.[2]

GM propose a syntactic account in which extraposition is permitted from indefinite NPs, but not from definite NPs. They consider relative clauses to be complements of the definite NP heads from which they have been extraposed, as defined by the head-complement relation given in (9) (GM p.4, example (11)).

(9) In a sequence of categories α_i, $\beta_i^1,\ldots, \beta_i^n$ in a structure Σ, $\beta_i^1,\ldots, \beta_i^n$ are complements to α_i only if α_i governs $\beta_i^1,\ldots, \beta_i^n$.

A complement to a head is thus defined as being a constituent that is governed by that head. The head-complement relation must apply at LF, in order to account for the fact that the NP and the extraposed relative that modifies it are construed together. An additional rule defines government (10) (GM, p.4 example (12)).

(10) α governs $\beta =_{\mathrm{df}} \alpha$, β are dominated by all the same maximal projections, and there are no maximal projection boundaries between α and β.

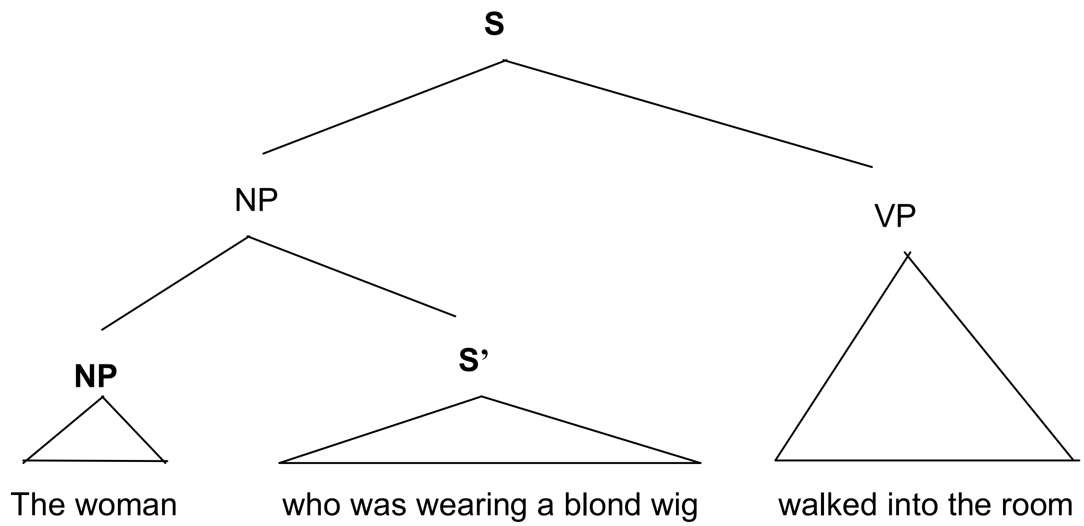
Taken together, the net result is that complements to NPs (i.e. extraposed relatives) must be dominated by all maximal projections (S-nodes) that dominate the NPs themselves. This of course must apply to unextraposed relatives as well. Consider the sentences in (11).

(11) a. The woman who was wearing a blond wig walked into the room.
?? b. The woman walked into the room who was wearing a blond wig.
c. A woman walked into the room who was wearing a blond wig.

In (11a), illustrated in Figure 1, all maximal projections dominating the NP *the woman* (in this case the **S**) also dominate the relative clause, which is the complement to the NP, in keeping with GM's head complement relation.

Figure 1. Structure of sentence (11a).

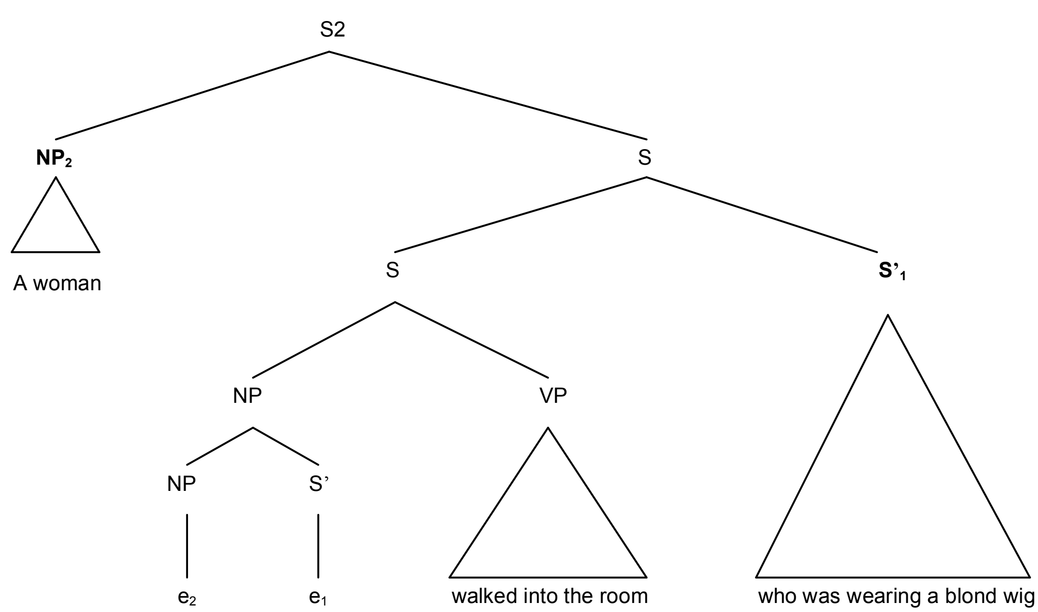
[2] There are other syntactic accounts of extraposition that will not be discussed here (for a brief review of some, see Wittenberg, 1987). Guéron, 1980 provides a syntactic and semantic account within the Extended Standard Theory, which is the starting point for her 1984 analysis with May described in Section 2 above. Baltin (1981, 1983, 1984) has an alternative structural account, formulating "generalized subjacency." Culicover and Rochemont (1990) develop a structural account, which does not involve movement and in which the structural configuration to be met applies at surface structure, in contrast to Guéron (1980) and Guéron and May (1984), whose analysis involves movement and whose constraints apply at logical form. However, Culicover and Rochemont do not address the definiteness restriction. For a semantic account within the framework of discourse representation theory, see Wittenberg (1987).



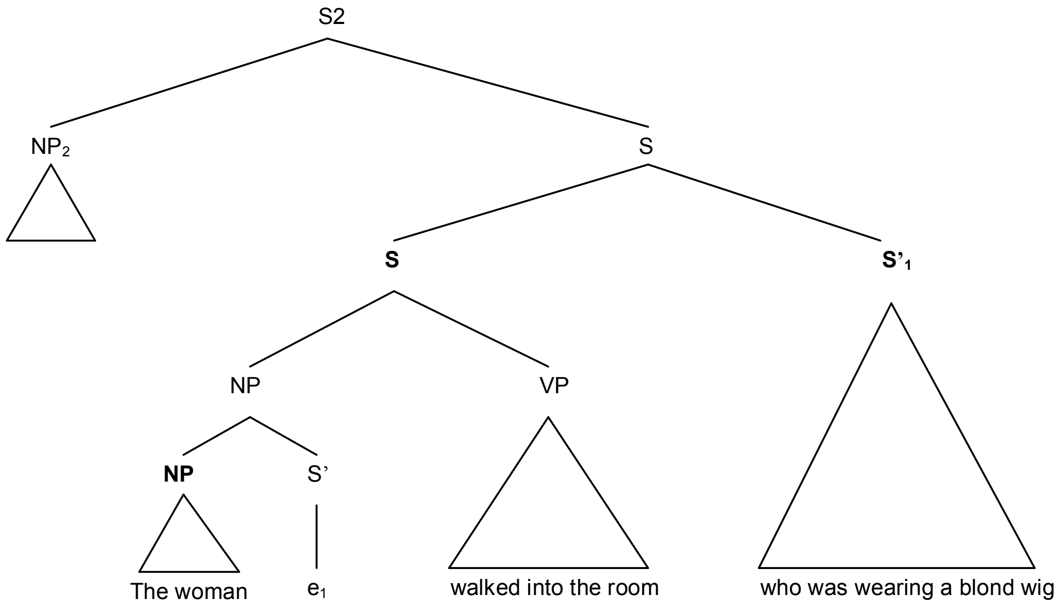
GM account for the difference between (11b) and (11c) by proposing a movement rule at LF which moves quantified NPs, but not deictic or definite NPs, leftward to a position adjoined to S. The movement rule will apply in (11c), such that the new S node created by the adjunction (S2, see Figure 2a) dominates both the NP and the relative clause that modifies it. In (11b), the rule will not apply, leaving the definite NP in situ (Figure 2b).

Assuming that the extraposed constituent adjoins to S, the extraposed constituent will no longer be dominated by the original S node, which dominates the definite NP subject.

Figure 2.
(a) Sentence (11c)



(b) Sentence (11b).



There are problems with GM's account. For example, using standard constituency tests, Culicover and Rochemont (1990) provide evidence showing that an extraposed phrase may be adjoined to VP and not to S. If they are right, GM's quantifier raising rule will not rule out sentences like (11b).

Reinhart's (1983, p. 150 ff.) analysis of sloppy identity suggests that definite NPs should also be subject to quantifier raising, in order to account for sentences such as (12).

(12) The exchange student enjoys her classes and so does the woman from IBM.

There are three possible interpretations of this sentence. In one interpretation, the referent of *her* is neither the *exchange student* nor the *woman from IBM*, in which case both the exchange student and the woman from IBM enjoy someone else's classes. In a second interpretation, the woman from IBM enjoys the exchange student's classes. In a third, the sloppy identity reading, the exchange student enjoys her classes and the woman from IBM enjoys not the exchange student's classes, but her own classes, which are not the same as the exchange student's. Expressing this interpretation at LF requires expressing the predication as a lambda abstraction (13) whose arguments are the definite NPs, as shown in (14) (to match the surface structure Reinhart places the arguments before the lambda abstraction instead of after them).

(13) λx (x enjoys x's classes)

(14) The exchange student (λx (x enjoys x's classes)) and
the woman from IBM (λx (x enjoys x's classes))

If definite NPs must be allowed to undergo quantifier raising, GM's quantifier raising rule will no longer account for the difference between extraposition from definite NPs and extraposition from indefinite NPs, in a sentence such as (15).

(15) ?? The woman hates her neighbors who lives across the street and so does the lady who lives on the corner.

Culicover and Rochemont (1990) argue against GM's quantifier raising rule in part by pointing out the following example (16) (Culicover & Rochemont, 1990, p. 36, fn. 28). GM's quantifier raising rule would not apply to the deictic *that*, and thus (16) would be wrongly ruled out by their account.

(16) That man just came into the room that I was telling you about.

An observation made in both Guéron & May (1984) and Culicover & Rochemont (1990) is that the acceptability of at least some of these structures seems to rely upon their interpretability with respect to the discourse context. Additionally, Culicover and Rochemont (1990, p. 30) point out that some examples involving extraposition require a certain stress pattern, making clear that they mean not lexical stress but a broader sentence level stress, whose exact pattern is a function of discourse context. Both observations suggest that specific characteristics of the discourse structure in which these sentences are felicitous warrant further examination.

Two studies (Huck and Na, 1990; Miller, 2001) consider extraposition specifically with respect to discourse structure. Miller (2001) analyzes *it*-extraposition, of sentential subjects and infinitival clauses, such as shown in previous examples (1) and (2). In most of his examples (all from corpora), extraposition is optional from a syntactic point of view, since the sentences have the same meaning in either structural configuration. His main finding is that if the content of the potentially extraposed phrase reiterates material in previous discourse context, it is not extraposed, but remains in its initial location (nearer the beginning of the sentence), whereas if the information contained in the extraposed phrase is going to be elaborated on in subsequent discourse, it is extraposed. He views this as a way to keep the discourse coherent, flowing from information already stated to information that is newer.

Huck and Na (1990, 1992) examine the context of sentences with extraposed relative clauses from NP subjects. Because the present study builds on the observations made by Huck and Na (1990), their account is discussed in some detail in Section 3.

3. A pragmatic account of extraposition of relative clauses.

Rochemont, (1986, Ch.4) claims that definite NPs from which relative clauses have been extraposed appear to require a stringent set of discourse conditions. He assumes that the structures themselves are well-formed, but whether or not they can be interpreted depends on finding the appropriate context for them, though he does not discuss what the possible contexts might be. Huck and Na (1990) attempt to do just that: determine the right context for sentences with extraposed phrases. Section 3.1 reviews the pragmatic account of extraposition proposed by Huck and Na, (1990), and the theory they adopt is outlined in Section 3.2. In Section 3.3, some problems with Huck and Na's (1990) analysis are described.

3.1. Huck and Na's (1990) claim regarding the definiteness restriction.

Like Rochemont (1986), Huck and Na (1990) (henceforth HN) take the view that the acceptability of sentences with phrases that are extraposed from definites is dependent on their discourse context. It follows that the restriction on their occurrence need not be accounted for in the syntax. Rather, their acceptability should be explained in pragmatic terms. Let us look at one of their examples (HN's example (6), p. 54).

(17) The guy just came in that I met at TRENO'S yesterday.[3]

According to HN, (17) is acceptable when some part of the extraposed relative has an intonational prominence[4]: in this case, the word *Treno's*. They describe a context in which the speaker has been talking about two people, one of whom he met at *Treno's* and one of whom he met somewhere else, e.g. at *Andrea's*. If the first of these two people walks in, the speaker could felicitously say (17); the emphasis on *Treno's* signals contrast with the person he met at *Andrea's*.

In the example just described, the NP subject refers to a person already in the common ground of the discourse - the speaker was already talking about this person. The information contained in the relative clause is not new, because the interlocutors know, from explicit mention in the conversation, that the speaker met one guy at *Treno's* and one at *Andrea's*.

HN reason that extraposed relatives will not work when there is a mismatch in information status between the definite NP and the relative clause. Citing Heim (1982), among others, they claim that use of the definite NP implies that its referent is familiar to participants in the discourse. More specifically, they state (HN: 60, n 14): "a definite NP is acceptable when the identity of its referent is calculable from the information given." In (17), the definite NP *the guy* has been explicitly mentioned at some point in the discourse, so the referent of the NP is present in the common ground. In addition, according to HN's description of the context for (17), both interlocutors know that the speaker met two people, one of those people at Treno's, so the property expressed by the relative clause, "meeting x at Treno's," is in the common ground as well. Since HN agree with Heim (1982) that definite NPs presuppose that the referent is calculable from discourse information, the information status of the definite NP will always be such that the referent of the NP is present in the discourse context somehow. If they assume this,

---

[3] Throughout this paper, critical intonational prominence on words will be expressed by printing the word in capital letters.

[4] HN actually use the word "stress," here and throughout their paper, instead of "intonational prominence." I am convinced that what they mean is not word-level (lexical) stress, but an actual pitch accent. For the purposes of this paper lexical stress will be distinct from the term "pitch accent." "Pitch accent" is what I believe Culicover & Rochemont (1990, p.10) to mean when they refer to a "broader sentence-level stress" which is apparently related to discourse context and required of certain sentences with extraposed phrases. I will use the term pitch accent to mean intonational prominence, over and above word-level stress. A pitch accent tends to be aligned to the stressed syllable of the word bearing it (Pierrehumbert & Hirschberg, 1990, Beckman, 1996).

then the only way for a mismatch to be generated, according to their view, would be to have the property expressed by the relative clause be new with respect to the discourse.

To summarize HN's position, extraposition of relative clauses from definite NPs is acceptable as long as the information in the relative clause is given in the discourse, so that it is congruent with the given nature of the definite NP.

3.2. Huck and Na's (1990) theory of focus.

To account for examples such as (17), HN need a way of determining when a phrase conveys new information, and when it conveys given information. Recall that one crucial factor in the acceptability of (17) was that a word in the extraposed relative clause, *Treno's*, needed to be uttered with a pitch accent to signal contrast between the person the speaker met at Treno's and the person the speaker met at Andrea's. HN note that contrastive information tends to be given in discourse. Recognizing the importance of intonation to signal contrastive (given) information in example (17), they develop a theory of focus to account for the acceptability of sentences with relative clauses extraposed from definite NP subjects. Following Rochemont (1986), they assume that extraposed phrases are in a focus position. Extraposition is one of the class of structures that Rochemont (1986, Ch 4) describes as constructional focus.

HN distinguish between informational focus (information that is new with respect to the discourse), contrastive focus (information that constitutes a different subpart of an utterance which is otherwise identical to some other utterance in the discourse), and interrogative focus (information being questioned). For the purposes of this paper, only informational and contrastive focus need be considered in detail. Whether or not focus is informational, contrastive or interrogative depends on their rules of focus interpretation.

3.2.1. Focus Interpretation

HN adopt a filing system similar to Heim's (1982) to model discourse, where a file card corresponds to an entity in the discourse.

3.2.1.1 Informational Focus

HN define informational foci as clauses which either cause new file cards to be added to the discourse filing system, or which cause a new proposition to be added to a card already created. As an example, imagine that (18) is uttered at some point preceding the utterance of (17) in the conversation.

(18) I met an unemployed ad executive at Andrea's party last weekend.
…..
(17) The guy just walked in that I met at TRENO'S yesterday.

In this discourse, the indefinite NP *an ad executive* is an informational focus, because it causes a new file card to be added to the discourse filing system. Until the utterance of this NP, the discourse referent does not exist in the common ground. All information about the ad executive, such as the fact that the speaker met him at Andrea's party, is entered on the card.

Index: 1
1 = 'ad executive'
1 is unemployed
speaker met 1 at Andrea's party last weekend

3.2.1.2. Contrastive Focus

A contrastive focus is that constituent in a new sentence, which differs from another constituent in an otherwise identical sentence already present in the discourse. (19) shows a sentence pair (HN's examples (14) and (16)) illustrating contrastive focus.

(19) a. Did a guy come in here who was holding a rabbit?
b. Did a guy come in here who was holding a duck?

(19b) is identical in its interpretation to (19a), with the exception that *duck* replaces *rabbit*, so *duck* is the contrastive focus of (19b).

3.2.2. Focus Assignment Rules

In addition to focus interpretation rules, HN adopt focus assignment rules from Rochemont (1986) and Selkirk (1984). One focus assignment rule, adopted from Selkirk (1984), states that a constituent bearing a pitch accent is a focus, and a focused constituent has a pitch accent somewhere in it.[5] Given HN's assumption that an extraposed phrase is in a focus position (Rochemont, 1986), it will necessarily have a pitch accent somewhere within. This rule allows them to guarantee the presence of a pitch accent in the extraposed phrase, to signal contrast with something else in the discourse.

To rephrase HN's argument, a relative clause can be extraposed from definite NPs as long as it is a contrastive focus and not an informational focus. If it is an informational focus, it conveys new information, and is therefore not given with respect to the discourse content. Since the content of the relative clauses must be given, the pitch accent within it, according to the focus assignment rules, would have to be one that signals contrast of the word bearing the pitch accent with some other entity in the discourse.

3.3. Problems with HN's Analysis

3.3.1. Extraposed phrases need not always have pitch accents in them.

The focus assignment rule that ensures that an extraposed phrase is necessarily focused and will necessarily have a pitch accent in it is apparently adopted by HN in

[5] Huck and Na (1990) only discuss intonationally marked focus, aside from noting Rochemont's constructional focus view of constituents that have undergone rightward movement.

order to account for the observation that phrases extraposed from definite NPs are typically uttered with a pitch accent somewhere in them.

There are exceptions to this, however. HN themselves give an example indicating that there are cases in which a phrase extraposed from a definite NP need not have such intonational prominence. (HN: 58-59, examples (14) and (19), repeated below as (20) and (21)).

(20) Did a guy come in here who was holding a duck?
(21) No, but a GIRL came in here who was holding a duck.

Here, the contrast is between the words *guy* and *girl*, conveyed by a pitch accent on the word *GIRL*. This word is not inside the extraposed relative. This forces them to add a qualification to their rule: namely, that if there is a focus somewhere else in the sentence, as in (21), the extraposed phrase itself does not have to contain one. Their rule about extraposed phrases is thus challenged.

Recall that HN assume that an extraposed phrase is necessarily focused, and that focused phrases necessarily have a pitch accent somewhere in them. Perhaps, then, there are cases in which extraposed relatives are not focused, and Rochemont's (1986) assumption should be dropped?[6] With the qualification motivated by (20-21), given a contrastive focus somewhere other than inside the relative clauses, there could be a contrastive focus in the relative clause, but there need not be (does that mean they could be focused, but need not be?). It is difficult to devise a counterexample to test this rule, given that it allows for two of two possibilities: there either is or is not a contrastive focus inside the relative clause.

3.3.2. Extraposed phrases can express information that is discourse-new.

Sentences (22) and (23) comprise a discourse in which the extraposed relative clause in (23) conveys information that is new with respect to the discourse. The setting of the conversation is Speaker B's house. Speaker A has been to Speaker B's house several times.

(22) *Speaker A*: Weren't there five bottles on that shelf when I was here the other day?
(23) *Speaker B*: Yeah, but during the earthquake, the TWO fell to the GROUND that were CLOSEST to the EDGE.

Speaker A remembers five bottles, but has no way of knowing what has happened to the two that are missing until Speaker B tells him. Thus, the information in the relative

[6] It is important to keep in mind that in English, it is difficult to utter a long string of words without any pitch accents at all (Beckman, 1996). Example (21) could be so uttered, especially at a rapid speech rate, though it is more likely that another pitch accent would fall within the extraposed relative. If there were a pitch accent in the relative clause, in addition to the pitch accent conveyed by the contrast in the matrix clause, it might convey focus, of some type, or it might simply be there because of the length of the utterance.

clauses is new to Speaker A. This discourse will be discussed in greater detail in the next section, along with other example discourses.

To summarize thus far, HN's claim that extraposed phrases are necessarily focused, and necessarily contain a pitch accent, is undermined by example (21). Further, it is not the case that the content of the relative clause must be given or in contrast with given information.

4. Extraposition from definites, reconsidered.

As noted previously, Rochemont, (1986, Ch.4) assumes that sentences in which relative clauses have been extraposed from definite NP subjects are well-formed, and that they require stringent discourse conditions in order to be acceptable. He does not discuss what the appropriate discourse contexts might be. Intuitively, he states that extraposed phrases are presentational in the sense that they provide information, though he does not go into detail about the nature of this information.

Using example discourses, I show that extraposed relatives serve a discourse function by providing information that answers the immediate (local) question under discussion (Roberts, 1996). In contrast to Huck and Na (1990), what is critical for extraposition is not the information status – given or new with respect to the discourse as a whole - of the content of the relative clause and its relationship to the same information status of the definite NP. It is true that the definite NP and the relative clauses must match. That is, if the definite NP is discourse-old, then the property expressed by the relative must also be discourse old, otherwise it would not be possible to identify the referent of the definite NP. In cases where the content of the relative clause is new, the definite NP must also be new with respect to the discourse.[7] However, this fact has no bearing on whether or not extraposition is acceptable, as will be demonstrated in Section 4.5.

This section begins with comments on intonation in Section 4.1, followed by a review of how definite NPs are licensed in Section 4.2. Section 4.3 discusses given-ness and newness of discourse referents from the point of view of the discourse and the hearer. Section 4.4 explains the question under discussion so that it may be used to account for the examples with extraposed relatives discussed in Section 4.5. The examples in Section 4.5 are fabricated, but Section 4.6 shows one example sentence which was found in a newspaper article, and which provides support for the main claim of this paper, that a relative clause extraposed from an NP subject answers the immediate question under discussion.

4.1. A word about intonation

The role of intonation in understanding spoken language is important, since intonation conveys information about the discourse context in which the sentence so uttered is felicitous. For a discussion of the role of prosodic focus as conveyed by pitch

[7] In the case of a new discourse referent being introduced with a definite NP, the existence of the referent is accommodated. This is described in greater detail in section 4.5, for example Discourses 3 and 4.

accent, see Roberts (1996) and Selkirk (1995). For a description of intonational contours and their meanings, see Pierrehumbert and Hirschberg, 1990. Others (Rochemont, 1986, Culicover and Rochemont, 1990, Huck and Na, 1992) have observed that sentences with extraposed relative clauses are acceptable if pronounced in a certain way, with pitch accents in certain constituents, and it is very likely that the pitch accents signal the relationship of the accented items to the discourse.

While intonation is significant, it is omitted from this study so that the properties of the information structure in the discourse can be worked out by themselves. Intonational prominence in all relevant examples will be indicated by putting words that bear pitch accents in capital letters. This is done in order to recognize that certain pronunciations of these sentences are necessary. However, an explanation of the exact functions of those pitch accents and other prosodic characteristics such as phrasing is beyond the scope of this paper.

4.2 Definite NPs

About the use of definite NPs, HN say only that the referent of the definite must be calculable from the discourse information. Roberts (2002) gives a detailed account of how definite NPs may be licensed. She defines two terms: strong familiarity and weak familiarity.

The plain term "familiarity" originates with Heim (1982), who says that definite NPs presuppose familiarity. For Heim (1982), this meant that the referent of a definite NP would be accessible to interlocutors, either because it had been introduced via explicit utterance of an indefinite at a prior point in the discourse, or because it was or became perceptually accessible to all interlocutors.

Roberts (2002, pp. 14-15) defines a taxonomy of familiarity in which strong familiarity and four sub-types of weak familiarity are distinct. A strongly familiar NP has as antecedent a discourse referent introduced by the explicit utterance of an NP. An example is given in (24)

(24) There is **a** squirrel that comes to my office window every day. This morning, **the** squirrel was carrying an acorn in its mouth.

Referents which are accessible in the discourse, but not licensed because of explicit mention of an indefinite NP earlier in the discourse are termed weakly familiar. There are four kinds of weak familiarity. An NP can be weakly familiar if the entity referred to is perceptually accessible to interlocutors. This includes cases in which entities enter or are in the discourse space, such as a bird flying by, a person approaching the speakers, or some event occurring within earshot and sight of the speakers. For example, if a group of people are having a conversation in a room lit by an overhead light, and this light flickers, it would be possible for one of the interlocutors to ask (25).

(25) What's wrong with **the** light in here?

An entity which is globally familiar in the general culture or at least familiar to the to the discourse participants because of common experiences is also weakly familiar. For instance, at the present time, it is possible to refer to "**the** war with Iraq" in a conversation among people who have been following the news.

An entity can be weakly familiar if it is introduced into the discourse because its existence is entailed by the discourse context. Example (26) illustrates this.

(26) Every faculty office has a computer in it. In Shari's office, **the** computer was making strange noises.

The first sentence in (26) may be uttered as is written, or it may be common knowledge among the interlocutors that every faculty office is equipped with a computer. In the second utterance, once we have begun by saying "In Shari's office,…" the shared knowledge of the interlocutors that Shari is a faculty member entails that her office must have a computer in it, and so it can be referred to with a definite NP.

The fourth way that weak familiarity is realized is by giving a functional interpretation to the definite description with a familiar and salient argument. The example given by Roberts (2002, p.3 (5)) for this case is the following:

(27) There is a statue on **the** dashboard of this car.

According to Roberts (2002), *the dashboard of …* is a relational function, whose argument is *car*. Other such functional relations exist, such as the broiler (of an oven), the furnace (of a house or other building which typically only has one furnace), the water heater (of a house), the hard drive (of a computer), the heart (of a living creature), and the steeple (of a church).

Roberts (2002) concludes this section of her paper by stating that referents of definite NPs must be weakly familiar, where weak familiarity subsumes strong familiarity. Hers is a precise account of what it means to be "calculable from the discourse information" as HN put it.

There are times when accommodation is necessary before an NP's referent is recoverable. Roberts (2002) offers the following example (28) (Roberts, 2002, p.11, (19)).

(28) John was murdered yesterday. The knife lay nearby.

There are a number of ways that people can be murdered, and once the hearer interprets the second sentence, s/he will accommodate the fact that John was stabbed to death. Once this has been accommodated, the context will entail the existence of a knife.

For the example discourses that follow in Section 4.4, the definite NP subjects will be familiar in one of the ways described above, except for two discourses in which the referent of the definite NP subject will be recoverable via accommodation, because its referent is being newly introduced into the discourse.

4.3. Given-ness vs. newness

It is useful to distinguish three types of given/new: given/new with respect to the discourse, given/new with respect to the hearer, and given/new with respect to the question under discussion.

Prince (1981) distinguishes between given/new with respect to the discourse and given/new with respect to the hearer. Information that is new with respect to the hearer is necessarily new with respect to the discourse, because we expect interlocutors to keep track of the information discussed in any conversation. However, information that is given with respect to the hearer need not be given with respect to the discourse. For example, a member of a family that owns a dog may begin a conversation by asking another family member (29).

(29) Have you fed **the** dog yet this evening?

The definite NP *the dog* is new with respect to this discourse, because this is its first mention in this particular conversation. With respect to the interlocutors (hearers), it is old information, since the referent of *the dog* is a known member of the family.

Information that is old with respect to the question under discussion is a part of the question under discussion (QUD), or on the QUD stack. Information that is new with respect to the question under discussion is at least a part of the answer to the QUD. In order to understand what it means to be old or new with respect to the question under discussion, it is necessary to explain how the question under discussion is defined.

4.4. The question under discussion.

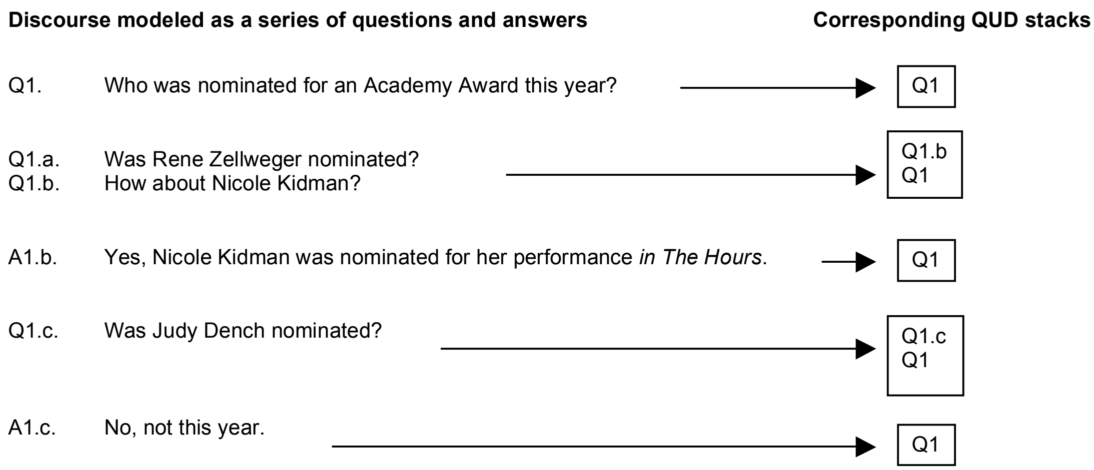
In Roberts' (1996) theory of information structure in discourse, discourse is modeled as a series of questions and answers to those questions. Her theory defines a "question under discussion" as well as an "immediate question under discussion," both of which are illustrated in this section. This section provides a short, basic description of these. For a detailed formal account of the theory, the reader is referred to Roberts (1996).

The **q**uestion **u**nder **d**iscussion, or QUD, is defined as a function from the set of questions and answers that make up a discourse to ordered subsets of accepted questions. When a question is asked and accepted (or accommodated, in the case of an implicit question), that question is added to an ordered QUD stack. This question stays on the stack until it is answered, or until it is deemed unanswerable by the speakers. If a sub-question of the first question is asked, it is also added to the stack. The stack is ordered such that questions asked later in the discourse are higher on the stack than questions asked earlier. This process is illustrated in Figure 3 with the following short discourse in (30).

(30) Q1. Who was nominated for an Academy Award this year?
Q1.a. Was Rene Zellweger nominated?

Q1.b. How about Nicole Kidman?
A1.b. Yes, Nicole Kidman was nominated for her performance in *The Hours*.
Q1.c. Was Judy Dench nominated?
A1.c. No, not this year.

Figure 3. A series of four QUD stacks at various points in the discourse (30).



Female Oscar nominees are the topic of this conversation, and since the interlocutors accept the topic of conversation, Q1 (the broader question under discussion) is added to the stack. The second QUD stack shown in Figure 3 is what the QUD stack looks like after utterance of Q1.b. Q1.a and Q1.b are sub-questions of Q1. According to Roberts' theory, answering a sub-question entails partially answering Q1. A complete answer to Q1 would include answering all of the possible sub-questions of Q1, or at least the set of sub-questions that interest the interlocutors. Q1.a. is asked but not accepted (because none of the interlocutors provide an answer), and so it is not added to the stack. Q1.b, on the other hand is accepted - we know this because it will be answered by the next utterance - and so it does get added to the stack. At this point in the discourse, Q1 is the question under discussion, and Q1.b is the <u>immediate</u> question under discussion.

When A1.b is uttered, it answers Q1.b. After utterance of A1.b, the corresponding question, Q1.b, is removed from the stack, and the QUD stack is as shown in the third box in Figure 3. Q1 is not removed, because it has not been completely answered yet. Another sub-question, Q1.c is asked. At this point in the discourse, Q1.c is added to the stack (the fourth box in Figure 3). A1.c answers this question, and after utterance of A1.c, Q1.c gets popped from the stack.

It is worth noting that real conversations are rarely this explicit and simple. For example, a person is unlikely to enter her office in the middle of the morning and ask (31) (though it isn't impossible to do so).

(31) Q1. Does anyone want to go and get coffee?

Q1a. Jane, do you want to get coffee?
Q1b. Kris, would you like to get coffee?
Q1c. How about you Anne, would you like to get coffee?

More likely the person asking the question will only ask the big question, Q1, and assuming that Jane, Kris and Anne are in the office and heard her, each will answer either yes or no. So the discourse is more likely to proceed as shown in (32).

(32) Q1. Does anyone want to go and get coffee?
A1a. No, thanks, I have to run to a meeting.
A1b. *Kris says nothing but looks up from her work long enough to smile and shake her head, indicating "No, thanks."*
A1c. Oh, yeah! - as soon as I finish this writing this message.

As (32) illustrates, the sub-questions may not be asked explicitly, but they are implicit, and can be modeled using the question-answer paradigm.

Often, there are no explicit questions asked in a discourse. Even the coherence and the flow of information of this kind of discourse (including monologues, which essentially consist of a person telling another person a long story) can be modeled by positing questions which all of the assertions that make up the discourse answer. This will be the case in the example discourses in Section 4.4.

As stated in the previous section, information can be given or new with respect to the broad question under discussion, and also with respect to the immediate QUD. If information is given with respect to the QUD, it is part of the QUD, or on the QUD stack. Information that is new with respect to the question under discussion is (at least a) part of the answer to the QUD. In the case of relative clauses extraposed from definite NP subjects, the extraposed relative must be new with respect to the immediate question under discussion. In other words, the content of the relative clause must answer the immediate question under discussion. The noun phrase head itself should be a part of the QUD. In order to avoid confusion with the terms discourse-given/new and hearer-given/new, I will use "informative and non-informative with respect to the QUD" rather than given/new with respect to the QUD. This choice of terminology is also consistent with Rochemont's (1986) intuition that extraposed relatives present information.

4.5. Analysis of example discourses.

Four sample discourses are illustrated in this section. A sentence with a relative clause extraposed from a definite NP subject is part of each discourse.[8] In analyzing the information structure of these sentences it will become clear that the extraposed relative is an answer to the immediate QUD, and that the NP subject that the relative modifies is part of the QUD.

[8] It is worth noting that extraposition is not obligatory in any of these examples. The unextraposed structures are grammatical in each example discourse.

Before proceeding to the four example discourses, HN's example (17), printed again as (33), is reconsidered using the question-answer paradigm described in Section 4.4.

(33) The guy just came in that I met at TRENO'S yesterday.

Recall that there are two persons in the common ground. The speaker met one person at Andrea's and one person at Treno's. If the second person enters the room, the speaker can felicitously utter (33), even without someone explicitly uttering the question under discussion (34).

(34) Which guy just walked in?

The definite NP is licensed because the referent is strongly familiar (as defined by Roberts, 2002): this person has been talked about earlier in the conversation. The noun phrase head of the relative clause, *the guy*, is already a part of the QUD (*which guy?*), and therefore not informative with respect to the QUD. What is informative is the property expressed by the relative clause, namely the property of having been met by the speaker at Treno's the previous day.

By simply specifying the given-ness of the definite NP subject and the given-ness of the information conveyed by the relative clause, the discourse function of the extraposed relative is missed. Use of the QUD stack allows a modeling of the discourse that is localized enough to show this function.[9]

Discourse 1 has an informational structure similar to HN's example. The definite NP subject is discourse-given.

Discourse 1. *Setting*: Terry, a doctor, is telling Jan about her trip out of town to a nationwide conference on heart disease prevention, which was attended by doctors from around the U.S. In addition to presentations, panel discussions, etc., recreational activities were scheduled and conference attendees could take part in these activities. One particular evening, a group of the conference participants chose to attend a football game, while several others went to a Tony Bennett concert. Terry was not feeling well, so rather than go to either event she returned to her hotel room and went to bed early. The next morning, she noticed that some of the doctors were in a disagreeable mood.

(35) *Terry* (continuing): I'm still not certain, but I think THOSE DOCTORS were sulking who had been at the FOOTBALL game the night before. Paul told me later that it was a bad game, and their team lost.

[9] Guerón (1980) observes that extraposition is more acceptable in sentences with presentational predicates, also known as verbs of appearance. The predicate *came in* in (33) functions here as a verb of appearance. The verbs in this class are described by Rochemont (1986) as serving no other purpose in the sentence than to set a scene for the subject, or introduce the subject into the discourse. An explanation for why extraposition is easier in sentences using verbs of appearance is beyond the scope of this paper. However, in the four example discourses in this section, verbs of appearance are not used, in order to ensure that the proposed analysis using the QUD is not dependent on their presence.

In this discourse, the NP *those doctors who had been at the football game* is weakly familiar. Terry has already told Jan that she was at a conference of doctors, and that one evening a group of doctors went to a game, another group went to a concert, and she was not part of either group. This example is like HN's example (33) in which two individuals have been under discussion. In fact, in (35), the deictic, *those*, is felicitous since Terry is singling out one of two groups. According to HN, then, the referent of this NP is given with respect to the discourse, as is the property expressed by the relative clause.

In Roberts' (2002) terms, the referent of the NP is weakly familiar. Recall that weak familiarity subsumes strong familiarity, and in this discourse it is likely that the referent denoted by this noun phrase is strongly familiar. For example, Speaker B may have uttered (36) at some time prior to the first sentence in (35).

(36) Some of the doctors went to a football game, and some went to see Tony Bennett, but I felt a migraine starting so I just went back to my hotel room and slept.

The immediate question under discussion which the extraposed relative provides an answer to is (37).

(37) Which doctors were sulking?

The predicate *sulking* recalls a fact that has already been mentioned, which is that not all doctors were in a good mood the morning after their outing. With respect to the question under discussion, it is not informative. (37) is not explicitly asked, but implied, and for Terry to tell a coherent story, she must provide an answer to the question.

The QUD stack at the point in the discourse just before (35) is shown in Table 1. Since Speaker B is telling her friend about the conference she attended, the broad question under discussion of this discourse is *How was the conference that you attended?* The immediate question under discussion, (37), is highest on the stack because it is the most recent question.[10]

Table 1. QUD stack for Discourse 1.

Immediate QUD	Which doctors were sulking?
Broad QUD	How was the conference that you attended?

The immediate QUD does not stay on the QUD stack for long, since the content of the extraposed relative in (35) provides the answer to it. At the end of the sentence, we are left with just the broad QUD on the stack, until the next question arises in the discourse.

[10] It is possible for there to be other sub-questions on the stack between the broad QUD and the immediate QUD, but it is not necessary to show this in order to illustrate my claim about extraposed relatives.

Another way to confirm the informativeness of the extraposed relative clause is to rephrase the first sentence in (35) using a cleft structure *It was X that Y*, in (38), where Y represents background information and X is an answer to "Who Y'ed?" Note that if X and Y in (38a) are reversed, as in (38b), the sentence is contextually inappropriate.

(38) a. I'm still not certain but I think it was the doctors who went to the FOOTBALL game that were sulking the next morning.
b. ?? I'm still not certain, but I think it was the doctors who were sulking the next morning who went to the football game.

In (38a), X, *the doctors who went to the football game*, is the focal constituent. This is consistent with the focal property of the extraposed relative *who had been at the football game* in (35). The background information Y, *that were sulking*, is already part of the question under discussion. The focal constituent of the *it*-cleft construction gives an exhaustive answer to the question under discussion (Roberts, 1998).

If the relative clause must be informative with respect to the immediate question under discussion, it follows that when the relative clause is not informative, the sentence will be unacceptable. This case is illustrated by Discourse 2.

Discourse 2. *Setting*: Same as for Discourse 1, including the information that some doctors went to a football game, and some went to a Tony Bennett concert, and that Terry went to bed early with a migraine. However, the information about the doctors being in a bad mood the next morning is excluded from this context, and instead, Terry reports (39) to Jan.

(39) ?? *Terry*: The next day, during the first coffee break, the doctors were SINGING and DANCING who had been at the TONY BENNETT concert.

This sentence is unacceptable, but the unacceptability does not come from the definite NP's information status. As in Discourse 1, the referent of the definite NP subject, *the doctors who had been at the Tony Bennett concert*, is weakly familiar (and most likely strongly familiar). According to HN, this sentence ought to be acceptable.

Since the *singing and dancing* in the predicate is the new information with respect to the hearer, Jan, and therefore also with respect to the discourse, an equivalent cleft sentence is (40). This cleft has a different structure than the previous one shown in (38a). The structure in (40) is *What Y did was X*, where Y represents background information and X is the answer to "What did Y do?"

(40) What the doctors who had been at the Tony Bennett concert did the next morning was sing and dance.

The question under discussion for which (40) is a response is (41). Another appropriate response to (41) would be (42), in which the relative clause is not extraposed.

(41) What did the doctors who had been at the Tony Bennett concert do the next morning?

(42) During the first coffee break, the doctors who had been at the Tony Bennett concert were singing and dancing.

The extraposed relative in (39) provides an answer for a question that has not been asked, namely, (43), and it is therefore infelicitous.

(43) Which doctors were singing and dancing?

This suggests that in addition to the information expressed by the NP, the information expressed by the predicate must be non-informative with respect to the question under discussion, as it was in Discourse 1.

Discourses 3 and 4 show that it is possible to extrapose relative clauses from definite NP subjects that introduce new referents into the discourse. The property expressed by the relative clauses must be congruent with the information status of the definite NP, so it too must be discourse-new. What is demonstrated with these discourses (contra HN) is that it is possible to have new information expressed by the extraposed relative clause. As demonstrated in Discourses 1 and 2, what is necessary for extraposition is that the information expressed by the extraposed constituent be an answer to the immediate QUD.

Discourse 3. *Setting*: Dan's dining room. The speakers, Bill and Dan, are friends, and Bill has been to Dan's house before. Bill can see three bottles on the shelving unit in Dan's dining room. The shelves have a lip on the front edge, to prevent items from falling off them. Bill remembers from a previous visit that there used to be five bottles on that shelf.

(44) *Bill*: Weren't there five bottles on that shelf when I was here the other day?
(45) *Dan*: Yeah, but during the earthquake, the TWO fell to the GROUND that were CLOSEST to the EDGE.

This being Dan's dining room, he will know that there are only three on the shelf (whether he is looking at the shelf or not), and Bill must be able to see the shelf in order to ask the question he just asked.

Both speakers know that five minus three is two, and thus the context entails the existence of the two missing bottles – their existence is weakly familiar.[11] Dan could

[11] This example is similar to Partee's (1970) example of the missing marbles.

(i) I dropped ten marbles and found only nine of them.
* It is probably under the sofa.

Roberts (2002, p. 39) points out that the problem with (i) is that the missing marble is weakly familiar, but that weak familiarity is not enough to license use of the pronoun to refer to the familiar dropped and missing marble. On the other hand, weak familiarity is enough to license a definite description to refer to the missing marble as shown in (ii) (Roberts, example (59)).

have therefore referred to the two missing bottles with a definite description as shown in (46).

(46) During the earthquake, the two missing bottles fell to the ground.

In fact, though, he did not do that. Instead, the discourse referent represented by the subject NP, *the two (that were closest to the edge)*, is new to the discourse (Prince, 1981). Bill, the hearer, can infer that this new discourse referent and the weakly familiar referent are the same. The information in the predicate *fell to the ground* is also discourse/hearer-new.

Bill's yes/no question can be added to the stack as is, but Dan is responding to a different question (47).

(47) What happened to the two missing bottles?

This question is not uttered explicitly, but is implicated by Bill's actual question (44). Beginning with Bill's utterance, the discourse can be modeled using Roberts' (1996) theory of information structure in discourse as shown in (48).

(48) Q1. Weren't there five bottles on that windowsill the other day?
A1. Yes.
Q2. What happened to the two missing bottles?

Because the information in the sentence (45) as a whole is hearer-new, it is necessary to show the ordering of information in B's response. First, a response equivalent to (45) using clefting is shown in (49).

(49) What happened to the two that are missing was they fell to the ground.
The reason those two fell to the ground was they were closest to the edge.

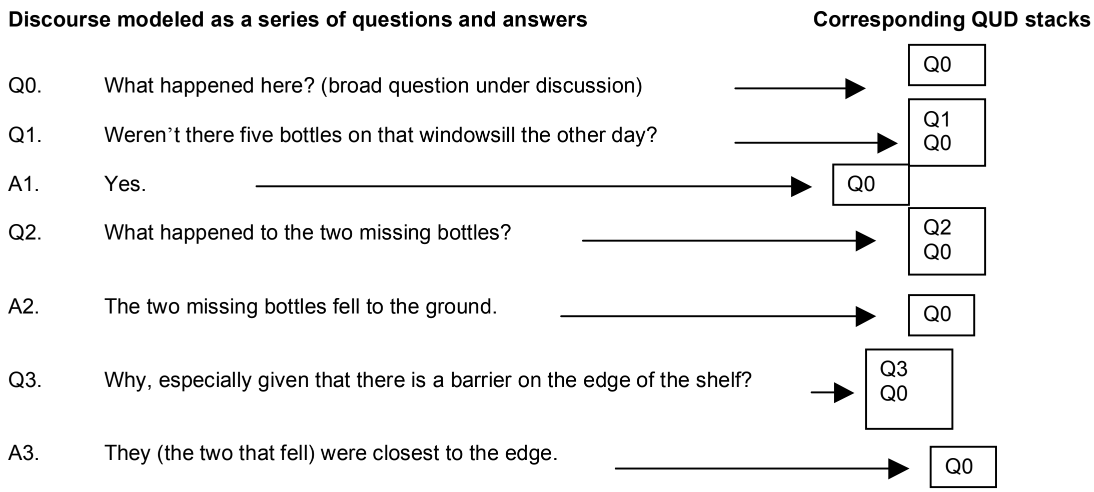
The equivalent response in cleft sentences shows what is considered background information and what is considered to be new. The first sentence in (49) is the answer to Q2 in the model (48). To explain the extraposed relative, we posit another implicit question, Q3, and the complete discourse is modeled in (50).

(50) Q1. Weren't there five bottles on that windowsill the other day?
A1. Yes
Q2. What happened to the two missing bottles?
A2. The two (missing bottles) fell to the ground.
Q3. Why (especially given that there is a barrier on the edge of the shelf)?
A3. They were closest to the edge.

(ii) I dropped ten marbles and found only nine of them.
The missing marble is probably under the sofa.

The QUD stack for this discourse changes as shown in Figure 4. In order for the extraposed relative to be felicitous, it must answer the immediate QUD, which it does by answering Q3.

Figure 4. QUD stacks for the discourse in (50).



Q1, Q2 and Q3 can all be considered sub-questions of the broad QUD, "What happened here?" If, after A3, this topic has been addressed to the speakers' satisfaction, then Q0 will also be removed from the stack after A3, and a new topic might be introduced.

This discourse model differs from the previous two in that a single sentence (45) is modeled as a discourse which includes two separate sub-questions Q2 and Q3 and the answers to those sub-questions, A2 and A3. The claim made in the previous section about the predicate's information having to be non-informative with respect to the QUD still holds true for this discourse if modeled in this way. By the time the beginning of the relative clauses is reached, the predicate's information *fell to the ground* has been added to the common ground, and there is a new immediate QUD on the stack.

> **Discourse 4**. *Setting*: a card game using a standard 52-card deck. There are a few possible ways to win the game. One person is explaining the rules to the other players.
>
> (51) If the ace of spades is drawn during anyone's turn, all players will be forced to show their hands. At that point in the game, the HAND wins that is the LOWEST in VALUE.[12]

The definite NP subject *the hand (that is the lowest in value)* is hearer- (and discourse-) new. Even though the word *hand* was just used in the plural in the previous utterance, the particular hand which is under discussion in the final sentence in (51) has not been mentioned. However, the referent of this NP is weakly familiar, because the fact that

[12] This example is slightly modified from one due to Daniel Büring, p.c.

each player has a hand is known to the players, and it is these hands of cards that are referred to in the speaker's first sentence.

Given that the object of card games is generally to win, the predicate, *wins*, represents a default assumption, and this hearer-new information is easily accommodated.

The larger (implicit) question under discussion here is *Which card or combination of cards wins the game?* The immediate question under discussion, which is answered by the extraposed relative in (51) is (52).

(52) Which hand wins?

The second sentence in (51) is paraphrased as a cleft sentence (53) to show the order of information.

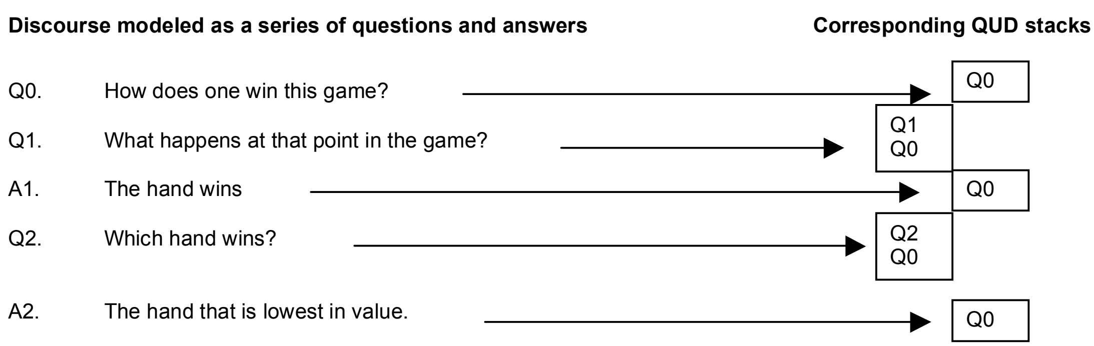
(53) What wins at this point in the game is the hand that is lowest in value.

As in Discourse 3, a single sentence can be modeled as a discourse. This is shown in (54).

(54) Q0. How does one win this game?
Q1. What happens at that point in the game?
A1. The hand wins
Q2. Which hand wins?
A2. The hand that is lowest in value.

The QUD stacks for the discourse in (54) are shown in Figure 5.

Figure 5. QUD stacks for (54).



Q1 and Q2 are sub-questions of the broad QUD, "How does one win this game?"

Discourses 3 and 4 demonstrate that definite NPs may refer to discourse-new entities that are weakly familiar and thus accessible to interlocutors. The property of any relative clause modifying such an NP would also have to be new with respect to the discourse. HN are correct in their claim that the information status of the relative must match the information status of its definite NP head, but it is not necessary for the definite NP to refer to a referent that is already present in the discourse, and further, this matching of information status has no bearing on whether or not a relative clause may be extraposed.

4.6. Found example.[13]

On the Opinion Page in the February 8, 2003 Columbus Dispatch is a piece written by Professor J.B. Quigley, which discusses the possible consequences of an invasion of Iraq. The first short paragraph describes President Bush's three main reasons for wanting to invade the country, and the second paragraph lists counterarguments to each reason. The third paragraph begins with the following sentence (55).

(55) If Bush's reasons are doubtful, then perhaps the critics are correct who see the true aim as gaining access to Iraq's oil reserves.

The writer does not use the word *critics* in the first two paragraphs preceding this sentence, and yet, the referent of this definite NP is weakly familiar. It is entailed by the context, because it is familiar in the global culture at least by those following the news: as Bush was preparing to go to war with Iraq during this time, many critics around the world protested his stated motivations.

The immediate question under discussion answered by the content of the relative clause is (56). To show this further, (55) is rephrased as a cleft sentence in (57).

(56) Which critics are correct?

(57) It is the critics who see Bush's true aim as gaining access to Iraq's oil reserves who are correct.

This example can be analyzed in the very same way that the sentences in Discourses 1-4 are. All examples demonstrate that extraposition of relative clauses from definite NP subjects is possible if the content of the relative clause is informative with respect to the immediate question under discussion. The definite NP is not informative with respect to the QUD, it is part of the QUD, as is the information expressed by the predicate in each example (at least by the time the extraposed relative is encountered, as in discourse (45), modeled in (50)).

5. Summary and Discussion

In summary, the definiteness restriction on extraposition from definite NPs is a function not of syntax, but of pragmatics. The structure must be allowable by any syntactic theory. Whether the sentence is acceptable depends on the relationship of the information conveyed by the extraposed phrase to the discourse context.

[13] A preliminary search of the Brown and Wall Street Journal corpora yielded no sentences with extraposed relative clauses from definite NP subjects, so example (55) appears to be a rare find.

HN observe correctly that the information status of the relative clause must match the information status of the definite NP. This must be so in order to ensure that the referent of the definite NP be identified. However, it is not necessary for the referent of the NP to be present by having been explicitly uttered in the discourse at the time of the NP's utterance. The referent may be weakly familiar.

This matching of information status has no bearing on the acceptability of extraposed relative clauses, and it is in this respect that the present study differs from HN. The claim here is that extraposed phrases critically must answer the immediate question under discussion, or QUD (Roberts, 1996), in order to be acceptable. Of course, extraposition is not obligatory, but it is possible as long as the relative clause's content is informative with respect to the immediate QUD. What this paper shows, then, is that a finer-grained analysis of the information flow in a discourse is necessary to explain the relationship between the content of the extraposed relative and the previous discourse. Whether that content is discourse-given, hearer-given, or hearer-new is insufficient to account for its informativeness with respect to a local point in the discourse.

This work is consistent with two other studies that examine syntactic operations from a pragmatic point of view. Miller (2001) showed using examples from corpora that whether or not a constituent was extraposed depended on the relationship of that material to the discourse. If the content of the extraposable phrase reiterated material from a prior point in the discourse, it remained unextraposed. If the phrase contained information that would be explained further in the discourse, it was extraposed. Miller (2001) examined It-extraposition only, in which the definiteness restriction does not come into play. However, it is interesting that a syntactic configuration should be chosen based on discourse context, specifically to keep the discourse coherent.

De Kuthy (2002) examined the NP-PP split in German. This phenomenon has also been previously analyzed from a syntactic point of view. De Kuthy showed that the acceptability of this construction depends on the discourse context. A PP modifying an NP may be separated from that NP as long as the information status of the PP does not match that of the NP. If one of the constituents expresses background information, the other must express new information.

The present study, together with Miller (2001) and De Kuthy (2002), suggest that other marginal constructions in language may be due to pragmatic factors, rather than syntactic constraints.

References

BALTIN, MARK R. 1981. Strict Bounding. *The Logical Problem of Language Acquisition*, ed. by C.L. Baker and J. McCarthy, 257-259. Cambridge, MA: MIT Press.

BALTIN, MARK R. 1983. Extraposition: Bounding vs. Government-Binding. *Linguistic Inquiry*, 14.155-162.

BALTIN, MARK R. 1984. Extraposition Rules and Discontinuous Constituents. *Linguistic Inquiry* 15.157-163.

CULICOVER, P.W. & ROCHEMONT, M.S. 1990. Extraposition and the Complement Principle. *Linguistic Inquiry* 21.23-47.

DE KUTHY, K. 2002. *Discontinuous NPs in German – A Case Study of the Interaction of Syntax, Semantics and Pragmatics*. Stanford, CA: CSLI Publications.

GUÉRON, J. 1980. On the Syntax and Semantics of Extraposition. *Linguistic Inquiry* 11.637-678.

GUÉRON, J. & MAY, R. 1984. Extraposition and Logical Form. *Linguistic Inquiry* 15.1-31.

HEIM, I. R. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. University of Massachusetts, Amherst: Doctoral dissertation.

HUCK, G. J. & NA, Y. 1990. Extraposition and Focus. *Language* 66.51-77.

HUCK, G. J. & NA, Y. 1992. Information and Contrast. *Studies in Language: Intl. Journal sponsored by the Foundation "Foundations of Language"* 16.325-334.

MILLER, P. 2001. Discourse constraints on (non)extraposition from subject in English. *Linguistics* 39.68-71.

PIERREHUMBERT, J. & HIRSCHBERG, J. 1990. The Meaning of Intonational Contours in Discourse. Chapter 14 in *Intentions in Communication*, ed. by J. Morgan, P.R. Cohen, & M.E. Pollack, 271-323. Cambridge: MIT Press.

PRINCE, E. 1981. Toward a Taxonomy of Given-New Information. *Radical Pragmatics*, ed. by Peter Cole, 223-55. San Diego: Academic Press.

QUIGLEY, JOHN B. 2003, February 8. The consequences of invading Iraq are dire. The Columbus Dispatch, A6. Columbus.

REINHART, T. 1983. *Anaphora and Semantic Interpretation*. London: Croom Helm, Ltd.

ROBERTS, CRAIGE 1996. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *OSU Working Papers in Linguistics **49***: *Papers in Semantics*, ed. by J-H. Yoon & A. Kathol. Columbus, OH: The Ohio State University.

ROBERTS, CRAIGE 1998. Focus, the Flow of Information and Universal Grammar. *Syntax and Semantics Volume 29: The Limits of Syntax*, ed. by P.W. Culicover and L. McNally. San Diego: Academic Press.

ROBERTS, CRAIGE 2002. Uniqueness in Definite Noun Phrases. Manuscript, The Ohio State University, Columbus, OH.

ROCHEMONT, M. 1986. *Focus in Generative Grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

SELKIRK, ELIZABETH O. 1984. *Phonology and Syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

SELKIRK, ELIZABETH O. 1995. Sentence Prosody: Intonation, Stress, and Phrasing. Chapter 16 in *The Handbook of Phonological Theory*, ed. by J. A. Goldsmith. Cambridge, MA: Blackwell Publishers.

WITTENBERG, K. 1987. Extraposition from NP as Anaphora. *Syntax and Semantics, Volume 20: Discontinuous Constituency* ed. by G.J. Huck and A.E. Ojeda. London: Academic Press.

A CONSUMER'S GUIDE TO CONTEMPORARY MORPHOLOGICAL THEORIES

Tom Stewart
Truman State University

1. Preface

When a linguist goes in search of "the best" theory in some domain, it is important for him or her to ask "best for what purpose?" Individuals will of course prefer to work with an approach that makes sense to them, one that complements—or at least does not contradict—other assumptions they hold with respect to grammatical theory. This is not to say, however, that "it's all relative," that theory choice is solely an aesthetic selection from among notational variants. The relative quality of a theory can be evaluated on empirical grounds, based on the particular predictions that the theory's assumptions entail.

In this study, intended as a consumer's guide, so to speak, I give characterizations of various morphological theories currently used by different linguists. For the reasons stated above, in the theory characterizations to follow, I have chosen to begin each section with a table by which one may readily compare and contrast some of the guiding assumptions in each theory. Criticism of each framework, both theoretical and empirical in nature, will be presented where available, and replies or adjustments in the literature will follow. A bibliography of leading publications for each framework concludes the respective section.

2. How to Interpret a Table

The first continuum, *morpheme-based* versus *word/lexeme-based,* concerns the basic units assumed to organize morphological activity. In a strongly morpheme-based

Tom Stewart. A Consumer's Guide to Contemporary Morphological Theories. *OSUWPL* Volume 58, Fall 2008, pp. 138-230.

theory, the *morpheme* is the atomic meaningful unit, and morphology is about how morphologically complex expressions come to have the meanings and attributes they do, thanks to these morphemic units. Morphological analysis, therefore, is analysis down to, and up from, the level of constituent morphemes. In a strongly word/lexeme-based theory, the *word* (subject to definition) is the organizing principle of morphological structure. Analysis below the word level, especially that which takes derived bases back to source roots is not of primary concern to such a theory. Derived lexical items may owe some part of their lexical character (semantics, grammatical category, phonology) to their source roots, but in word/lexeme based theories, the exhaustive analysis into parts is often (but not always) seen as an excess, a hypersegmentation which goes beyond the requirements of syntax at least, since rules of syntax are generally presumed not to care about the internal constituency of the words they manipulate (the Lexicalist Hypothesis (Chomsky 1970)).

The second continuum, *formalist* versus *functionalist*, has to do with a broader perspective on what linguistic theory and analysis are supposed to accomplish. This is therefore a fundamental distinction which may shape the types of phenomena one chooses to address, what sort of data constitute real counterexamples to theoretical claims, and what role evidence external to the grammar (language acquisition, psycholinguistic testing, sociolinguistic patterning, and typological evidence) is given in support or as counterevidence to a theory. Formalist approaches focus primarily on rules, constraints, and units which are particular to language structure, usually with the goal of capturing "all and only" those generalizations relevant to the characterization of linguistic competence. Functionalist approaches are interested more in contextualizing language as cognitively and socially grounded behavior. Functionalist analyses tend to be more tolerant of gradient behaviors, appealing often to constraint satisfaction, trade-offs, relative frequencies, etc. For these reasons in particular, functionalist discussion draws formalist fire for being fuzzy, vague, and indeterminate. Formalist approaches receive criticism for being artificially "neat" in the data they consider, abstracting over variation, and ignoring the language user as part and parcel of the language-use equation. This distinction might equally be termed "micro" versus "macro" theorizing, respectively.

The third continuum, *in grammar* versus *in lexicon*, refers to the "location" of morphology in the architecture of a grammar. Theories which place morphology in the grammar may do so as its own component or sometimes distributed among independently motivated components, typically syntax or phonology. Much work in generative morphology has taken an 'in-grammar' approach to morphology, according very little role indeed to the lexicon, other than as a repository for idiosyncrasy (e.g., Di Sciullo and Williams 1987). An approach which puts morphology in the lexicon, on the other hand, has a very different perspective on just what the grammar does. The lexicon is a repository for most if not all lexical knowledge, predictable or not, and the complex lexical entries interact with grammatical structures in as many distinct ways as grammatical structure requires. What the former approach gains in reducing redundant lexical listing it loses in its failure to naturally characterize inflectional paradigms, for example. The latter approach, on the other hand, presents the opposite problem, a rich and rather redundant lexicon, but an accordingly streamlined grammar. Issues of storage versus computation are relevant at this level, with computation being the focus of 'in-grammar', and storage the emphasis of 'in-lexicon'. This is not necessarily an either-or proposition, however, since it is possible, according to the Split-Morphology Hypothesis, e.g., to handle derivational morphology in the lexicon and inflection post-syntactically (Anderson 1982). Theories assuming some version of the Split-Morphology Hypothesis will be marked in the center column of this continuum.

The fourth continuum, *Phonological formalism* versus *Syntactic formalism,* is not entirely independent of the third, but neither is it fully predictable from it. Approaches which place morphology in the grammar will, for consistency's sake, tend to formalize morphological rules to be as similar as possible to the rules assumed for the relevant adjacent component of grammar. Word-Syntax (Lieber 1992; Di Sciullo & Williams 1987), for example, makes great use of hierarchical structure and percolation, whereas Lexical Morphology (Kiparsky 1982a) formalizes lexical and post-lexical phonological rules in similar ways, distinguishing them by domain of application, rather than making a formal distinction in rule construction. Approaches which place morphology in the lexicon, yet nonetheless use a formalism akin to some component of the grammar may do so for expository purposes, but there is less motivation on theoretical grounds to do so. A mark in the center column for this continuum will indicate a qualitatively distinct rule format for morphology.

Borrowing terminology from Stump (2001:2-9), the fifth continuum, *incremental* versus *realizational*, focuses on the input/output conditions on the morphological component. A choice along this dimension will entail a very different picture of just what morphology "does." In an incremental approach, the meaning and other attributes of morphologically complex expressions are built up gradually as a more or less additive process (thus 'incremental'). This addition can happen, metaphorically speaking, either through the concatenation of morphemes or through the application of morphological rules. From this perspective, every attribute or element of meaning not present in a lexical root must be added to that root in the morphology.

In a realizational approach, by contrast, the input to morphology is more abstract. A lexical base (whether root, lexeme, or lexical stem) and some set of properties (appropriate both to that base and to the context in which the complex expression finds itself) jointly determine the morphophonological 'spell-out' of the fully inflected word in that context. Incremental methods are most appropriate in describing languages where there is an overt exponent for all and only the meanings and attributes of the word in question, e.g., especially highly agglutinative languages, like Turkish. Where the overt morphology does not match one-to-one with the meanings and grammatical functions of the word as a whole, i.e., where the overt morphology either overdetermines or underdetermines the whole, an incremental approach will be forced into either abstract or *ad hoc* elements in the analysis, either phonetically null (zero-)morph(eme)s, or rules which apply but effect no discernable phonological change (roughly, a zero-derivation).

There are many undesirable consequences of countenancing null elements in an analysis, even if their "distribution" is constrained, not the least of which is learnability, i.e., how does a child know a zero when s/he "hears" one, and how does a child recognize which zero s/he "heard"? Realizational frameworks can, in principle, avoid the zero-morph trap because the association between a word and its morphosyntactic features is the input to the morphology, i.e., the meaning 'licenses' the presence of particular exponents, the meaning does not 'depend' on the introduction of meaningful pieces. What an incremental approach gains in concreteness of representation, it loses when faced with morphemes without meaning (so-called 'empty morphs') or meanings without exponents, as mentioned above. On the other hand, what the realizational perspective gains in formal versatility, i.e., empirical coverage, it loses in its apparently unnecessarily complicated treatment of the most transparent morpheme-like instances of morphological composition. This would not be such a problem if edge affixation were no more common than other types of morphological marking. As it is, however, the disproportionate amount of concatenative morphology found cross-linguistically looks rather like an accident on the realizational approach.

Now, it is certainly possible to think of more theoretical distinctions that one could use in the classification of morphological theories, and likewise it might also be possible to make do with fewer distinctions. This set of five, however, allows for some interesting similarities and differences to come out, and the dimensions are substantial enough that any given linguist can quickly identify the theory or theories which best match their own predispositions.

Accordingly, in what follows, I survey 13 current theories of morphology: A-Morphous Morphology, Articulated Morphology, Autolexical Syntax, Categorial Morphology, Distributed Morphology, Lexeme-Morpheme Base Morphology, Lexical Morphology and Phonology, Natural Morphology, "Network Model", Network Morphology, Paradigm Function Morphology, Prosodic Morphology, and Word Syntax, and discuss each according to the classifying features just described. I then, in the appendices, show how each theory would handle the facts from two well-known morphological phenomena: inflection of nouns in Scottish Gaelic, and verb agreement in Georgian.

3. A-Morphous Morphology

Morpheme-based					☑	**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar			☑			**In lexicon**
Phonological formalism		☑				**Syntactic formalism**
Incremental					☑	**Realizational**

Many of the assumptions which coalesced in the form of Anderson (1992) are laid out in an extended series of articles stretching back over at least fifteen years. While continually pointing out a resurgence of interest in morphology as a field of inquiry in linguistics after an extended drought, Anderson just as repeatedly asserts (1977:17) that "the notion of a separable morphological 'component' is probably untenable." The name A-Morphous Morphology is intended to directly challenge the traditional role of the morpheme as a primitive in word structure, focusing instead on lexical roots or stems, and operations applied thereto. Halle and Marantz (1993:112-13) are baffled by this move and fundamentally misappraise the claims made by Anderson (1992), assuming that morphemes must be included in Anderson's structural representations, the morphemes' phonological representations then removed, and (incredibly) reinserted later in the derivation, for how else could morphosyntactic information get into a syntactic structure than through the insertion of meaningful pieces (i.e., morphemes), however temporarily or covertly? Explicitly acknowledging the asymmetry between lexical roots and morphological operations constitutes something of a paradigm shift (no pun intended) in morphology, and between the word-based and morpheme-based camps is quite a large chasm (cf. Aronoff 1976, 1994; Anderson 1988b:162-64; 1992:48-72; Zwicky 1992:338).

In A-Morphous Morphology (and its immediate predecessor, the Extended Word and Paradigm (EWP)[1] framework, from which there is little discernable break), primary attention is given to inflection (Anderson 1977, 1988a, 1992:ch. 4-6). In EWP, for example, derivation, cliticization, and compounding each get one chapter to inflection's three chapters.

The lexicon in A-Morphous Morphology is not the minimal "idiosyncraticon" (Zwicky 1992:338) inherited from Bloomfield (1933:269), but rather an un-list-like collection of knowledge that a speaker may have, governed by rules of varying generality (Anderson 1992:183). Anderson takes the relevant word-like unit to be the *stem* (using the word *lexeme* rarely, if at all), derivation to be a lexicon-internal phenomenon (cf. LMBM, below; contrast PFM, also below), and inflection to "fall 'outside the lexicon' in the sense that [inflectional rules] represent knowledge not of particular words, but rather of the form taken by words as a consequence of the syntactic structure in which they appear" (Anderson 1992:183-84). The model of the grammar (see Anderson 1982:594), then, entails the Split Morphology Hypothesis (Perlmutter 1988; Booij 1993; cf. Beard 1995), with inflection effectively 'in' the syntax and derivation in the lexicon.

An interesting perspective may be gained through the comparison of the attributes of inflectional and derivational word-formation rules (WFRs) in A-Morphous Morphology (Anderson 1992:123, 185):

<u>Inflectional WFRs</u> are characterized by:

a. A formal Structural Description, specifying conditions on **S** (the lexical stem in the input) and conditions on **M** (the aspect(s) of the morphosyntactic context realized by the particular WFR); and

b. A formal Structural Change, which may involve "not only affixation but also other phonological changes such as metathesis, substitution, deletion, etc." (123);

whereas <u>Derivational WFRs</u> are characterized by:

c. A formal Structural Description, specifying the class of input stems the rule can apply to and any additional conditions;

d. A formal Structural Change, specifying the alteration the rule performs in creating the phonological form of the derived stem from the form of the input stem;

e. A Syntactic Structural Description and Change; and

f. A Semantic Structural Description and Change.

[1] The name EWP is perhaps at odds with the limited usefulness Anderson (1992:79) ascribes to the notion *paradigm*.

Inflectional WFRs are not additive, or 'invasive', beyond the level of the phonological form (contrast the rules of AM, below). Because of the inflectional feature content of a particular syntactic terminal node, its associated morphosyntactic representation (MSR) 'licenses' (not Anderson's term) the introduction of inflectional exponents via inflectional WFRs, i.e., inflection *presupposes* the morphosyntactic representation, rather than *creating* it, as the metaphor goes in morpheme-based frameworks. Derivational WFRs can potentially effect a broader range of changes in the input, but this is done without reference to particular (morpho-)syntactic contexts.

In contrast with the Word Syntax and LMBM positions (both below), A-Morphous Morphology accords little significance to word internal derivational hierarchical structure, since syntax apparently does not have or need access to that sort of information (the Lexicalist Hypothesis (Chomsky 1970)). The intricate relationship of inflection and syntax, however, leads Anderson (1992:84) to conclude that the Lexicalist Hypothesis must be relaxed in inflection, although it may safely be assumed to hold in derivation. Compounding is a hybrid case in A-Morphous Morphology (Anderson 1992:292), because there is motivation for a syntactically-accessible hierarchical structure in headed compounds, in contrast with ordinary derivation: "The formation of compounds seems to involve a genuinely syntactic combination of lexical elements below the level of the word."

Probably the most noteworthy and controversial aspects of A-Morphous Morphology have to do with the implementation of inflectional WFRs so that the "inflectional formatives of a word [are placed] in their correct relation to one another" (Anderson 1992:123). The null hypothesis is that no special ordering mechanisms will be required, and an unordered list of morphosyntactic features will be sufficient to direct the phonological realization automatically. This cannot be the case, however, for two reasons:

1. one and the same inflected word may bear two or more distinct values for the same morphosyntactic feature (e.g., agreement in Person and/or Number for multiple arguments of a verb)(Anderson 1977:23), and
2. of two or more contextually motivated inflectional rules, there are numerous cases cross-linguistically where only a subset of these rules actually apply, implying a disjunctive relation between particular rules (Anderson 1986:7-8).
   Rather than a full conjunctive application of all applicable rules, or the more limited (but still reasonable) expectation that every feature be realized at least once in the inflected form, the actual details of realization require that some provision be made in an adequate grammar for rule ordering.

In response to the first issue, Anderson (1977:21) proposes that words in syntactic contexts have *morphosyntactic representations* (MSRs, mentioned above), i.e., inflectional feature matrices whose contents are internally unordered by default, but which gain layers just in case "a rule of the grammar assigns features to an element, and that element already carries specifications for those features" (see also Anderson 1992:94). For example, an MSR with complex [–F +G], if further assigned the value [+F], will not unify to *[+F –F +G], but rather to the layered structure [+F [–F +G]], with any and all duplicate features (whether they bear contrastive values or not) appearing in an outer layer with respect to earlier-assigned and unduplicated feature-value pairs. Layering is in principle unlimited, but there is apparently no practical need for more than three layers in any one MSR. Similarly, there is no overt constraint that layering is limited to agreement (or so-called Φ-)features and so, to the extent that layering is not invoked except in cases of repeated or conflicting person, number, gender, case, or animacy specifications, this generalization is missed. If layering is triggered during the sequential creation of an MSR (it must be sequential in order to determine, in cases of duplicate features, which instance is inner, and which outer), the inherent features of a possessed noun should be inner with respect to those of a possessor and correspondingly, the internal arguments of a verb should be inner with respect to external arguments. In Anderson (1977:21), the token offer at an alternative formalization is made, more specific features "[±1st person possessor], [±plural possessor], etc., but this would be of little interest." This is true, certainly, and a fairly *ad hoc* response to the situation, but it is owing to the binary nature of features in A-Morphous Morphology (cf. *n*-ary features in, e.g., PFM, below).

In the case of Georgian "inversion" (e.g., Harris 1981, 1984; Anderson 1984), Anderson (1992:141-56) proposes a "purely morphological transformation" whereby an inner layer of the MSR is moved to an outer position. Thus, inflectional WFRs which happen to be keyed to particular layers (i.e., have particular layers specified as part of their Structural Description) will be effectively 'tricked' into applying to a different layer, producing the observed agreement marking mismatches (see Appendix B for some discussion). In order to 'force' features into particular layers, however, Anderson (1992:147) invokes a dummy placeholder, apparently the only instance of a zero in A-Morphous Morphology. In its favor, the zero is purely formal, and has neither semantics nor reference (cf. zeroes in DM, Word Syntax, below). Such uses of the MSR device allow A-Morphous Morphology to engage in a measure of "virtual Relational Grammar," while technically avoiding a backwards reach into syntax proper.

On the issue of disjunctive rule ordering, A-Morphous Morphology relies on a version of the Elsewhere Condition (*EC*; Anderson (1969), Kiparsky (1973), not to mention Pāṇini). The Pāṇinian Principle, often mistaken for the full EC, is a precedence principle, whereby the most narrowly defined of a set of competing rules (alternatively, morphemes; see DM, below) precedes the other competitors in application, and thus rules may apply conjunctively or disjunctively and still respect the Pāṇinian Principle (cf. PFM, below, in which disjunctive application is derived independently). Anderson's (1986:4, 1992:132) EC formulations include a (weak) disjunctivity rider:

> "...whenever one rule is more specific than another in the sense that the forms subject to the first constitute a proper subset of those subject to the second, the application of the more specific rule precludes the later application of the more general, less specific one."

Anderson (1992:132, fn. 30) notes that this formulation entails disjunctive application only if the more specific rule applies, and applies <u>first</u>. Subtly, therefore, this EC allows four of the five logically possible outcomes of trying to apply two rules, a specific one (**S**) and a general one (**G**) (Janda (n.d.):3):

——	(neither applies)
G	(only G applies)
S	(only S applies—G, which must be applicable given the 'proper subset' definition above, is blocked by the EC, i.e. *S, G)
G, S	(G applies first, S is not blocked by the EC, and therefore applies as well)

This condition is claimed to account for disjunction between rules, between a stem and a rule, and between stems as well (Anderson 1986:4, 1992:133-34). This principled disjunctivity is not empirically justified, however[2], and so Anderson (1992:129) adds not only the device of stipulated *rule blocks* (the rules within which blocks may, but need not, realize the same or similar inflectional properties), but also the option of extrinsic (*ad hoc*) rule ordering within these blocks, substantially weakening the predictive power of the account overall. The result is an observationally adequate description, but there is little insight into why the observed order obtains rather than any number of readily describable alternative patterns. Similarly unaccounted for is the tendency for disjunctively related exponents to have similar if not identical linear placement restrictions with respect to the stem (cf. PFM, below).

Although it makes rather less use of the *word* and the *paradigm* than one might expect from a "word-and-paradigm" type of theory, A-Morphous Morphology makes a number of important and provocative contributions in its denial of the relevance of the morpheme as a basic unit of language. A-Morphous Morphology borrows some trouble by adopting the Government and Binding (GB) approach to syntax which is not particularly morphology-friendly. Unlike LMBM and DM (both below), however, A-Morphous Morphology does not focus on the formal interface as much as the logical necessities such an interface would entail, and is therefore somewhat vague. Anderson borrows Chomsky's (1981:92) dismissive phrase "merely a matter of execution" on two occasions—instead of taking a stand on when "lexical insertion" should happen (i.e., at D-Structure (DS) or S-Structure (SS))(1992:91, fn. 16), and when tentatively considering whether Generalized Phrase Structure Grammar (GPSG) might not have a better account of agreement than GB (1992:109). Of course, this is a theory of morphology, and so some hand-waving with respect to syntactic-theoretic detail is no great flaw. The architecture of grammar given in Anderson (1982:594) is much more specific, and already in that article it is suggested that SS is the locus of "lexical insertion" and that derivation was 'in the lexicon', but the diagram clearly includes "inflection" in a component marked "Phonology." This is misleading, however, because MSRs are present at SS, however, so 'inflection' here must refer to feature realization, the application of inflectional WFRs.

As an analytic tool, the formalism of A-Morphous Morphology is generally transparent, and accommodations are made for both affixal and more processual operations. Trouble spots are generally restricted to truly controversial areas (e.g., the EC and language-specific ordering). That compounds, clitics, and morphophonology are treated as well in Anderson (1992) is especially helpful, although it remains clear that

---

2 For counterexamples, see Janda and Sandoval (1984); for extended discussion and further counterexamples, see Janda (n.d.).

Anderson's answer to his own (1982) question "Where's morphology?" is an ambivalent "everywhere, yet nowhere," that is, in many places, not one single place.

Anderson, Stephen R. 1977. On the formal description of inflection. *Chicago Linguistic Society* 13.15-44.

Anderson, Stephen R. 1982. Where's morphology? *Linguistic Inquiry* 13(4).571-612.

Anderson, Stephen R. 1984. On representations in morphology: Case, agreement, and inversion in Georgian. *Natural Language and Linguistic Theory* 2.157-218.

Anderson, Stephen R. 1986. Disjunctive ordering in inflectional morphology. *Natural Language and Linguistic Theory* 4.1-31.

Anderson, Stephen R. 1988a. Inflection. *Theoretical Morphology*, ed. by Michael Hammond and Michael Noonan, 23-43. San Diego: Academic Press.

Anderson, Stephen R. 1988b. Morphological theory. *Linguistics: The Cambridge Survey* [vol. 1], ed. by Frederick J. Newmeyer, 146-91. Cambridge: Cambridge University Press.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

4. Articulated Morphology

Morpheme-based				☑		**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar		☑				**In lexicon**
Phonological formalism			☑			**Syntactic formalism**
Incremental	☑					**Realizational**

Of all the frameworks to be considered here, Articulated Morphology (AM) stands with the least representation in the literature. It is noteworthy, however, because it is set out in significant detail in Steele (1990, 1995), and it has a unique combination of attributes. It is "amorphous" in the sense that affixation is secondary with respect to lexical stems, yet AM is incremental in that no information (beyond the root) is present in a complex construction that the application of some rule did not put there. Operations are parts of rules, and these rules are applied so as to create an output of the intended sort. AM is limited to the inflectional domain (Steele 1995:261), and so some of the questions a more general theory faces are not addressed here. Since the framework presupposes the operation of derivation, and since the rules are construed as applying to stems, AM may fairly safely be characterized as word/lexeme-based in its orientation. Inflectional rules in AM are strictly information-adding—they cannot replace or delete information already present in a morphological object to which they apply.

Because rules must be sensitive to the informational content of their inputs, the question of extrinsic ordering of rule application is neatly sidestepped. A rule applies to an object of a certain type and augments its informational content, possibly modifying its phonological shape in the process (262). Since AM is not morpheme-based, it is no problem to construct a rule which adds information but performs no phonological operation whatsoever. This sort of rule is an identity function on the phonological level, so a "zero-morph" is not introduced into the structure, which results in the following metatheoretical bonus (Steele 1995:288-89):

Potawatomi: rule adding [singular] in the transitive animate

X		*X*
[ ] [ANIM: +]	→	[ ] [ANIM: +]
Person: -speaker, -hearer		Person: -speaker, -hearer
		Number: singular

AM is based on three principles (Steele 1995:271-72):

1. Associate a stem with that informational subpart specifically identified with the stem, in the absence of the inflectional operations at issue (i.e., stems are informationally reduced);

2. Analyze inflectional operations as adding information to the morphological object they are performed on (i.e., rules result in feature specifications; they may add values for as-yet-unspecified features present in their input, they may add feature-value pairs not at all present in the input, or they may do both at once);

   and

3. Classify morphological objects according to the kind of information they present, and classify operations according to the kind of object they are performed on and the kind of information they add.

Steele claims that these principles allow "the organization of a morphological system in AM [to be] entirely *intrinsic*, driven by the fact that the operations effecting phonological modifications also introduce a distinctive kind of information to a particular kind of morphological object" (272-73). While other theories try very hard to prevent access to the internal structure of bases/stems beyond the outermost layer (LM&P, see below), AM invests its rules with rich conditions on application, stating what rules must have already applied before the rule in question may apply.

The AM approach may work so long as, for example, exponents of Person occur consistently inside a rule introducing Number. The rule introducing Number could (and would) be written with a reference to the pre-existing specification for Person. If some Person exponents appear before and others after, this would effectively block the application of the rule introducing Number in the latter cases. Responses to this situation could be of at least two sorts—either the introduction of a second (back-up) rule introducing the Number information without the crucial Person information in the base, or else rendering the Person information in the base optional. In the latter case, however, the intrinsic ordering effect is lost, since the rule could then apply optionally before or after the rule introducing Person information. (Georgian comes to mind—see Appendix B—as an example of a language where exponents of the same basic category appear in radically different positions with respect to the stem.) Steele (1995) avoids dealing with prefixal subject Person markers in Potawatomi, for example, by claiming that these are likely proclitics rather than inflectional markers, and thus outside the domain of inflection (273, fn. 14; 278-79).

AM rules formally consist of a *domain* (input conditions) and a *co-domain* (output conditions), as essentially a before-and-after photo set of the representation (Steele 1995:276). There is no acknowledgement that the rules are consequently highly redundant, that if one subtracts the domain contents from the co-domain, the difference would be simply an inflectional morpheme. The contrast with A-Morphous Morphology is quite apt, since its realizational emphasis on fully-specified representations and minimal rules is exchanged in AM for minimal representations and enriched inflectional rules. Steele makes a virtue of AM's capacity for allowing "in principle, any number of distinctions in the morphological types" (279)—Potawatomi requires three or four, depending on whether the person prefixes/proclitics are in or out of inflection (277, 279):

> <u>Stem</u>: "a morphological object lacking Number"
> <u>Extended Stem</u>: "a transitive object that has one fewer N[umber] attributes than arguments"
> <u>Word</u>: "an object whose arguments are all associated with the property of Number"
> <u>Indexed Form</u>: "a morphological object where both Person and Number are saturated"

These definitions are not only specific to the Potawatomi language; they are specific to <u>verbs</u> within the Potawatomi language. Nouns and adjectives would certainly require different definitions within the same language, and all of these are subject to cross-linguistic redefinition. Languages with fusional agreement markers would, presumably, rule out the extended stem type of object since the same rule would always have to introduce both Person and Number at once. The result is that there is no definition of *stem* and *word* independently available, separate from language-particular systems. If operations are classified "according to the kind of object they are performed on and the kind of information they add" (Steele 1995:272, part of principle 3 above), this inextricably ties the operations to language-specific details as well, and thus both the operations and the morphological object types lack all but the most abstract generality. This is not a descriptive advantage. A-Morphous Morphology is derided for being able to express the following "simple generalizations" (279-80):

1. Suffix_1: stem → stem
2. Suffix_{2a}: transitive stem → extended stem
3. Suffix_{2b}: stem → word

4. Suffix$_3$: $\left\{\begin{matrix}\text{intransitive stem}\\ \text{extended stem}\end{matrix}\right\} \rightarrow \text{word}$

Steele apparently does not take into account that these "simple generalizations" are effectively telling us what each suffix does, i.e., this is a morpheme-based set of statements. It is neither surprising nor a defect in A-Morphous Morphology that it cannot make these statements.

AM's intrinsic rule ordering would be more impressive if there were any relation to the distribution of inflectional properties as handled by the syntax. The proper application of rules in a derivation presupposes knowledge of the properties of the goal state, the triggers for the application of these rules. Once they get started, as promised, the rules will not apply until they are supposed to, sequentially speaking, but there is feigned ignorance of the overall goal state that any given inflectional derivation is intended to achieve. The logic of this incremental (re-)creation of a known end-state in the morphology is directly attributable to the too-literal interpretation of the derivation metaphor[3]. Other problems include the necessary introduction of stipulated negative conditions on the domain for a significant number of the rules proposed (Steele 1995:291, 294, 296), floating feature values where problematic feature unifications are foreseen (287), and *ad hoc* 'avoidance strategies' to block formally predicted but actually unattested effects of rules, e.g., the AVOID 3RD PERSON strategy (287), something that the inflectional component should never have to do, unless it were trying to take over functions more appropriately located in the semantics and syntax.

The formal and potential empirical difficulties that AM faces (even in the analysis of Potawatomi, the data set which was chosen specifically to show off the advantages of the theory) go a long way toward inadvertent self-incrimination. AM offers little prospect of yielding cross-linguistically comparable descriptions, but if a linguist is interested in describing individual languages on their own terms, without reference to meaning or structure outside of inflection, then AM might be suitable.

Steele, Susan. 1990. Pass it on! A combinatorics-based approach to feature-passing. *Chicago Linguistic Society* 26.407-21.

Steele, Susan. 1995. Towards a theory of morphological information. *Language* 71(2).260-309.

5. Autolexical Syntax

Morpheme-based		☑				**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar				☑		**In lexicon**
Phonological formalism				☑		**Syntactic formalism**
Incremental		☑				**Realizational**

[3] Word Syntax also pays special attention to overcoming the same concern.

Autolexical Syntax directly addresses the interface between syntax and morphology. Sadock (e.g., 1988) has assembled a modular, but or non-serial modular non-serial, theory of grammar, in which semantic, syntactic, and morphological modules operate separately, yet simultaneously constrain the class of well-formed expressions in a language. In this way, a potential expression may be semantically interpretable, but not syntactically parsable, or vice versa, and in either case the expression would be ruled out. The same is true with respect to morphological structure. The suggestion is, then, that one can 'troubleshoot' any ungrammatical expression and trace the source of the problem to one or more of the components. More recent work in the framework (e.g., Singer 1999), however, has invoked the violable-constraints approach of Optimality Theory to allow for variable effects of violations of the requirements of the three components, undercutting the restrictiveness of the original model. Although Sadock uses the term *lexeme*, he never directly defines it. He clearly includes affixes, clitics, bound roots, and stems in this category, and thus his definition must be something closer to the traditional *morpheme*, although morphologically complex stems are treated as units by the rules of morphology. This stand puts Autolexical Syntax at or at least near the morpheme-based pole of the first continuum.

Sadock (1991) proposes a subsystem (not a *module* in the sense of Fodor (1983)) which he calls the *Interface*. This subsystem has "direct access to all varieties of grammatical information" and uses this information to coordinate "the several representations produced by the autonomous modules"(36). The lexicon is a part of the interface subsystem, and it, too, does not constitute a module in its own right. A grammatical (i.e., well-formed) expression of any size in a language corresponds to a triple $\{r_{syn}, r_{sem}, r_{mor}\}$ of acceptable outputs from the three components posited in this grammar (20; cf. triples in Categorial Morphology, below). The lexical entry for any "lexeme" in this theory is a set of three representations, one for each component, and these three representations define the grammatical use of the "lexeme" (30). For example, *dog* would have the following lexical entry:

dog
syntax = N[0]
semantics = F-1
morphology = N[-0]

This means that *dog* is a noun, bar-level 0, a function of one variable on the semantic level, and a noun stem, from the point of view of morphology. Somewhat counter-intuitively, the minus (-) on a morphological bar-level representation is simply a marker of the morphological domain of analysis, i.e., a greater negative integer does not mean a smaller morphological unit, but rather a larger one, such that [-0] is a stem, [-1] is a(n inflected) word, and [-2] is a "super-word", i.e. a word plus an attached clitic element. Lexical stems and larger expressions—those "placed" by the syntax—have a specific syntactic representation. Affixes, on the other hand, have a semantic and a morphological representation, but no syntactic representation, therefore they are inaccessible to (not manipulable by) the rules of syntax. While Sadock does not make the claim that all morphology is concatenation, he sets aside non-concatenative processes to be handled (at another time) autosegmentally, in the manner of Prosodic Morphology (McCarthy 1981, see below).

The formal focus in Autolexical Syntax is clearly on (primarily binary-branching) tree structures, and the representations within lexical entries are interpretable only with respect to such tree structures. Much of Sadock (1991) is devoted to cases where the structure implied by one component is at odds with that of another, such as the cases of clitics, which attach morphologically to constituents other than syntax or semantics might suggest (48), or incorporation, whereby arguments, which are syntactic atoms in the general case, are morphologically proper subparts of other words (79). The precise details of these analyses are not important here, besides the general clue that the mixed behavior is ironed out within the simultaneous triple representation, as facilitated by principles of the Interface subsystem. As Sadock rightly points out, such a "simultaneous treatment ... is precluded in a hierarchical model of grammar, where the output of one component is modified by the next component downstream" (51). Both clitics and incorporation have been chronic sources of aggravation and fascination in grammatical theory, and so it seems a genuine advance to have a fairly unified account of them.

Sadock (1988:281) proposes a classification scheme for "lexemes":

	I	II	III	IV	V	VI	VII	VIII
Syntax	+	+	+	+	–	–	–	–
Semantics	+	+	–	–	+	+	–	–
Morphology	+	–	+	–	+	–	+	–

where + means "has a representation in that module" and – "has no representation in that module." Of these eight classes, class VIII is ruled out in principle as being empty in every regard, a victim of "intermodular suicide" (281). Sadock identifies instances of classes I, III, V, and VII, i.e., those classes which have at least a morphological representation. Class I is exemplified by the ordinary lexical stem, like *dog* above, with a representation at all three levels. Class III is exemplified by pleonastic and purely functional elements like dummy *it*, infinitive *to*, and complementizer *that* (280). Class V is a derivational affix, which is semantically a property expression, a function on the meaning of the stem, e.g., the German diminutive *–chen*, as in *Hühnchen* 'chicken, pullet', but which has no independent representation in the syntax (281-82).

-chen
syntax = nil
semantics = Prop
morphology = N[M1, Ntr]/N

'M1' refers to a particular morphological rule in Sadock (1988:274), category-changing derivation on a stem, and the slash formalism is parallel to that used in Categorial Grammar. The 'Ntr' condition on the affix is a condition imposed on the output of affixing *–chen* to a stem, i.e., the result will be of neuter gender. In this sense, Sadock claims, the affix is the head of the construction (akin to the ideas of Williams 1981, but with a more substantial, rather than positional, definition of 'head'). Class VII is a stem forming element like the *–s* in the non-head of certain German compounds, e.g., *Freiheitskämpfer* ‘struggler for freedom’, where the *–s* is not simply the genitive marker (gen. *Freiheit*).

> -*s*
> syntax = nil
> semantics = nil
> morphology = N[M2, CF]/N[F]

Such elements have no syntactic or semantic representation, they are present in the morphology only, for the creation/marking of a combining form (CF), a stem (formed by lexemes of class N[F]) for use in compounding only, by means of morphological rule 2 (M2; Sadock 1988:247, 282). A more complete classification takes the five attested classes of lexemes and contrasts them in terms of the general formal content of their representations in the three modules (289):

	Stem	Inflectional	Derivational	Incorporating	Clitic
Syntax	X[0]	———	———	$[_{Y[1]}X[2]\underline{\ \ }]$	$[_{Zn}X[2]\underline{\ \ }]$
Semantics	Property or Relation	———	F(X[0])	F(X[2])	F(X[2])
Morphology	X[-0]	$[_{X[-1]}X[-0]\underline{\ \ }]$	$[_{Y[-0]}X[-0]\underline{\ \ }]$	$[_{Y[-0]}X[-0]\underline{\ \ }]$	$[_{W[-2]}Y[-1]\underline{\ \ }]$

These formulations are intentionally abstract on Sadock’s part, and the variables allow for a range of instantiations of each type.

A more satisfying contextualization of the “lexemes” is to be found in the set of intermodular defaults based on Sapir (1921), at least in spirit. These are predictions that hold in the general case between a representation in one module and that in another.

1. Prop or Rel → X[-0]
   If a lexeme is semantically a property or a relation, then it is a morphological stem.
2. X[-0] → X[0]
   If a lexeme is a morphological stem, then it is a syntactic atom.
3. X[M1] → semantics = F(Y[-0])
   If a lexeme is subject to morphological rule 1 (as -*chen*, above), then it has the semantics of a function on a stem.
4. X[M2] → semantics = nil
   If a lexeme is subject to morphological rule 2 (inflection, (274)), then it has no independent semantic representation.
5. X[M*n*] → syntax = nil
   If a lexeme is subject to any morphological rule whatsoever, then it has no independent syntax.

These are defaults only, of course, and particular “lexemes” in various languages can override these defaults, but at a cost. The prediction of Autolexical Syntax is that “the more deviant a form, the rarer it is both among languages of the world and within the lexicon of a particular language” (289). This of course raises the question of the nature of such a “form” that both exists within one language and has a cross-linguistic frequency as well.

In a more general assessment, Autolexical Syntax commits one to a fairly idiosyncratic architecture of the grammar, and the interactivity of an omniscient and omnicompetent interface subsystem is somewhat worrisome as a theoretical construct (see Sadock's (1991:20) Figure 2.1 for a graphic 'black hole' metaphor). The theory is at its best when it takes on clitics and incorporation, but its take on more commonplace morphology is rather less insightful. Whereas many theories which posit components make the components themselves do most of the work and an interface (if any) tidies up around the edges; in Autolexical Syntax, by contrast, the components are essentially abstract formal filters on what takes place in the arena of the Interface, where representations are compared and lexemes inserted in structures.

Spencer (1993:151) wonders in a review quite pointedly, "do we need the machinery of Autolexical Syntax to account for all this?" Spencer's answer is "no," and it does indeed seem that even though the formalism is not in itself unnecessarily powerful, the character of the Interface, inasmuch as the three representations of any expression need not in principle have much of anything to do with each other, belies the well-considered formalism. The account of cliticization also has some empirical challenges (Spencer 1993:149-50) in the area of 2P clitics in Serbo-Croatian. A user of Autolexical Syntax must be aware that even though Sadock invokes a number of major theories (GPSG (Gazdar et al. 1985), Prosodic Morphology (McCarthy 1981), Word Syntax (Selkirk 1982), Stratificational Grammar (Lamb 1966), Montague/Categorial Grammar (Dowty 1979)), and claims to be borrowing from them at several points, the overall Autolexical picture is not readily compatible with the broader range of assumptions found in any of the sources, and so an Autolexical analysis will be somewhat "in a world of its own"—a potentially stimulating world, but isolated nevertheless.

Sadock, Jerrold. 1985. Autolexical syntax: A theory of noun incorporation and similar phenomena. *Natural Language and Linguistic Theory* 3.379-440.

Sadock, Jerrold. 1988. The Autolexical classification of lexemes. *Theoretical Morphology*, ed. by Michael Hammond and Michael Noonan, 271-90. San Diego: Academic Press.

Sadock, Jerrold. 1991. *Autolexical Syntax*. Chicago: University of Chicago Press.

6. Categorial Morphology

Morpheme-based		☑				**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar			☑			**In lexicon**
Phonological formalism					☑	**Syntactic formalism**
Incremental		☑				**Realizational**

Schmerling (1983) calls for a return to fundamentals among practicing Montague Grammarians, particularly those who were practicing the category theory (Montague 1973) without involving Montague's (1970) particular brand of linguistic metatheory, which is indeed quite different from the standard assumptions in other theories of grammatical structure. Schmerling notes that, from the perspective of Montague (1970), the theoretical framework "has distinct phonological, syntactic, and semantic *systems*, while invoking neither 'morpheme' nor 'levels'" (Schmerling 1983:222). Schmerling takes the core of a language to be a set of expressions A and an indexed set of operations. The set A contains not only the basic expressions (i.e., morphological simplexes), but also "all the expressions derived from these by repeated application of the operations; it contains nothing else" (223). Schmerling characterizes her version of Montague Grammar as a formalization of the Item-and-Process (IP) approach to word formation (although the lexicon she defines is more populous than a morpheme-based theory typically requires)(223). It is not about the position of discrete meaningful pieces (à la Item-and-Arrangement (IA)), but rather operations, separate from the words they participate in defining (224). The remaining parts of language are "an assignment of category indices to the basic expressions ... and a set of rules to assign category indices recursively to derived expressions" (223). A category-assigning rule is tripartite, containing (1) the index of the operation employed in the rule, (2) the index of the input categories, and (3) the index of the output category of the rule (223-24).

The idea of operations applying at the edges of expressions, despite Schmerling's de-emphasis on concatenation, is a common occurrence in Categorial Grammar. The pattern of functors taking arguments, and together forming a larger expression of a distinct category is the bread and butter of the theory, so to speak. Schmerling (1983) in particular talks about things that happen in response to cliticization, an example of "internal modification of an expression at its periphery" (226). Mutations and alternations, as operations, are assumed to be triggered by edge concatenation (226-27). This assumption is in trouble on empirical grounds for untriggered mutations and ablaut, e.g., English *man*/*men*. Cases like these involve affixation only under remarkably abstract assumptions, and actually support her early argument that morpheme-as-thing (IA-type) analyses are unnecessarily limited. Her approach to portmanteau forms such as French *du* and *au* involves a substitution operation of the "amalgamated" form for the sequence, *de* + *le* and *à* + *le*, respectively (228-30). Cliticization is similarly to be handled by a substitution of the clitic group for the host (226).

At a more concrete level, if we take any given operation to be the equivalent of any other, then non-concatenative morphology is no different from concatenative morphology. If, on the other hand, we consider the relative power and latitude of a substitution operation as opposed to an operation which takes an expression as an argument and *does something to that expression*, it seems that non-concatenative operations do not achieve equality in Schmerling's model.

The next major step in Categorial Morphology is Hoeksema's (1985) dissertation. Written without reference to Schmerling (1983), Hoeksema acknowledges that there is more to morphology than edge affixation, but decides to forgo those complications until after a solid theory of Categorial concatenation is in place. Hoeksema takes the more conservative approach to Montague metatheory, whereby expressions are represented as triples: a phonological projection, a categorial projection, and a semantic projection, formally (12):

For every lexical entry L: L = $<p_p(L); p_c(L); p_s(L)>$

Hoeksema (1985) is not particularly interested in phonological details, so the phonological projection, where mentioned at all, is typically just the standard orthographic form of the expression in question. Similarly, the details of the semantic projection are left fairly underspecified—where necessary, the semantic projection takes the form of expressions of intensional logic (13). The categorial projection, Hoeksema's (1985) true interest, is given significantly more detailed discussion. Basing the "word syntax," as he puts it, on the general framework of Categorial Grammar, "the categorial representations will be members of the set defined by the recursive statement" (13):

X is a category iff:
(i) X is a member of the set of primitive categories PC
(i.e., N, NP, and S); or
(ii) X is of the form V/W, where V and W are
categories; or
(iii) X is of the form V\W, where V and W are
categories.

Now the primitive category set is truly minimal, and it entails some rather complex derived categories at times, e.g., (NP\S)/NP = transitive verb, i.e., an expression such that, if it finds an NP to its right, will form an expression NP\S, which in turn, if it finds an NP to its left, will form an S (17). The information is "in there," but it takes some patient unpacking.

Hoeksema (1985:17-22) has a clear morphemic bias, since he defines one-place versus two-place operations, based on whether concatenation is involved (two-place) or not (one-place). Again, as with Schmerling, this makes concatenative and non-concatenative morphology qualitatively different. One-place operations include substitutions and zero-conversion (alias *transpositions*), whereas two-place operations include affixation and compounding (17-18). One-place operations are set aside almost entirely for the remainder of the book (subsequent chapters focus on compositionality and different types of compounding).

It will be useful at this point to summarize the approach to affixation. The two-place operations employed in the Categorial Morphology of Hoeksema (1985:19) are *right-cancellation* (RC) and *left-cancellation* (LC), common in Categorial Grammar:

RC (A/B, B) = A LC (A, A\B) = B

These operations, incorporated into lexical rule schemata of prefixation and suffixation are as follows (19):

Pref (v, w) = $<[p_p(v) + p_p(w)]$; RC $(p_c(v), p_c(w))$; $p_s(v)(p_s(w))>$
Suff (v, w) = $<[p_p(v) + p_p(w)]$; LC $(p_c(v), p_c(w))$; $p_s(w)(p_s(v))>$

Using these schemata, phonological projections are simply concatenated, categories are cancelled and resolved into new, derived categories, and semantic functors take scope over their arguments. This is fine as far as it goes, and Hoeksema (1985) has other fish to fry, so to speak, so the present focus should turn to Hoeksema and Janda (1988), where operations other than affixation take center stage.

In Hoeksema and Janda (1988), now in light of both Schmerling (1983) and Hoeksema (1985), the basic Categorial Morphology formalism is presupposed. From the very first expository section, 'Addition', context sensitivity beyond the purely categorial is assumed. Prefixation and suffixation, jointly referred to as *extrafixation*, are the only even potentially context-free operations (Hoeksema and Janda 1988:204). Addition operations which are context sensitive may be sensitive to phonological properties of their arguments (phonological constraints on the English suffix *–en* in *soften*, *tighten*), of prosodic constituents of varying sizes and qualities (e.g., consonants, vowels, clusters, syllables, stressed vowel/syllable etc.)—infixes are regularly placed with reference to one of these categories, rather than with reference to a morpheme boundary *per se*. Infixes and certain clitics are generally placed just within the edges of expressions, and a mechanism proposed by Bach (1984) called 'wrapping' is invoked to handle these cases. The first step is to distinguish the first and last elements in a string from the non-first and the non-last, respectively.

Let x be the string $x_1 \dots x_n$.
(i) FIRST (x) = x_1
(ii) RREST (x) = $x_2 \dots x_n$
(iii) LAST (x) = x_n
(iv) LREST (x) = $x_1 \dots x_{n-1}$

Once these basic operations are defined, the operations R[ight]WRAP and L[eft]WRAP can be defined in terms of them:

RWRAP (x, y) = FIRST (x) y RREST (x)
LWRAP (x, y) = LREST (x) y LAST (x)

The disposition of y with respect to the discontinuous parts of x needs to be determined, especially in the case of clitics, but also prosodically in general for issues of syllabification or metrical foot assignment, e.g., and so the further complex operations are defined (209):

(i) LWRAP-pref (x, y) = (LREST (x) (y LAST (x)))
(ii) LWRAP-suff (x, y) = ((LREST (x) y) LAST (x))
(iii)RWRAP-pref (x, y) = (FIRST (x) (y RREST (x)))
(iv)RWRAP-suff (x, y) = ((FIRST (x) y) RREST (X))

This allows the placement of a morpheme in second position (iii & iv) or in penultimate position (i & ii), with prosodic or other dependency to the left (ii & iv) or to the right (i & iii). As may be seen from the above, Hoeksema and Janda (1988) are very much about responding to empirical challenges with independently motivated formal mechanisms in an enriched version of Categorial Grammar and (especially prosodic) phonology.

As suggested by Schmerling (1983:223), the operations in Hoeksema and Janda (1988:212ff) are indexed with respect to the level of analysis at which they apply (e.g. segments, syllables, words, phrases). The potential power of this indexation may be worrisome to some, but at least the levels mentioned are independently available in any general theory of grammar. A distinction between operations and the morphological rules which employ them is useful (cf. Zwicky 1987a), especially for cases where the same or very similar operations figure in multiple rules (German Umlaut, Gaelic Initial Lenition; see Janda and Joseph (1986)). In this way also, a single rule may perform multiple operations, so as not to unnecessarily fragment operations which pattern together (cf. PFM, below).

Although there are many other details available in Hoeksema and Janda (1988), it will suffice to mention a pair of related predictions which follow automatically from the formal nature of Categorial Morphology. "Rules that combine RWRAP and suffixation and rules that combine LWRAP and prefixation do not occur" (213), and "Prefixation (suffixation) on level x is sensitive only to the properties of the leftmost (rightmost) constituent on that level" (218). Fula consonant mutation would seem to cast doubt on the latter prediction (Lieber 1992:166):

waa 'monkey'	**w**aa-ndu	Class 11
	baa-ɗi	Class 25
	mbaa-kon	Class 6

Although these are otherwise apparently well-founded generalizations, it should be noted that they are both phrased with respect to extrafixation, despite the article's explicit focus on process morphology.

Categorial Morphology has a long and respected ancestry, although it has not particularly caught on outside of the company of practicing Categorial grammarians Since it is a challenge to motivate this approach without first motivating a Montague view of linguistic metatheory, there are some inevitable obstacles to the accessibility of an analysis cast in this framework. As Hoeksema and Janda (1988) show, however, there is room under the umbrella for more than concatenation (compare Word Syntax, below), and this is clearly a(n unanticipated) bonus in empirical coverage.

Hoeksema, Jack. 1985. *Categorial Morphology*. New York: Garland. [1984. Rijksuniversiteit Groningen dissertation].

Hoeksema, Jack, and Richard D. Janda. 1988. Implications of process-morphology for Categorial Grammar. *Categorial Grammars and Natural Language Structures*, ed. by Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler, 199-247. Dordrecht: Reidel.

Schmerling, Susan F. 1983. Montague morphophonemics. *Chicago Linguistic Society* 19(P).222-37.

7. Distributed Morphology

Morpheme-based	☑					**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar	☑					**In lexicon**
Phonological formalism			☑			**Syntactic formalism**
Incremental				☑		**Realizational**

Together with Word Syntax and LMBM (both below), Distributed Morphology (DM) hopes to lay claim to the morphological interface of choice with GB–Principles and Parameters Syntax. Primarily a theory of inflection, DM adds a component to the traditional T- or Y-diagram of the grammar, placing *Morphological Structure* (MS) between S[urface] S[tructure] and P[honological] F[orm] (Halle and Marantz 1993:114). In this way, mismatches between syntactic, morphological, and phonological constituency can be accommodated before phonological implementation (115). The name, Distributed Morphology, is intended to “highlight the fact that the machinery of what traditionally has been called morphology is not concentrated in a single component of the grammar, but rather is distributed among several different components” (111-12). Word formation, they claim, can take place at any level of grammar, but they recommend only methods based on syntactic movement of heads (112). This is consistent with a post-SS component dealing with inflectional implementation and little else.

Inflection in DM is the result of lexical insertion of individual abstract morphosyntactic features in (sub-)terminal nodes under X^0. As many nodes are created under X^0 as there are inflectional categories to be realized, plus one for the lexical stem. “Morphological operations” apply to these morpheme-nodes, uniting those which are realized by a single fused exponent; morphemes with multiple exponents are ‘fissioned’ and the pieces moved to their respective positions. In DM, therefore, it is important to arrive at the right number of (sub-)terminal nodes for correct (lexical) insertion of inflectional morphemes. The question of what triggers the creation of (sub-)terminal nodes under X^0, something one might want to attribute to position or function in a syntactic construction, never arises in DM (perhaps because it is too obvious?), but the resulting metaphor is one of building structures to suit prospective residents (the inflectional properties), then remodeling to permit cohabitation (fusion) or separation (fission). With fusion, relative order is of little concern because one feature moves to be with another. Fission likewise operates without regard to ordering—it clones a node, and then separate positioning rules determine where the co-nodes end up. All the while, one knows ‘what to do’ because one knows ‘what’s about to happen’, that is, which morphemes are to take up residence in these structures. Halle and Marantz (1993:115) refer to fission and fusion as “well-motivated” operations, but this is only true on the assumption that abstract morphosyntactic nodes are atoms that must be created individually and then dealt with before it is too late, i.e., before PF. This is not at all a necessary assumption, but it is consistent with much of the Government and Binding (GB) emphasis on minimal units, extensive abstract structure, and computation.

Morphosyntactic features are represented as binary in DM, but their use is largely *ad hoc*, with features of any sort ([±strong] next to [±past] next to [±participle]) as needed. This gives the desired impression that this is a unified picture, despite the lexical-class character of the first, the morphosyntactic character of the second, and the arguably purely morphological nature of the third.

Vocabulary insertion in DM is quite late, into abstract, well-formed syntactic structures, on the condition that the features present in the (sub-)terminal nodes are non-distinct from the morphemes to be inserted (Halle and Marantz 1993:121-22). The phonological features of all morphemic material are inserted at MS and not before (122). The particular morphemes inserted may trigger 'morphologically conditioned phonological rules' called 'readjustment rules' in DM. Since there is no phonological material to act on before MS, it makes sense that such readjustments are subsequent to vocabulary insertion.

Halle and Marantz (1993:121) suggest that "the most striking difference between SS and MS derives from the systematic difference in the type of features found in the terminal nodes in the two structures." A more significant difference is that SS is a state, a structure, and MS corresponds to a derivation of indefinite complexity between SS and PF. MS is not simply the creation of (sub-)terminal nodes—it includes morphological operations, node placement, vocabulary insertion, allomorph selection, and readjustment. For Halle and Marantz (1993:114), "MS is a syntactic representation that nevertheless serves as part of the phonology." Why even a pretense of modularity, then? When it is convenient, MS is the representation after nodes are created but before fission and/or fusion, however, if one says in the same breath that vocabulary insertion happens "at MS" (122), then it clearly must be subsequent to the morphological operations, otherwise no fusional morphemes or multiple exponents could be inserted.

Halle and Marantz (1993:121) claim to subscribe to the Separation Hypothesis of Beard while giving the credit to Chomsky (1965), because there is a separation between the creation and manipulation of abstract nodes, on the one hand, and the phonological side of vocabulary insertion on the other. They "extend this separation to stems (lexemes) as well as affixes" (172, fn. 10), which shows their own peculiar definition of *lexeme*, and which furthermore is distinct from the position taken in Halle (1990), where particular lexemes were inserted at DS and inflectional affixes at MS. The thoroughly abstract position of freely generated syntactic structures, freely inserted morphosyntactic features, and freely inserted vocabulary items to be sorted out by a range of co-occurrence constraints and other filtering devices (Halle and Marantz 1993:121) makes late insertion possible, although an instinct to insert at least the major category items earlier to somehow give direction to the derivation is understandable. Halle and Marantz admit that there is "insertion" of vocabulary items at SS (122, quotes in original), but without any phonological substance. It turns out, then, that the information contained in the final construction is "there" all along, and that the requirements of PF necessitate a certain amount of "last minute" (i.e., MS) busy-work. This makes morphology seem more like a repair strategy (or set of strategies) than an integral aspect of a grammar.

Because in DM morphosyntactic features are attributes of terminal and (sub-)terminal nodes only, stem selection is sensitive to the addition of particular (potentially phonetically null) affixes. The selection of a past stem *rang*, for example, is determined by the presence of a (sub-) terminal node bearing the feature [+past] in which no overt morpheme is to be inserted. If [+past] were a feature of the head V, the appropriate stem could be selected without this appeal to inter-morpheme dependency, and the zero-morpheme could be dispensed with altogether. The assumptions of DM's MS, however, entail that [+past] have its own node, and that this is separate from the stem node. Localizing morphosyntactic features in the (sub-)terminal nodes under the X^0 nodes allows DM to avoid 'spell-out' rules found in rule-based realizational theories (Halle 1990:155), but the morpheme-based realization in DM requires an "intraword constituent structure" that is not part of the rule-based alternatives (A-Morphous Morphology, above, and PFM, below).

DM characterizes the choice among inflectional morphemes as one of competition, a common metaphor in realizational theories. As mentioned above, however, vocabulary insertion in DM is context-sensitive, only possible after the atomic (sub-)terminal nodes of MS have been resolved into the required content and position. Only at this point, therefore, can the context be identified with certainty and the correct morphemes even begin to compete for insertion. As also observed with respect to A-Morphous Morphology (above), and both LM&P and PFM (below), a principle of proper inclusion precedence, the so-called Pāṇinian Principle, is appealed to in DM as well (Halle and Marantz 1993:123). Competition is relevant, of course, only among actually insertable morphemes, i.e., those compatible with the insertion context, non-distinct from the features present in the (sub-)terminal node. In DM, the criterion for precedence is appearance "in the most complex, most highly specified context" (123; cf. the criteria of A-Morphous Morphology and PFM).

Despite the variety of manipulations available within MS, Halle and Marantz (1993:124) still find it necessary to appeal to rule blocks consisting of all morphemes realizing the same features. This move is redundant in frameworks that index rules to lexical classes (cf. PFM, Network Morphology), but it is necessary here, since DM implementation has blinders on[4] with respect to the stem node (or anywhere else the inherent attributes of the lexeme in question might be located). A sample rule block is given here to demonstrate three things: (1) how strongly motivated Halle and Marantz are to make even questionably phonetically similar effects part of a single rule (*beat-en* vs. *go-ne*), (2) how they have organized the block as a *position class*, even though the claim is one of content-oriented block organization, and (3) that DM must appeal to lexemes and morphologically defined classes thereof, despite their focus on the morpheme level (126):

[+participle, +past] ↔ /-n/ / X + ___
where X = *hew*, *go*, *beat*, ...
[+past] ↔ Ø / Y + ___
where Y = *beat*, *drive*, *bind*, *sing*, ...
[+past] ↔ /-t/ / Z + ___
where Z = *dwell*, *buy*, *send*, ...
[+past] ↔ /-d/

4 This narrowness of focus is a fleeting thing in DM, since stem allomorphy is claimed to be sensitive to the featural content of the other (sub-)terminal nodes, yet the inflectional class of the stem is not accessible to the insertion of affixes, necessitating the rule blocks.

[+participle]	↔	/-ing/
[3sg]	↔	/-z/
	↔	Ø

This is simply the entire set of verb inflecting suffixes, organized into a single block, connected by (only sometimes valid) bi-directional implications. Given DM's assumptions about the featural content of MS (sub-)terminal nodes at the point of vocabulary insertion, the /-ing/ and /-z/ affixes are not in serious competition with the others. The condition on insertion that a morpheme not be "featurally distinct" from the node into which it is to be inserted would technically allow the /-ing/ into the competition, but the simple fact that it is <u>never</u> used in the realization of a past participle in English suggests that a more careful formulation of either context or rule would eliminate such spurious competitors. Note that stem allomorphy is handled entirely separately from suffixation, in the readjustment rule division, even though the context in question ([±past, ±participle]) is responsible for triggering the choice of stem allomorph as well as the choice of suffix (see Halle and Marantz (1993:128) for sample readjustment rules).

The DM framework has very little to recommend it. Generalizations are fragmented, structure can be created and manipulated (and possibly deleted with no perceptible sign of ever having been there) by the notoriously powerful device of transformation, zeroes abound in representations, and the readjustment rules are *ad hoc* clean-up operations. While a theory must have adequate descriptive power, the conflicted internal logic of the DM assumptions makes MS a potentially very messy 'place' to be, with an unusually great need for representation-tweaking. Pullum and Zwicky (1991:387) claim that DM "represents a rejection of the proposals in *Aspects* (Chomsky 1965) and most subsequent work on the morphology-syntax interface, and a reversion to some of the earliest work in generative grammar."

Embick, David, and Rolf Noyer. 2007. Distributed Morphology and the syntax-morphology interface. *The Oxford Handbook of Linguistic Interfaces*, ed. by Gillian Ramchand and Charles Reiss, 289-324. Oxford/New York: Oxford University Press.

Halle, Morris. 1990. An approach to morphology. *Proceedings of the 20th annual meeting of the Northern-Eastern Linguistic Society*, 150-84.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. *The View from Building 20*, ed. by Kenneth Hale and Samuel Jay Keyser, 111-76. Cambridge, MA: MIT Press.

http://www.ling.upenn.edu/~rnoyer/dm/

8. Lexeme-Morpheme Base(d) Morphology[5]

Morpheme-based				☑		**Word/Lexeme-based**
Formalist			☑			**Functionalist**
In grammar			☑			**In lexicon**
Phonological formalism			☑			**Syntactic formalism**
Incremental				☑		**Realizational**

Lexeme-Morpheme Base Morphology (LMBM) is a complicated and ambitious theory of language. It can be called a theory of language because of the role its originator sees for a morphological component. “All the borders between all linguistic modules [are] defined as morphological interfaces comprising algorithms which convert the representations of one module to those of the other” (Beard 1995:389). From this description, it might seem that LMBM would be a morphologist’s paradise, since it makes the grammar apparently morpho-centric. It is an ambitious theory because its implementation requires a revision of almost every traditional component of grammar, so even though the sequence of assumptions cohere, there are many unorthodox assumptions concerning categories, morphological realization (the Morphological Spelling (MS) component), the nature and content of the lexicon, and considerably more as well. The assumptions require much exposition and justification, and so LMBM’s adoption in a particular analysis is almost guaranteed to run into confused resistance from the uninitiated majority.

In LMBM, a base component creates hierarchical structures which stand as general potential inputs to both the lexicon and the syntax. The content of such structures is some number of basic (underived) *lexemes* (defined as the major categories N, V, and Adj only). A subset of a putative universal set of 44 basic grammatical functions are assigned to nodes in the structure (Beard 1995:391-95). Derived lexemes are created in the lexicon from the base-generated structures through an *amalgamation* metaphor, whether through head to head raising or through bracket erasure. If derivation is not opted for, then every node in the base-generated structure must be accounted for (somehow filled, with a lexical or a functional head, in GB—but not LMBM—parlance) according to the general rules of GB syntax. It is crucially important to note that the output of the syntax and the lexicon is quite abstract, and the only phonological content is the underlying phonological representations of the basic lexemes in the structure. Morphological information, by which is meant anything that is realized by bound morphology or closed class free morphemes (including adpositions, pronominals, auxiliaries), is spelled-out in the MS component.

Ordering of affixes is determined based on the assumption that grammatical features in representations are ordered. Inherent features of the lexeme are spelled out first, then those of any derivational functions picked up in the lexicon, and then finally any grammatical features which were acquired by virtue of syntax (i.e., inflection). The MS component need not ‘see’ the layering of features, it is simply that the ordering is

[5] Earlier work on this theory indeed used “based” in the name, but in more recent work , e.g. Beard (1995), an increasing role for the base component in the architecture of the theory led Beard to alter the name, although many people have apparently not noticed the change, since references in the literature as often have the ‘d’ as not.

determined in up to three distinct stages, and the Affix Ordering Generalization is consistent with this layering (to the degree that the Generalization holds up empirically). Since all that the MS component gets as phonological input is the stem of the lexemes, it follows that morphological realization proceeds from the 'inside-out'. Beard (1995:54-55) casts doubt on Bybee's (1985; see "Network Model," below) *relevance* hierarchy as a universal category order, but at the same time has his own universal set of categories to propose; this can hardly be a coincidence.

There are several ways in which LMBM tries to "have it both ways," theoretically speaking. In order to account for those aspects of structure which are shared between derivational morphology and syntax, Beard strengthens the notion of the base component, which serves as the common input to both components. In order to keep the effects which motivate the Split Morphology Hypothesis without losing the generalization that many of the same sorts of marking processes are used in both inflection and derivation, LMBM posits the late-applying MS component which formally implements all of the grammatical functions and features distributed in the lexicon and the syntax (the *Integrated Spelling Hypothesis* (Beard 1995:101). In this way, derivation and inflection are functionally distinct, but formally united.

In LMBM, the notion of Case, which has been widely used in GB syntax (but with little independent motivation that did not overlap with either thematic roles or hierarchical structure), is redefined as a purely morphological notion. Given the universal set of grammatical functions, these functions are expressed by various syntactic constructions and morphological markings. Because the relation between grammatical function and morphological Case is typically not one-to-one, Case is seen as a morphological means of spelling out, in part or in whole, grammatical functions (Beard 1995: 254). These grammatical (i.e., not semantic) functions serve a crucial role in LMBM, and so it is important that a practitioner of LMBM accept the validity of the grammatical functions as a closed and universal set.

Agent	Means	Location	Prolation
Patient	Route	Goal	Proximity
Subject	Manner	Origin	Opposition
Object	Ession	Inession	Perlation
Possessivity	Duration	Adession	Circumession
Possession	Iteration	Anteriority	Termination
Measure	Accordance	Posteriority	Concession
Material	Purpose	Superession	Distribution
Partitivity	Exchange	Subession	Exception
Distinction	Cause	Transession	Privation
Absolute	Sociation	Intermediacy	Thematicity

(white = primary functions; light gray = primary spatial functions; darker gray = secondary functions)

LMBM assumes that any nominal entity in a sentence bears one (or two) of the above functions. A nominal may bear two functions if one is primary (spatial) and the other secondary, e.g., [Goal[Posterior]] *He went **behind the camera***.

An important innovation in LMBM is the disposal of several syntactic categories. This change is entailed when grammatical morphemes, both free and bound, become part of morphological spelling and consequently do not require a structural position in syntactic trees. It has long been noticed that there are functional parallels between prepositions and Case marking. Beard takes this as an indication that the functional parallels motivate a formally unified treatment. The tradition of classifying adpositions as [–N, –V] lexical items, despite their closed class status, has been misguided, according to Beard, who suggests that adpositions and Case marking co-operatively serve to identify grammatical functions of NPs, and thus that there are no syntactic PPs at all. This is a strong claim, with a *prima facie* counterexample in the Celtic so-called "inflected prepositions," but it does follow from the cross-linguistic distribution of adpositions with respect to case marking (extended argumentation in Beard 1995:229-77).

LMBM maintains a strict distinction between abstract grammatical functions and the formal pieces involved in the realization of those functions, i.e., the *Separation Hypothesis*. The separation in LMBM is more than just a logical conceit—the architecture of the grammar directly reflects this separation, since the grammatical functions are available even in the base component, but no phonological representations other than the stems of lexemes is available until the (post-syntax, post-lexicon) MS[6] component.

It must be acknowledged that LMBM takes the spirit of the GB post-syntactic level of Phonological Form (PF) very seriously. LMBM finds itself caught between two goals:

(1) to serve as a replacement to Word Syntax (see below) as a morphological interface with GB syntax, and
(2) to remain true to the several ways that LMBM architecture uncompromisingly deviates from the GB architecture.

Aspects of the latter goal include, for example, the fact that the base component would replace D-structure; the grammatical functions would more than replace GB's Case and Theta theory; and the reassignment of all function words to the MS component would fundamentally change tree-structure. These are large and sweeping revisions that would not go down smoothly in GB circles.

By translating grammatical functions into an abstract set, LMBM hopes to achieve cross-linguistic applicability in a way that theories which have a richer array of lexical categories and structural positions often do not. If the grammatical functions do indeed prove a viable approach, the focus of work in syntax and morphology would likely, almost necessarily, change extensively. There are some apparent logical problems of sequencing, such as having both a generative lexicon and a generative syntax, and the switching back and forth from one component to the other that sentence-building in LMBM would seem to require. There is also the apparent countermodular need for the base to have access to the stock of lexemes in advance of submitting the base-generated

---

[6] LMBM's *MS* ≠ DM's *MS*.

output to the lexicon. Even though the proposed meshing of LMBM with GB theory would require only “a modest adjustment” in GB, according to Beard (1995:361), LMBM has a distinct agenda as far as linguistic theory and investigation go. LMBM leaves an autonomous syntax with considerably less to work with than GB is used to having.

Although Beard (1995) does occasionally mention speakers of languages, the metaphors are more generally in terms of the automatic implementation of systems of deductive algorithms, the mapping function between components that morphology serves. Rhetoric can get a little mystical sometimes: “The lexicon has two options... If the lexicon chooses the former tack...” (339-40). Considerable thought has gone into both big picture and small picture issues in LMBM, but it seems that the revolution in orientation that LMBM's acceptance would require stands as a serious obstacle. That said, it takes an open mind (and not much of a vested interest in the pre-eminence of syntax) to fully engage this theory, but this is only because the framework contains a great many challenges to the conventional wisdom about what words and affixes are like.

Beard, Robert. 1986. Neurological evidence for Lexeme–Morpheme Based Morphology. *Acta Linguistica Academiae Scientarium Hungarica* 36.1-22.

Beard, Robert. 1987. Morpheme order in a Lexeme/Morpheme-Based Morphology. *Lingua* 72.1-44.

Beard, Robert. 1988. On the separation of derivation from morphology: Toward a Lexeme/Morpheme-Based Morphology. *Quaderni di Semantica* 9(1).3-59.

Beard, Robert. 1995. *Lexeme–Morpheme Base Morphology*. Albany, NY: SUNY Press.

http://www.facstaff.bucknell.edu/rbeard

9. Lexical Morphology and Phonology

Morpheme-based		☑				**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar					☑	**In lexicon**
Phonological formalism	☑					**Syntactic formalism**
Incremental	☑					**Realizational**

The literature on Lexical Morphology and Phonology (LM&P) is at the same time rich and convoluted. It represents a convergence between a morphological approach (level ordering) and a phonological approach (rule strata) with similar but not always identical theoretical assumptions about causes and effects in morphophonology. No piece of LM&P writing is complete without a box-and-arrows representation of modules in the grammar, because much of the concern in LM&P is getting the surface facts right with as general a rule set as possible, or as unified an underlying representation as possible, or (somewhat contradictorily) both of these at once.

LM&P holds fast to the one-meaning, one-form principle in the construction of underlying representations for morphemes. From the concatenation of abstract (morpho-)phonemic entities, there arise questions of deviation between presumed underliers and the surface pronunciation. LM&P assumes that all but the most recalcitrant alternations are effected by a rule of some sort. The recurrent question when looking at a morphologically complex expression in LM&P is “which came first?” The linear order of affixes is taken to reflect in some measure the sequence of sound-structural rule application to a base. On the (controversial) assumption that all sound-structural rules are primarily (or entirely) phonological, the issue of modularity in grammar arises again and again. Morphological processes add material, and then phonological process ‘iron out’ the discrepancies between what biuniqueness would predict and what actually occurs.

The method outlined above would be relatively simple if there were never any interaction between alternations associated with one affix and those with another, or if the changes observed were clearly phonetically motivated. To the degree that rules from different “levels” are interleaved and phonetically arbitrary (synchronically at least), LM&P has had plenty of grist for the theoretical mill.

Cyclicity of rule application has been a longstanding issue in LM&P because on the one hand, morphologically complex expressions are assumed to be built from the inside out in layers, represented by labeled bracketing, but on the other hand, there are numerous rules which would seem to apply in conjunction with the addition of a number of distinct affixes, rather than being affix-specific. LM&P assumes that some rules must apply cyclically, because certain rules apply multiple times, but cannot be handled by purely phonological conditioning. Cyclic rules apply in the lexicon, as words are built, and non-cyclic rules apply across the board postlexically. The two rule types have certain general attributes, e.g., lexical rules apply only in derived environments and are subject to lexical exceptions, whereas postlexical rules are exceptionless, general rules. Within the (universal construct of a) lexical rule component, rules are assigned to distinct language-specific strata, according to their behavior. Ideally, of course, the number of strata should be minimal, since the assignment of already idiosyncratic rules to such strata involves extensive stipulation. Most descriptions of languages use two rule strata, with the notorious exception of Halle and Mohanan (1985) who invoke four strata to handle English, alternating levels of rules and readjustments (cf. DM, above). An alternative approach to this from a more morphological perspective is the assignment of *affixes* to strata, and then having the sound-structural rules be concomitants of morphological rules. This is more in line with the work of Siegel (1979) and Allen (1978), the approach called *level-ordering*. The prime investigators in LM&P, however, are phonologists, especially Kiparsky (1982a and b, 1985), and this influential group, along with the formalism carried directly over from generative phonology, kept the ‘meaning’ side of morphology to a minimum. Kiparsky (1982a:39) warns with (trisyllabic) gravity that an appeal to “morphologization” (quotes original) is “the most unfortunate treatment of all,” that it constitutes a claim “that there are as many ‘Trisyllabic Shortening’ rules as there are suffixes that can trigger the process.” This last statement clearly establishes LM&P as morpheme-based and incremental. It is an Item and Process theory of morphology (if not Item, Arrangement, and Process).

Booij and Rubach (1985) suggest that there is a further lexical component, a post-cyclic lexical rule block that, as the name applies, follows the application of all cyclic lexical rules, yet still participates in determining the shape of particular words, and therefore is distinct from the postlexical rule block as well. This move leads them into a position where they must posit functionally parallel rules in different components, a problem which they dismiss on the grounds that the repetition is not 100% (15-18). Making redundant formulation an all-or-nothing issue, however, is an innovation with Booij and Rubach, since generalizations can be lost in sometimes very subtle ways.

Booij and Rubach's investigation of clitics in Dutch and Polish leads them to the claim that not only are clitics in the lexicon, they are affixed to bases in the lexicon, because they correlate with sometimes quirky alternations in the shape of the host (35ff.). Given what is known about the promiscuity of clitics vis-à-vis the distribution of affixes, this means that in the lexicon is an entry for the combination of every clitic and every potential host element in the language, a massive expansion of the lexicon. This claim is maintained despite their positing a separate operation of cliticization in the syntax, for all and only those clitics that do not correlate with alternations in their hosts (i.e., those that are phonologically uninteresting)(50). Rather than unify cliticization in the phonology, Booij and Rubach claim that there are lexicalized and non-lexicalized clitic-host combinations. This is equivalent to saying that only the parts of words which show alternations are ever "in the phonology"—that phonology only exists when it is actively altering something.

The justification for positing strata and for the assignment of particular rules to particular strata is grounded in surface sound-structural effects. Despite bracketing conventions, there is a strong tendency for words and morphemes to fade into the background. They represent the raw material for the operation of the rules, but they have little other reason for being in LM&P. Indeed, Kiparsky claims that "the output of every cycle is a lexical item" (Kiparsky 1982a:23). At the same time, and with no apparent irony, Kiparsky (1982a:46) suggests that every lexical entry itself constitutes an identity rule, which, because of its specificity, blocks alternative realizations of the same meaning, thanks to the Elsewhere Condition (the most narrowly specified of competing applicable rules precedes—and precludes—the application of all other competitors)(Pāṇini, Anderson 1969; Kiparsky 1973). A subtle distinction here (one which is probably too subtle for its own good) is that between the (monomorphemic) lexical *entry* and the (possibly derived) lexical *item*. In Kiparsky's minimally redundant lexicon (1982a:25-26), and with morphemes being sometimes rule, sometimes thing, it is easy to lose sight of what is 'in' and 'out' of the LM&P lexicon.

As alluded to at the beginning of this section, not only have the lines and arrows been drawn and redrawn in LM&P, the sense of what exactly the lines divided and the arrows related has changed. An acknowledged forerunner of LM&P, Chomsky and Halle (1968, i.e., *SPE*) appealed to different sorts of boundary markers in the phonological representations, on a par with phonemic segments, which phonological rules could refer to at no cost. The above-mentioned approaches of Siegel (1979) and Allen (1978) kept the boundary markers, but made them something that classes of morphemes were sensitive to, determining legal attachment sites, and thus creating level-ordering. LM&P replaces distinct boundary markers with distinct types of rule application (lexical and postlexical), and posits distinct components in the grammar (which includes the lexicon) to oversee the proper application of the rules. In this way rules are limited to strata within components, and the insertion of particular morphemes serves to trigger the application of certain lexical rules. In general, it can still be said that *–ity* in English is a "stratum 1 affix," but this is only determinable on the basis of the stratum 1 phonological rules which its insertion triggers, e.g., Trisyllabic Shortening, Obstruent Voicing: *brief* ~ *brevity*.

In the resulting picture, with the burden (apparently) shifted out of morphology and onto two species of phonology, there is very little insight into morphology beyond its effect on sound structure, i.e., morphophonology. The prediction that stratum X affixes will appear outside stratum X-1 affixes is no explanation for the affixes' presence on their particular stratum— occasionally there is a separate correlate for affixes which seem to pattern together (Latinate affixes in English), but the primary and overriding factor for generalizations about morphology is the behavior they exhibit with respect to units of sound. As a theory of morphology, LM&P is oblique at best, because the whole enterprise serves to enlarge phonology at the expense of morphology. Underlying representations are abstract, despite Kiparsky's (1982a) recurrent references to constraining abstraction, and exceptions to general rules are worked out via manipulations of the underlier, rather than questioning the rule formulation. While it is impossible to deny that there are many morphophonological subregularities in the lexicon of most any language, the claim that morphophonological patterns is a fundamental organizing principle of the lexicon ignores the many more accessible patterning principles (inflectional, derivational, semantic, syntactic) that are logically prior to phonology, even a phonology with an embassy in the lexicon.

Archangeli, Diana. 1987. An overview of the theory of Lexical Phonology and Morphology. *MIT Working Papers in Linguistics* 7.1-14.

Booij, Geert, and Jerzy Rubach. 1985. Postcyclic versus postlexical rules in Lexical Phonology. *Vrije Universiteit* [Amsterdam] *Working Papers in Linguistics* 14.

Kiparsky, Paul. 1982a. Lexical Morphology and Phonology. *Linguistics in the Morning Calm*, ed. by Linguistic Society of Korea, 3-90. Seoul: Hanshin.

Kiparsky, Paul. 1982b. From Cyclic to Lexical Phonology. *The Structure of Phonological Representations* [part 1], ed. by Harry van der Hulst and Norval Smith, 131-75. Dordrecht: Foris.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2.85-138.

Mohanan, K. P. 1986. *The Theory of Lexical Phonology*. Dordrecht: Reidel.

10. Natural Morphology

Morpheme-based		☑				**Word/Lexeme-based**
Formalist					☑	**Functionalist**
In grammar		☑				**In lexicon**
Phonological formalism		☑				**Syntactic formalism**
Incremental		☑				**Realizational**

Motivated in large part by the school of Natural Phonology (Donegan and Stampe 1977, 1979), Natural Morphology (NM) is a functionally-oriented call for more precise distinctions among sound-structural rule types and components of grammar. This may seem somewhat odd, however, when considered in light of the gradient model of grammar which NM ultimately proposes. Just as Natural Phonology had distinguished the automatic from the non-automatic in phonology, so too does NM seek to distinguish rules of morpho(pho)nology from both automatic phonology and morphology proper. Dressler (1985b:3-4) holds the view that there is an interface between morphology and phonology, namely *morphonology*, which is not in itself a component, yet does not belong to either of its neighbors. He distinguishes *morphonological rules* (MPRs) from *allomorphic rules* (AMRs) on a rather vaguely defined criterion of *productivity*. Dressler clearly distinguishes the segments involved in morphonological alternations from their domain of application, and attempts to separate rule types according to phonological, morphological, "lexical," stylistic, and other conditions on application. His diagnostics result in dense taxonomy of rule types, and for that reason if for no other, Dressler's (1985b) scheme has the feel of a flock of pigeons clamoring for their pigeonholes. A sound structural rule in this framework can have many sorts of conditions beyond the phonological, and this poses no formal or logical problem, because the wide range of linguistic and extra-linguistic constraints are "in the model." There is an attempt to counter this expressive power, however, in a rather arcane and arbitrarily demarcated system of "demerits" (scores of 1-5) that are assigned to a rule as a mark of its relative "naturalness," according to generality of application, phonetic distance between alternants, and so on.

As an introduction to the theory of NM, however, Dressler's (1985b:260ff) chapter 10, "Towards an explanatory model of morphonology: On the interaction of Natural Phonology and Natural Morphology within a semiotic framework," constitutes a belated but helpful sketch. Dismissing the Chomskyan goal of describing grammatical competence as reductionist, as incapable of accounting for "facts of language change, acquisition, impairment, variation, etc." (261), Dressler turns to the business of establishing a "counter-model" (NP/NM) to the formalist paradigm, rather than quixotically hurling isolated counterexamples at it with no real hope of falsification.

NM is avowedly functionalist in its orientation, considering not only the description of language data, but also the purpose of each element in the context of its use. Dressler wisely puts forward some of the logical pitfalls of functionalist argumentation[7]:

[7] So as to keep the focus on morphology and off of metatheory in general, the reader is referred to Dressler (1985b:270-71, §10.3.13) for counterarguments.

(1) circularity (markedness, naturalness),
(2) *ad hoc* devices (unboundedly many goals to be served simultaneously by language use in a finite context), and
(3) teleology in variation and change (quasi-mystical 'group-minds' or 'community grammars').

Two driving assumptions shape the entire NM system of morphological description, prediction, and explanation:

(1) The goal of language use is effective communication, and
(2) Language is a semiotic system in the sense of Peirce (1965).

Words are primary signs in NM, morphemes are secondary signs ("signs on signs"), and phonemes are tertiary sign. With signs as an organizing principle in the service of communication, the clear demarcation of the constituent signs in a string best facilitates the interpretation of the signs and the recovery of the primary meaning. It is predicted, therefore, that the more sign-like a morpheme is, the more efficient it is as a means of communication. Segmentability being systematically favored, any process or rule which serves to obscure morpheme boundaries (e.g., much of morphonology) is predicted to be contrary to the goal of clear and efficient communication, and thus there will be pressure from within the system to 'iron out' the alternation and thus to converge on a constant form-meaning correspondence, i.e., like a good sign should (e.g., 300-06). Where such convergence does not obtain, the explanation is presumably to be found in a conflicting function which inhibits the (re)unification.

Mayerthaler (1988) leans strongly toward universal functions which all languages must address, e.g., the symbolizing or encoding of semantic concepts. In order to accomplish a meaningful characterization of universal naturalness, he draws most of his supporting data from language change and language typology. In §1.3 (pp. 8-15), Mayerthaler draws a number of broad distinctions concerning the relative markedness of related pairs of semantic concepts regularly expressed in language (e.g., definiteness/indefiniteness, animate/inanimate, present/preterit, etc.) and determines that "prototypical speaker attributes" (including the "here and now," 1st person pronominals, etc.) are universally less marked, and thus "the more important and constructive its role is for the organization of natural languages" (15). For this to have any empirical content, there must be some translation, some correspondence of semantic naturalness (unmarkedness) in the form of language. This is what Mayerthaler calls *Optimal Symbolizing*. If morphology is sign-based, the more semantically marked a feature to be symbolized is, the more 'featured' (essentially, longer) the symbolization will be ("What is semantically 'more', should also be constructionally 'more'" (19)). Since the text of the English translation is of notoriously poor quality, I will extract the useful, though terminologically confounded, scheme of symbolizing types (18):

A. Featureless (no overt marking)
B. Featured (some overt marking)
   1. Additive Featured (increased content)
      a. Particle Additive (affixed)
      b. Modulator Additive (segmental lengthening)

2. Modulator Featured (segmental substitution)

As signs go, then, the optimal symbolization is B.1.a., or affixation, and the longer the better for the symbolization of a semantically marked element (cf. Latin positive *longus*—comparative *longior*—superlative *longissimus*). Zero-conversions and non-concatenative morphology are of course predicted to be inefficient symbols, so their use, especially in a systematic way, is a puzzle for NM (subtractive morphology is not addressed). Mayerthaler is quick to point out that the predictions of NM are always relative tendencies, rather than categorical statements. This correspondence between markedness and symbolization is a type of (weak) homomorphism, and Mayerthaler terms it *constructional iconism* (17-25). Homomorphism of this sort is a desirable condition from the perspective of NM, but it is admittedly an idealization which must often be disrupted in the service of competing linguistic (and extralinguistic) functions.

Dressler (1985b:301) modifies the simplistic 'bigger is better' sign evaluation metric of Mayerthaler's with an appeal to the practicalities of perception and production—an efficient sign "must be neither too big nor too small." Along with iconicity in the form of a sign, Dressler also stresses the value of a biunique relationship between the signifier and the signified, so that either is readily accessed from the other (301). The combined pressure of iconicity and biuniqueness motivate the prediction that operations which fuse or delete morphemes, whether in part or in total, are diachronically unstable and synchronically rare (306). Ambiguity in the input or output of any rule is a strike against it as a natural rule, and thus homophonous morphemes are to be disfavored ("homophonous zeroes" all the more so!)(313). Dressler lays out the following seven point scale, ranked in descending order of *morphotactic transparency* (=naturalness) (316-17):

I. Intrinsic allophonic phonological rules (PRs) intervene,
II. Extrinsic allophonic PRs, resyllabification,
III. Neutralizing PRs intervene
IV. Morphonological rules (MPRs),
V. MPRs with fusion,
VI. Allomorphic rules (AMRs), and
VII. Suppletion.

Dressler notes that rules often change type over time (cf. Janda 1986), and contrary to all expectations of naturalness, the change tends to be in the direction of decreasing transparency.

Biuniqueness and productivity are thought to go hand in hand, with the former implying the latter (Dressler 1985b:329). In this way, the addition of new words to the lexical stock, which should employ the most productive means available, ought to involve the application of the clearest (i.e., unique, and perhaps transparent, too) signifier for the signified in question. This reasoning involves a vicious circle of course, but through an appeal to the diachronic loss of transparency in rules, NM can allow for, if not actually account for the development of polysemous morphemes and the rise of new productive morphemes displacing the old.

Almost coextensive with, but inversely related to, the scale of morphotactic transparency (excluding suppletion) is the scale of *indexicality*—as a rule becomes more context sensitive, the presence of the output 'points' more clearly to the presence of its conditioning environment, and the greater the phonetic distance between input and output, i.e., the greater the change the rule effects, the more indexical the rule is (thus intrinsic allophonic rules have almost no indexical value). Although NM assumes fuzzy transitions from one rule type to another, it is nevertheless a modular theory, such that the application of PRs presupposes the application of MRs. A subtle consequence of this modularity is the quantum leap in the indexicality of a rule once it becomes morphologized. MRs 'precede' PRs, and thus they have a certain priority over PRs. MRs furthermore have semanticity, which phonemes and allophones (in themselves) do not. For these reasons, it is suddenly much less troubling that morphotactic opacity increases over time, since indexicality and semanticity increase correspondingly (Dressler 1985b:309-11, 333-34).

Wurzel (1989) turns the focus specifically on inflectional morphology, from the perspective of systems as coherent and consistent wholes. Not to dismiss the role of language typology, but rather to take individual languages as extensions of types, Wurzel refers to *System Defining Structural Properties* (SDSPs), which organize and lend stability to inflectional systems. Since inflectional classes are based on paradigms, and paradigms are based on inflectional markers, and markers in a given language "are not part of any universal inventory of markers" (63), introducing inflectional classes into a discussion of morphological naturalness is inevitably challenging. Wurzel speaks of morphological norms at the language-specific level, rather than in terms of naturalness in general, e.g., in Modern German, because the weak verb formation is increasingly common and the only productive rule for new verbs in the language, the weak pattern is (currently) the norm for German (64-65). He couches the range of SDSPs as parameters (75):

a. an inventory of categorial complexes and categories assigned to them,
b. the occurrence of basic-form inflection or stem inflection,
c. the separate as opposed to combined symbolization of categories of different categorial complexes,
d. the number and distribution of formally distinct inflectional forms in a paradigm,
e. the types of markers occurring, and their relations to the categorial complexes concerned, and
f. the existence or nonexistence of inflectional classes.

Given these SDSPs, one may construct a "typological characterization and classification of inflectional systems" (75). In very congruent systems, the SDSPs act almost as laws, while in more mixed systems, the SDSPs stand more as defaults, i.e., as what happens when no extraordinary circumstances come into play (82; cf. Zwicky 1986). System congruence (=*congruity* in Wurzel) is not something that can necessarily be assessed through cursory inspection. Rather, it involves extensive and exhaustive comparisons, e.g. (83):

1. of (abstract) marker types (e.g., suffixes),

2. of particular markers,
3. of the number of distinct inflectional forms in different paradigms,
4. of co-occurrence of various markers, and
5. of all the different markers realizing particular inflectional categories.

Given the SDSPs and the mass of empirical data that one would collect in discerning a particular language's set of parameter settings, it is not hard to imagine that a system-internal pressure toward increased congruence is proposed as a motivator of morphological change. It is also not surprising, given the focus in NM on conflicting functional motivations, that these SDSPs come into conflict with system-independent considerations, i.e., the more universally oriented issues identified by Mayerthaler (1988).

In summary, NM predicts that the most efficient morphological system will exhibit iconicity and biuniqueness to the highest degree possible, avoiding syncretism and avoiding zero-marking on all but the most basic (semantically least marked) expressions. To the degree that languages do countenance syncretism and zero-derivation, this is claimed to be the result of conflict with other systematic pressures, and further that such language states are rare, unstable, and subject to change at the earliest opportunity. Although the testing ground for these intuitively plausible hypotheses is based on the description of synchronic morphological systems, the methodological focus is always on comparison with some other language state, to evaluate the relative naturalness of the states. Indeed, as a theory of synchronic morphology, there seems to be something missing in NM. Mayerthaler himself states (1988:4) "[W]e do not believe in the possibility of a synchronic linguistics in the sense that it would be possible to write an adequate grammar excluding the dimension of time." It is perhaps not an accident that a theory that holds multiple gradient scales as central organizing principles focuses on language variation and change. As a tool of the typologist, the historical linguist, or even the dialectologist, NM would surely have an appeal in its functionalist and system-wide ("macro") orientation.

Dressler, Wolfgang U. 1985a. Typological aspects of natural morphology. *Acta Linguistica Academiae Scientarium Hungarica* 35.51-70.

Dressler, Wolfgang U. 1985b. *Morphonology: The Dynamics of Derivation* [Linguistica Extranea, Studia 12]. Ann Arbor, MI: Karoma.

Dressler, Wolfgang U. 1985c. On the predictiveness of Natural Morphology. *Journal of Linguistics* 21.321-37.

Dressler, Wolfgang U. (ed.), Willi Mayerthaler, Oswald Panagl, and Wolfgang U. Wurzel. 1987. *Leitmotifs in Natural Morphology* [Studies in Language Companion Series, vol. 10]. Amsterdam/Philadelphia: John Benjamins.

Mayerthaler, Willi. 1981/1988. *Morphologische Natürlichkeit/ Morphological Naturalness* [Linguistica Extranea, Studia 17]. Ann Arbor, MI: Karoma.

Wheeler, Max W. 1993. On the hierarchy of naturalness principles in inflectional morphology. *Journal of Linguistics* 29(1).95-111.

Wurzel, Wolfgang Ullrich. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer.

11. "Network Model"

Morpheme-based			☑			**Word/Lexeme-based**
Formalist					☑	**Functionalist**
In grammar					☑	**In lexicon**
Phonological formalism			☑			**Syntactic formalism**
Incremental		☑				**Realizational**

Joan Bybee, who was a driving force in the Natural Generative Phonology (Hooper 1976, 1979) movement, has proposed the "Network Model", a functionally-oriented view of morphology, seen first and foremost as an organized system. In the preface to her (1985) book, Bybee observes that it might appear strange to some that her attention had moved from a very concrete approach to phonology to settle on "a different set of issues" (v). On the contrary, an approach to morphology follows naturally from the careful division of morphophonemics from articulatory- and perceptual-based phonology ('phonology proper'). By emphasizing that morphophonology is morphologically conditioned, and therefore part of the domain of morphology, Bybee likens the arbitrary nature of much of morphophonology to the arbitrariness found throughout morphemics, *l'arbitraire du signe*.

As a functionalist theory, the concern is not with descriptive segmentation of morphemes, because there are simply too many deviations from a one-to-one form-meaning correspondence. Bybee's goal is "to propose certain principles in a theory of morphology whose goal is to explain the recurrent properties of morphological systems, including fusion and allomorphy, which are traditionally viewed as problems [many-to-one and one-to-many meaning–form connections, respectively], in terms of the general cognitive and psychological characteristics of human language users" (3). With cognition as a concern, psycholinguistic experimentation is an important source of evidence for the claims of the Network Model. Similarly, because the goal is to explain recurrent patterns, cross-linguistic data from linguistic typology is also of importance. Many formalist theories, by contrast, tend to de-emphasize evidence of these sorts, because they introduce gradient patterning, rather than neat categorical behavior.

In contrast with the Word Syntax approach (below), the Network Model suggests that the lexicon is not merely structured, it is richly structured, with connections at many levels. Phonological connections, syntactic (categorial and subcategorial) connections, and semantic connections link words and parts of words simultaneously. Multiple links constitute lexical associations of differing strength and character, and generalizations about lexical subclasses can refer to constellations of links, including links from different grammatical domains (cf. Jackendoff 1975). For example, much of the exposition of the network model in Bybee and Moder (1983) is done in terms of the ablauting strong verbs in English, such as *sing*/*sung* and *string*/*strung*. Bybee and Moder show that the oldest members of this class were monosyllables ending in a velar nasal. Later additions to the class have diversified this condition, allowing for a final velar and/or nasal, as in *dig*/*dug* and *spin*/*spun*, respectively. Rather than deriving morphologically complex words by rules *per se*, Bybee's model appeals to patterns among the various links in the lexicon to identify morphological patterns. Thus, morphological analysis is radically not about Items and Arrangements or Items and Processes, nor is it about Words and Paradigms. Morphological analyses are implicit in the lexical connections that individual speakers make in their own lexicons. Patterns defined by links can be referred to as *schemata*, either *source-oriented* or *product-oriented*, as conditions guiding the coining and interpretation of novel forms (Bybee 1985:129; Bybee and Moder 1983:255). The individualization of morphological analysis is not a surrender to chaos and unpredictability, however, since the empirical experience of speakers acquiring and processing their language, especially within the same community, is very likely to be comparable. With comparable experience, the reasoning goes, will come motivation for largely coinciding lexical structure. In this way, quite contrary to the 'ideal speaker-hearer' approach often appealed to in (Chomskyan) linguistic theorizing, the Network Model is based in the experience and general cognitive processes of natural language users.

Bybee (1985) claims that derivational and inflectional morphology are not qualitatively distinct phenomena, but rather "a gradual...distinction, the basis of which is *relevance*..." (5). Not only, then, is morphology restricted to the lexicon, but also form and function are distinguished quite clearly, although in practical terms, each dimension on its own is gradient. "The semantic relevance of an affix to a stem is the extent to which the meaning of the affix directly affects the meaning of the stem" (4). This is potentially a vague and variable gradient, but Bybee purports to avoid "ethnocentrism" by drawing claims about relative relevance from a cross-linguistic comparison of fifty languages in widely different language families and geographic regions, thereby escaping (to the greatest degree possible) genetic or contact confounds (8). From this typological evidence, Bybee claims support not only for the categories which she posits along the continuum, but also for the relative ranking of each grammatical meaning.. Specifically, with reference to verb morphology, "the categories of valence, voice, aspect, tense, mood and agreement are ranked for relevance to verbs in that order" (4-5). Part II of Bybee (1985:137-205) discusses in depth what is understood by "aspect," "tense," and "mood" in her model, in order to clarify the categories for further testing and to pre-empt spurious counterexamples which might follow from differing definitions of what constitutes a tense, for example.

Bybee's assumption of a cline of relevance allows her to make predictions about exponent form, on the one hand, and sequencing on the other. Bybee (1985:4-5) claims that exponents of more relevant grammatical meanings will be found closer to the verb stem than will those of less relevant meanings, and more relevant exponents are more likely to involve morphophonological alternations in the affix, the stem, or both (Bybee 1985, 33-43).

The Network Model has a variety of independently proposed solutions to problematic issues in deviations from one-to-one form-meaning matching (Carstairs 1987). Morphophonology is considered a historical relic of earlier phonetically-motivated alternation now housed in the morphology, rather than something to be processually recapitulated in putatively synchronic phonological rules. Fusional morphology is similar, and is claimed to follow from frequent cooccurrence of morpheme pairs. Affix genesis is rooted in semantic bleaching and phonetic erosion (without calling this by the name "grammatic(al)ization"). These explanations, in their broadest senses at least, are generally agreed upon in historical circles.

Perhaps surprisingly, Bybee (1985:50-58) finds a place for the basic-derived distinction in the lexicon, the natural occasion to appeal to a scale of relative (un)markedness. Unmarked (i.e., zero-marked) word forms are predicted to be semantically unmarked, or at least no less semantically marked than the most unmarked word form in the paradigm. Markedness is of course a concept which is frequently criticized for cross-investigator inconsistencies, but the typological and psycholinguistic bases for Bybee's analysis, including for example, the sequence of acquisition of forms, do manage to add some weight to her argumentation.

Claims in this area are not without difficulty, however. Bybee claims considerable support for the claim that semantically unmarked forms are "morpho-phonemically simpler" than more marked forms (6). Certain stem allomorphy facts from Sanskrit would seem to deviate from this prediction, in that for those paradigms with weak and strong stems, and especially for those with three grades of stem, the weak(est) stems (in white) are found in oblique cases to the exclusion or near-exclusion of the direct cases, where the strong stem (in gray) generally predominates, e.g., in the masculine forms of the possessive adjective *bhagavant*, 'fortunate' (Stump, 2001: 170):

	Singular	Dual	Plural
Nominative	*bhagava:n*	*bhagavant-a:u*	*bhagavant-as*
Accusative	*bhagavant-am*	*bhagavant-a:u*	*bhagavat-as*
Instrumental	*bhagavat-a:*	*bhagavad-bhya:m*	*bhagavad-bhis*
Dative	*bhagavat-a:*	*bhagavad-bhya:m*	*bhagavad-bhyas*
Ablative	*bhagavat-e*	*bhagavad-bhya:m*	*bhagavad-bhyas*
Genitive	*bhagavat-as*	*bhagavat-os*	*bhagavat-a:m*
Locative	*bhagavat-i*	*bhagavat-os*	*bhagavat-su*

Even though it would appear that the nominative singular is the only "zero-marked" form in the paradigm, the weak stem, which is always "morphophonemically simpler" is used in the oblique forms, and never in the nominative singular. This is primarily a suggestion for a redefinition of *basic* versus *derived*, however, since on grounds of predictability, the weak stem is usually predictable from the strong stem, but the reverse is less reliable, meaning that "morpho-phonemically simpler" can be a misleading diagnostic for the directionality of derivation. It would seem, therefore, that in the Network Model, the interlexical connections might more reliably point to a basic form than a guideline framed in terms of the relative number of phonemes.

A complex perspective has emerged from Bybee's particular program. This involves the difference between regular and irregular, productive and nonproductive, and type versus token frequencies of words. In Bybee and Moder (1983:251), irregular inflectional forms, particularly those involving morphophonological alternations, are claimed to be "scheduled for leveling, since they disrupt the one-to-one correspondence between sound and meaning." In Bybee (1995), however, an explanation for the endurance of certain disruptive alternations is explained with reference to token frequency (*lexical strength*, in Bybee's terminology): "irregulars will tend to regularise unless they are sufficiently available in the input to create a strong lexical representation. Thus if the irregular past has low token frequency and is thus more difficult to access, a regular form might be created" (428). The more frequently a verb is used, the more able it is to sustain irregularity in its paradigm, should any such irregularity exist. The verb *to be* is cross-linguistically very likely to show some irregularity in its paradigm, and Bybee's claim is that the reason is the frequency with which forms of the verb *to be* are used in everyday speech. The pressure of conventional usage 'outweighs' the pressure of regularity.

The sometimes elusive notion of *productivity* is also a function of frequency in the Network Model, but in this case it is type frequency, the proportion of the vocabulary in the relevant grammatical category which participates in a particular pattern. The higher the type frequency, the more likely the class is to act as a default, and consequently the more likely it is to be employed for analogy, as in cases of doubt, neologism, and language acquisition. The chances for the expansion of the pattern's input set increase as a result, and this means an increase in the pattern's productivity. Formalist theories, on the other hand, tend not to worry as much about pattern frequencies overall, with the exception of theories such as Network Morphology and PFM, in which default patterns play a central role in rule application.

The many parallels between the Network Model and the concerns of Natural Morphology (for which see above), just as there are parallels in the perspectives of Natural Generative Phonology and Natural Phonology, their respective inspirations include the types of evidence that each allows, the larger systemic questions which each seeks to address, and the focus on the isomorphic sign as a driving influence in change. For these reasons, it seems a little strange that two schools cite each other's work almost not at all (with the exception of Dressler 1985b, who offers three Bybee references). Bybee offers "Network Model" as a tentative theory-name, but not until Bybee (1995:428). Since similarities with the Network Morphology program (named in 1992/93; see next section) are very limited, perhaps the Network Model may need another title. As for the appeal of the Network Model as it stands, however, its attention to cognition and typology make it a likely stimulus for new research programs in psycholinguistics, and its clear and falsifiable predictions make it a standing challenge to those engaged in the description of synchronic systems and diachronic changes.

Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*, [Typological Studies in Language, vol. 9]. Amsterdam/Philadelphia: John Benjamins.

Bybee, Joan L. 1988. Morphology as lexical organization. *Theoretical Morphology*, ed. by Michael Hammond and Michael Noonan, 119-41. San Diego: Academic Press.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5).425-55.

Bybee, Joan L., and Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language* 59(2).251-70.

12. Network Morphology

Morpheme-based				☑		**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar					☑	**In lexicon**
Phonological formalism				☑		**Syntactic formalism**
Incremental					☑	**Realizational**

Network Morphology has been developed by the (University of) Surrey Morphology Group. An integral part of the theory is the computer language DATR (Evans and Gazdar 1996), which was designed with lexicon modeling in mind. The lexical knowledge modeled in Network Morphology is based on the common computational principles of hierarchy and inheritance. Network Morphology lexica are strongly hierarchical, and individual lexical entries are typed feature matrices, analogous to representations in Head-Driven Phrase Structure Grammar (HPSG; Pollard and Sag 1994) and Generalized Phrase Structure Grammar (GPSG; Gazdar et al. 1985), adjusted for direct computational implementation as lines of programming code. The authors of articles in Network Morphology move frequently between feature notation and tree diagrams, which is helpful to a reader who may not always be able to picture the dependency relations in the compressed featural format.

Based on the concept of feature inheritance, the Network Morphology lexicon begins at the very top with the type *word*, which branches into subtypes according to syntactic categories. New subtypes are motivated each time there is a subset of lexemes which differs from the default feature set in some systematic way. A subtype must have some specific feature value which differs from the larger class; this feature value overrides the feature value the subtype would inherit by default from the supertype. In this way, dependent types largely cohere with their parent types, and sister types cohere in the defaults they jointly inherit from a common parent node. Lexical classes and subclasses are thus defined, and this allows generalizations to refer to individual nodes or hierarchically related nodes. Simultaneously, this suggests that generalizations will not hold over disparate classes, i.e., those not so related in the hierarchy (this suggestion is not exactly true, but there is a systematic way proposed to handle it, discussed below).

Corbett and Fraser (1993:126, 136) provide a more concrete example. The declension classes in Russian are generally claimed to number three or four. An example paradigm for four typical nouns will show the reason for the ambivalence (Corbett and Fraser 1993:115).

	I *zakon* ‘law’	II *komnata* ‘room’	III *kost’* ‘bone’	IV *v’ino* ‘wine’

nom sg	*zakon*	*komnat**a***	*kost'*	*v'in**o***
acc sg	*zakon*	*komnat**u***	*kost'*	*v'in**o***
gen sg	*zakon**a***	*komnat**i***	*kost'**i***	*v'in**a***
dat sg	*zakon**u***	*komnat**e***	*kost'**i***	*v'in**u***
inst sg	*zakon**om***	*komnat**oj***	*kost'**ju***	*v'in**om***
loc sg	*zakon**e***	*komnat**e***	*kost'**i***	*v'in**e***
nom pl	*zakon**i***	*komnat**i***	*kost'**i***	*v'in**a***
acc pl	*zakon**i***	*komnat**i***	*kost'**i***	*v'in**a***
gen pl	*zakon**ov***	*komnat*	*kost'**ej***	*v'in*
dat pl	*zakon**am***	*komnat**am***	*kost'**am***	*v'in**am***
inst pl	*zakon**am'i***	*komnat**am'i***	*kost'**am'i***	*v'in**am'i***
loc pl	*zakon**ax***	*komnat**ax***	*kost'**ax***	*v'in**ax***

It has been noted that declensions I and IV are formally quite similar, contrasting clearly with both II and III (Corbett 1982). Network Morphology allows for the capturing of gradient similarities with a hierarchical approach to the lexicon. The following hierarchy (adapted from a tree diagram in Corbett and Fraser 1993:126) shows the formal affiliation of Russian declension classes:

I. Nominal
- A. Adjective
- B. Noun
  1. N_O (traditional o-stems)
     - a. N_I, e.g., *zakon*
     - b. N_IV, e.g., *v'ino*
  2. N_II, e.g., *komnata*
  3. N_III, e.g., *kost'*

This hierarchy captures "the fact that there are four main declension classes [in Russian], but that the differences between N_I and N_IV are not as great as those between either of them and the other declensional classes" (127). N_O is a "super-node" from which N_I and N_IV inherit their shared properties (127).

Since Network Morphology revolves around type hierarchies, it is important to note that each *node* in the network corresponds to a class of lexemes, characterized by common attribute-value pairs, called *facts*. Facts are inheritable downward in the network, unless overridden by specific facts listed at an intervening node in the *path*, down to and including the node in question. For this reason, facts about inflectional classes are composites of inherited facts and stipulated sub-class-specific facts. In order for a declension class to be 'well-typed', the composite of facts must constitute a complete set of rules of inference (i.e., facts) for a full inflectional paradigm appropriate to the lexeme-class[8].

[8] Parallels to HPSG are many here. The work of Riehemann (1997) is also compatible in its hierarchical approach to derivational patterns.

To turn this hierarchy briefly and (somewhat) simplistically into a DATR representation (adapted from Corbett and Fraser 1993:135-36):

```
NOMINAL:  <stem> == "<infl_root>"                     1
      <phon stem hardness> == hard                    2
      <mor stem hardness> == "<phon stem
                             hardness>"               3
      <mor acc> == "<mor nom>"                        4
      <mor acc pl animate> == "<mor gen pl>"          5
      <mor acc sg animate masc> == "<mor gen sg>" 6
      <mor dat pl> == ("<stem pl>" "<mor
                      theme_vowel>" _m)               7
      <mor inst pl> == ("<stem pl>" "<mor
                      theme_vowel>" _m'i)             8
      <mor loc pl> == ("<stem pl>" "<mor
                      theme_vowel>" _x).              9

NOUN:      < > == NOMINAL                            10
      <mor loc sg> == ("<stem sg>" _e)               11
      <mor nom pl> == ("<stem sg>" _i)               12
      <mor gen pl> == "< "<mor stem hardness>"
                      mor gen pl>"                   13
      <soft mor gen pl> == ("<stem pl>" _ej)         14
      <mor theme vowel> == _a                        15
      <syn cat> == n.                                16

N_O:       < > == NOUN   % traditional o-stems       17
      <mor gen sg> == ("<stem sg>" _a)               18
      <mor dat sg> == ("<stem sg>" _u)               19
      <mor inst sg> == ("<stem sg>" _om).            20

N_I:       < > == N_O                                21
      <formal gender> == masc                        22
      <mor nom sg> == "<stem sg>"                    23
      <hard mor gen pl> == ("<stem pl>" _ov).        24
```

This fragment (of a fragment) of a grammar is designed to show both default inheritance (< > == X) and the node specific facts which introduce new information (<formal gender> == masc)[9]. Using fact-indices (the line numbers at right, above), and given the following lexical entry for the noun *zakon* ‘law’:

```
      Zakon          < > == N_I
                     <gloss> == law
                     <infl_root> == zakon            25
                     <sem animacy> == inanimate
```

the rules of inference used in the inflected forms in a paradigm of class N_I are as follows:

	I *zakon* ‘law’	Rules used
nom sg	*zakon*	1, 23, 25
acc sg	*zakon*	1, 4, 23, 25
gen sg	*zakon**a***	1, 18, 25
dat sg	*zakon**u***	1, 19, 25

---

[9] Feature values stipulated at a node can also override default values, e.g., for N_IV (not shown), <mor nom pl> == (“<stem pl>” _a), which overrides the value (“<stem pl>” _i) it would otherwise inherit from NOUN (Corbett and Fraser 1993:137).

inst sg	*zakon**om***	1, 20, 25
loc sg	*zakon**e***	1, 11, 25
nom pl	*zakon**i***	1, 12, 25
acc pl	*zakon**i***	1, 4, 12, 25
gen pl	*zakon**ov***	1, 2, 3, 13, 24, 25
dat pl	*zakon**am***	1, 7, 15, 25
inst pl	*zakon**am'i***	1, 8, 15, 25
loc pl	*zakon**ax***	1, 8, 15, 25

Network Morphology also permits *rules of referral* (Zwicky 1985, 1992; Stump 1993a), whereby systematic formal parallelisms not handled by defaults are formalized as a stipulated referral to another form in an analogous paradigm, e.g., for N_III, the value for <mor nom sg> is referred to the corresponding value under N_I, whereas the value for N_III's <mor gen sg> is referred to that of N_II. These referrals are ways of expressing parallelisms not predicted by hierarchical inheritance patterns.

Network Morphology offers a rich formal system for the representation of lexical patterns. It was designed with computational implementation in mind, and so there is a practical advantage for choosing this framework. It is clear from the above examples and from the hierarchical lexicon approach in general that all morphology is handled in the lexicon—derivation mapping from one lexeme to another, and inflectional patterns handled through defaults and overrides as one moves down the path from the most general lexical class to specific lexical entries. This clearly implies that Network Morphology is realizational in its approach, since the formal markings are values for abstract attributes in the feature representation. Once a large enough grammatical fragment is built and particular lexical entries are introduced into the model, the program's output is the full inflectional paradigm of each lexeme, marked with <syn gender> and <syn animacy> values (see Corbett and Fraser 1993:139-41).

The formalism and level of detail needed for computational implementation might be off-putting or even irrelevant for some potential consumers of morphological theory. Some might also question the license to split subtypes of subtypes with no defined limit. The inheritance metaphor, however, makes clear predictions, and the possibility of computational implementation of grammars compiled using this model make for a very appealing (virtually) empirical check on the correctness of predictions. Since correct output does not necessarily guarantee the optimal description, Network Morphology's reliance on default inheritance supplies the impulse to minimize redundancy in the lexical representations. As a descriptive tool and as a computational input, Network Morphology is designed with the future of linguistic research in mind.

Brown, Dunstan. 1998. "Stem indexing and morphonological selection in the Russian verb: A Network Morphology account." In Ray Fabri, Albert Ortmann, and Teresa Parodi, eds. *Models of Inflection* [Linguistische Arbeiten 388]. Tübingen: Max Niemeyer Verlag. pp. 196-224.

Brown, Dunstan, Greville Corbett, Norman Fraser, Andrew Hippisley, and Alan Timberlake. 1996. “Russian noun stress and Network Morphology.” *Linguistics* 34: 53-107.

Corbett, Greville G., and Norman M. Fraser. 1993. “Network Morphology: A DATR account of Russian nominal inflection.” *Journal of Linguistics* 29: 113-42.

Fraser, Norman M., and Greville G. Corbett. 1995. “Gender, animacy, and declensional class assignment: A unified account for Russian.” In Geert Booij and Jaap van Marle, eds. *Yearbook of Morphology 1994*. Dordrecht: Kluwer. pp. 123-50.

http://www.surrey.ac.uk/LIS/SMG/surrey_morphology_group.

13. Paradigm Function Morphology

Morpheme-based					☑	**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar					☑	**In lexicon**
Phonological formalism			☑			**Syntactic formalism**
Incremental					☑	**Realizational**

Paradigm Function Morphology (PFM) is a lexeme-based realizational theory of inflectional morphology. Stump’s work owes much to the theory and metatheory of Arnold Zwicky, and PFM spells out in detail some of the leading ideas of Zwicky’s (e.g., 1987b) Interface Program. Although PFM’s introduction in the literature is generally taken to be Stump (1991), some important precursors may be gleaned in his less formally-oriented (1990) article:

> The proposed framework embodies a conception of the boundary between inflection and derivation that is wholly at odds with the split morphology hypothesis. In particular, this framework does not treat inflection as an extralexical phenomenon but instead presupposes that all morphological processes operate in the lexicon. It does not presume that all rules of derivation inherently precede all rules of inflection but instead allows some intermixture of inflection with derivation ... inflectional and derivational processes are distinguished according to the kinds of expressions that they produce.... (116-17)

With the exception of Matthews (esp. 1972) and Carstairs (e.g., 1983, 1987), PFM gives unusual prominence to the paradigm as an organizing principle in morphology. Many theories have a nodding acquaintance with the paradigm, but treat it as an epiphenomenon, something with pedagogical or perhaps only curiosity value (e.g., Anderson 1992:79-80). This does not mean that PFM treats the paradigm as a primitive, however.

The *paradigm* is a set of cells defined by the universal and language-specific co-occurrence restrictions on morphosyntactic features and their permissible values. Every *cell* in the paradigm, therefore, corresponds to a complete well-formed set **σ** of morphosyntactic *properties* (i.e., feature-value pairs). For the paradigm of a lexeme **L**, the form **Y** which occupies a given cell is the realization of the corresponding set σ on the root **X** of **L**. The eponymous *paradigm function* (PF) is a mapping from a root-pairing <X, σ> to a (word-)pairing <Y, σ>, that is, to an inflected word with the property set appropriate to the cell it appears in. A paradigm function is in turn "defined in terms of more specific *realization rules"*—the individual rules of morphology realizing the language's morphosyntactic properties" (Stump 2001:33).

The formalism of PFM is both rigorous and interpretable within the traditions of formal linguistics. That said, however, there are a few barriers to clear interpretation. Each and every realization rule, for example, bears a triple subscript: the rule block ***n*** that the rule belongs to, the proper subset **τ** of σ that the rule participates in realizing, and the lexeme class **C** whose paradigm function the rule participates in defining.

Rule format: $RR_{n,\ \{\tau\},\ [C]}(\langle X, \sigma\rangle) =_{def} \langle Y', \sigma\rangle$

All of this appears before the root-pair <X, σ> is encountered, and well before the effect of the rule on the root X, i.e., **Y'**, is encountered. If strict attention is not paid to a sometimes quite long string of subscripted shorthand abbreviations, it can be difficult to keep track of the point being made in each rule (Stewart 2000).

For example, from Old Norse:

$RR_{\mathbf{2,\{MOOD:indic,VOI:act,TNS:past,PER:3,NUM:pl\},[V]}}(\langle X,\sigma\rangle) \quad =_{def} \langle X\mathit{u},\sigma\rangle$

Stump (2001) has made the articles which preceded it more accessible.

In PFM, rules of all sorts, and consequently PFs as well, are seen as static well-formedness conditions holding between lexical roots and stems, between stems and inflected words. This is in keeping with other non-derivational approaches to linguistics, e.g., HPSG, but the different perspective can be misleading if one takes the descriptive model to represent a derivation in the traditional sense of the word. The step-by-step demonstration of rule evaluation is therefore more on the lines of a logical proof, but the fact that a proof generally looks not unlike an incremental building up of complex morphological structure (at least in the horizontal dimension) certainly renders it an 'apparent derivation'.

A key concept in PFM is that of the *rule block*, mentioned in passing above. Stump (2001:33) likens the block to Anderson's (1992:129) use of the same term. An important difference exists, however, between the two conceptualizations. A-Morphous-type blocks were motivated as a response to cases of disjunctive rule application; there is no independent motivation or principle which allowed the rule block to cohere. PFM blocks, by contrast, correspond to the traditional notion of a *position class*, whereby "rules belonging to the same block compete for the same position in the sequence of rules determining a word's inflectional exponence" (Stump 2001:33). "Same position" here is more literally construed than the disjunctions in Anderson (1986, 1992), such that a PFM block of realization rules corresponds to a "slot" in a word's sequence of inflectional affixes. PFM rule blocks, therefore, are organized according to the distributional facts of exponence, and not of the more abstract notion of disjunctive rule application. PFM gets disjunctive application for free, as it were—since no more than one exponent can appear in a given slot, no more than one rule from the same block may apply in the definition of a given PF.

Reference to slots while at the same time eschewing morphemes as objects opens PFM for some criticism, because (as happened to the MSRs in A-Morphous Morphology, above) zeroes can take up residence in vacant positions. To counter this possibility, and in keeping with the "function" mentality, absolutely no structural zeroes are allowed for in PFM. Where no rule in a block is applicable to <X, σ>, a universal realization rule applies, the Identity Function Default, mapping the input onto itself (Stump 2001:53, 143):

Identity Function Default (IFD):
$RR_{n, \{\}, U}(<X, \sigma>) =_{def} <X, \sigma>$

Here, n ranges over all rule blocks, { } is the empty set of morphosyntactic properties, and U is the class of all lexemes. The IFD, therefore, is effectively the last rule in every rule block, guaranteeing that a proof never fails because some slot in the PF was undefined for lack of an applicable rule. There is no question of "adding Ø"—the IFD evaluation of the block is "no change."

On the issue of rule ordering, PFM denies the need for extrinsic rule ordering. By Pāṇini's Principle (no disjunctivity rider required, cf. A-Morphous Morphology, above): given any complete set s of morphosyntactic properties appropriate to a particular lexeme class and any lexeme in that class, "the value of the ... PF for the root-pairing <X, σ> is always the result of applying the NARROWEST APPLICABLE RULE" in each of the blocks mentioned in the PF schema (Stump 2001:52). A PF schema identifies which rule blocks are involved in the definition of the form realizing the set σ on the root X of lexeme L, e.g.:

$$PF(<X, \sigma>) =_{def} Nar_3(Nar_2(Nar_1(Nar_0(<X, \sigma>))))$$

Narrowness, then, is evaluated between realization rules in terms of the relative specificity of the set of morphosyntactic properties realized by each rule. This is the method for enforcing the Pāṇinian Principle, i.e. proper subset precedence.

The Identity Function Default is, for PFM, the "default default," meaning that where no special case is called for, the IFD takes over. The default-override relation is crucial in PFM, as it is in Network Morphology (above). Defaults are what lexemes in a particular class 'inherit' by virtue of class membership, provided that they are not simultaneously members of a more select class (a proper subset of the larger class, of course) which is subject to a special override rule. The Narrowness relation is simply a principled (rather than extrinsic or arbitrary) and formal way of deciding, between two realization rules, both of which are applicable in a given case, which would override the other (subject to further override by some third rule, narrower still than either of them). Defaults are therefore layered, and the prediction is that the Pāṇinian Principle will always be adequate for the unique determination of the narrowest applicable rule, given the joint assumptions that blocks are position classes and that no block is ever undefined thanks to the IFD (Stump 2001:21-25).

Lexemes each possess a *stem set* consistent in number of stems at least with the other members of the same lexeme class. In the general (non-suppletive) case, stems will be related to the root or another stem by rules of *stem formation*, purely formal operations (Stump 2001: 183-86). If stems occupying corresponding positions in the stem sets of comparable lexemes in distinct inflectional classes are not characterized by parallels in general phonological shape, purely morphological (*morphomic*, in the sense of Aronoff (1994:22-29)) rules of *index assignment* come into play, marking stems as 'strong' versus 'weak', e.g., or assigning arbitrary numerical indices (Stump 2001: 190-194)[10]. Rule block 0 in any given language is a block of *stem selection* rules that identify the morphosyntactic properties each stem may (partially or wholly) realize (Stump 2001: 175-79). In this way, regular (and/or productive) stem-internal non-concatenative marking may be handled by stem formation rules, and the Separation Hypothesis is still respected, since rules of selection and formation are in principle independent.

The evaluation of particular realization rules is stated as a default phonological entity, which implies that the default shape of the exponent may be overridden under specific circumstances. An unordered set of morphophonological rules constrains the evaluation of each realization rule in any instance of its application. For any given application of a randomly chosen rule, any number of morphophonological rules (including none) may affect the phonological shape of the rule's evaluation. Where whole blocks of realization rules or an identifiable subset of rules in a block is subject to one or more particular morphophonological rules, a *morphological metageneralization* may be stated concerning those rules to account for this subregularity (cf. *meta-templates*/*meta-redundancy-rules* in Janda and Joseph (1992a and b, 1999)).

PFM is more limited in its scope than many of the other theories considered here, for example, in that only the barest intimations of how to handle derivation and compounding, let alone cliticization, have appeared to this point (Stump 1995; but see Spencer 2004 for a proposed extension). No particular theory of syntax has been assumed as an input to PFM, although it has been identified as a promising interface for HPSG by Kathol (1999). Although PFM has been compared to A-Morphous Morphology as coming from a similar theoretical perspective, a much closer affiliation is to be found with Network Morphology (above) in the shared reliance on features, defaults, lexical classes and subclasses, and the paradigm as an organizing principle. One clear distinction there is PFM's tying rule blocks to position classes directly, whereas this does not seem to be captured in the Network Morphology approach. An empirical examination in this area might well prove a useful line of study to determine the necessity/redundancy of such an assumption.

---

[10] If a stem is used in the realization of all and only the occurrences of some morphosyntactic property, say {TNS:past}, it may be (mnemonically) useful to use an index which reflects this use, i.e., identify a "past stem." This does not, of course, entail that all indices for the particular stem set must bear functionally-defined indices. From a realizational perspective, function-based names can give a (misleadingly) morphemic cast to an element of form.

Stewart, Jr., Thomas W., and Gregory T. Stump. 2007. “Paradigm Function Morphology and the Morphology/Syntax Interface.” In *The Oxford Handbook of Linguistic Interfaces*. Edited by Gillian Ramchand and Charles Reiss. Oxford University Press, pp. 383-421.

Stump, Gregory T. 1990. “Breton inflection and the Split Morphology Hypothesis.” *The Syntax of the Modern Celtic Languages* [Syntax and Semantics 23], ed. by Randall Hendrick, 97-119. San Diego: Academic Press.

Stump, Gregory T. 1991. "A paradigm-based theory of morphosemantic mismatches." *Language* 67(4): 675-725.

Stump, Gregory T. 1992. "On the theoretical status of position class restrictions on inflectional affixes." In Geert Booij & Jaap van Marle, eds. *Yearbook of Morphology 1991*. Dordrecht: Kluwer. pp. 211-41.

Stump, Gregory T. 1993a. "On rules of referral." *Language* 69(3): 449-79.

Stump, Gregory T. 1993b. "Position classes and morphological theory." In Geert Booij & Jaap van Marle, eds. *Yearbook of Morphology 1992*. Dordrecht: Kluwer. pp. 129-80.

Stump, Gregory T. 1993c. “Reconstituting morphology: The case of Bantu preprefixation.” *Linguistic Analysis* 23(3-4): 169-204.

Stump, Gregory T. 1995. "The uniformity of head marking in inflectional morphology." In Geert Booij & Jaap van Marle, eds. *Yearbook of Morphology 1994*. Dordrecht: Kluwer. pp. 245-96.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: CUP.

14. Prosodic Morphology

Morpheme-based	☑					**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar		☑				**In lexicon**
Phonological formalism	☑					**Syntactic formalism**
Incremental	☑					**Realizational**

Prosodic Morphology is an outgrowth of Autosegmental Phonology (Goldsmith 1976). Proposals made in McCarthy’s (1979) analysis of Classical Arabic, and distilled somewhat in McCarthy (1981), gave rise to an approach that escapes the limitations of the two-dimensional trees of Word Syntax (see below). In McCarthy (1981) the task is to accommodate non-concatenative morphology into the same basic scheme as concatenative morphology. In order to accomplish this, McCarthy invokes the abstract multidimensional representations, or tiers, found in Autosegmental theory. If every morpheme is represented on its own tier, root and non-root morphemes are more parallel at the formal level. The asymmetry comes in the form of a prosodic skeleton, to which the segmental and/or featural content of the morphemes is mapped on an independent basis. This allows for the retention of discrete morphemes, while allowing the parts of these morphemes to appear discontinuously in the output string, a result not readily permitted in representations of two dimensions (or fewer). Thus, e.g., the Classical Arabic form *kattab* "cause to write (perfective active), the morphemes are /k-t-b/ "write", /-a-a-/ (reducible to /a/ under assumptions of spreading) [perf. act.], and CVCCVC [causative]. McCarthy (1981:385) exemplifies fifteen abstract morphological classes for the (majority) triconsonantal roots of Classical Arabic, choosing to refer to the classes by the established Hebrew term, *binyan*(*im*).

The analysis in McCarthy (1981) requires a number of stipulative exceptions to "unmarked" patterns of association between segments and the skeletal slots, e.g., in cases where the middle of three consonants spreads, rather than the more usual "one-to-one, then spread from the last attached segment to fill the remainder of appropriate slots" (which would give **katbab* instead of attested *kattab*, mentioned above). The device of *preassociation* allows for certain overrides of the unmarked association patterns, whereby one could say "attach edge segments, then fill remainder by spreading as yet unattached segments." McCarthy proposes this, with the functional explanation that failing to ensure that at least edge elements are attached before spreading may have the consequence of obscuring the root's identity (a foreshadowing, perhaps, of faithfulness and opacity concerns in his later Optimality Theory work)(McCarthy 1982:204-05, 213-14, 221).

Marantz (1982) capitalizes on the descriptive success of McCarthy's framework, testing it on reduplication data from several languages. Whereas McCarthy (1981) used the skeletal tier as a sort of output template to be filled in, Marantz (1982:437) suggests that affixes as well as stems can be segmentally underspecified, that "most reduplication processes are best analyzed as the affixation of a consonant–vowel (C–V) skeleton, itself a morpheme, to a stem. The entire phonemic melody of the stem is copied over the affixed C–V skeleton and linked to C and V 'slots'." Defining a complete copying operation from which 'leftovers' can be 'stray-erased', and the segments or features within which can be overridden by preattached values (Marantz 1982:444), perhaps excessively powerful, but given that there are languages which use total reduplication, a single universal operation based on the limiting case is actually conceptually simpler. The fact that other languages reduplicate no more than one or two segments in all cases undercuts the universal appeal somewhat, but there is a case to be made either way (cf. the l-reduction approach to reduplication in Hoeksema and Janda (1988:221-25)).

A real advantage of Marantz's (1982) presentation is the involvement of a richer and independently motivated prosodic hierarchy (also developed in Halle and Vergnaud 1980) in the description of the different abstract shapes that affixes can take. The limiting case, "normal affixation" is the addition to a stem of a morpheme which is fully specified, all the way to the segmental level, borrowing nothing from the content of the stem (Marantz 1982:456). Yidin^{y} reduplication copies the first two syllables of the stem, regardless of their segmental (C–V) composition (453). The more frequently encountered reduplication types are somewhere in the middle, then, with a specific C–V skeleton, and perhaps some limited segmental and/or featural preassociation (449). From this perspective, morphological operations and different morpheme types are formally united in a plausible way. Perhaps the start from nonconcatenative processes led to this more evenhanded treatment of the two types, concatenative and non-concatenative. The skewed relative frequency of “normal affixation” versus the much less common reduplication cross-linguistically is unpredicted, however.

Akinlabi (1996) gives an indication of the survival of the approach into the Optimality Theory paradigm. Akinlabi, although dealing with putative morphemes which are no larger than features or sets of features, hopes to account for these as edge-oriented affixes. In the constraint-based framework of Optimality Theory (McCarthy and Prince 1993, 1994), constraints which align prosodically-defined elements such as syllables, feet, and (prosodic) words are commonly employed to describe positional affinities between one level of the prosodic hierarchy and another when, all else being equal, independent positioning might be assumed. Akinlabi seeks to adjust the terrain, positing constraints which ALIGN particular morphemes to particular prosodic constituent edges (243):

Featural Alignment
ALIGN (PFeat, GCat)
A prosodic feature is aligned with some grammatical category.

What this fails to take into consideration, and what McCarthy (1981, 1982), Marantz (1982) and Halle and Vergnaud (1980) before them failed to emphasize, is that *morpheme* is not part of the prosodic hierarchy. Because the phonological material in a given word owes its existence, in the general case, to some element of meaning or grammatical function, and there is therefore some dependency between a morpheme and its spell-out (“Pfeat is the featural spellout (or content) of the morphological category in question” (Akinlabi 1996:243), Prosodic Morphology sees no obstacle to positing a hierarchy:

root > morpheme > syllable > C–V skeleton > segment > feature

The comparability of morphemes and syllables is limited, since meaning attends the one but not the other. The question of where (or whether) to place ‘foot’ in the above shows the grafting of one dimension into another. To base an analysis on correspondences between the phonological and the morphological, especially when one is presuming to propose universal constraints (as OT analyses explicitly presume), is to open oneself up to criticism of allowing too liberal a formal representation. Because these levels are not always spelled out exhaustively in the examples given in Prosodic Morphology (although McCarthy 1982:213, e.g., comes close), it is easy to ignore the questionable telescoping that is going on.

For Akinlabi, the placement of a featural affix is part of the lexical entry of that morpheme; determining whether it is a prefix (i.e., placed relative to the left edge of the stem) or a suffix (relative to the right) is based on evidence for directionality of autosegmental association. A featural suffix, for example, will tend to have its effect at the right edge of the stem, but depending on the relative strength of feature cooccurrence constraints and faithfulness constraints, the suffix may be forced further in from the edge, or else be blocked from applying. Within the formalism of OT this is fairly ingenious, despite some of the questionable underlying assumptions. As a brief example, Chaha labialization is claimed to be a featural suffix, realized on the rightmost stem consonant which may be labialized (coronal consonants may not be labialized, although labial and velar consonants can). The feature links once only, to the rightmost licensing consonant, potentially linking to an initial consonant if there are only coronals after it in the word. In case of a stem with only coronal consonants, the feature does not link and thus is not phonetically realized. Given these details, a constraint hierarchy of *COR/LAB>>PARSE>>ALIGNR (249). A particular coup for this approach is the factorial typology given and exemplified in an appendix to the article (283):

Co-occurrence	>>α>>	Alignment	>>β>>	Parse	*Nuer continuancy*
Co-occurrence		Parse		Alignment	*Chaha labialization*
Parse		Co-occurrence		Alignment	*Japanese mimetics*
Parse		Alignment		Co-occurrence	*Aka voicing*
Alignment		Co-occurrence		Parse	*Athapaskan [–continuant]*
Alignment		Parse		Co-occurrence	*Aka voicing, Zoque palatalization*

Except for the Japanese mimetics, however, one would hardly know this was a morphological analysis. The categories realized by the various featural affixes are backgrounded throughout the article, in an effort, it would seem, to cast this as nothing other than phonological theory. Simultaneously, therefore, Akinlabi (1996) displays the inheritance from the earlier work in Prosodic Morphology and regresses theoretically to a more concatenative ideology.

Prosodic Morphology, although undergoing some significant transformations in its transition into constraint-based (OT) analyses, is an approach that the phonologically-minded may take to readily. Despite the several caveats in the above, there is clearly something of value in this method of representing the phonological aspect of morphology. One must remember, however, that the insights of multi-tiered representations can collapse into the same plane if viewed from a different angle.

Akinlabi, Akinbiyi. 1996. "Featural affixation." *Journal of Linguistics* 32: 239-89.

McCarthy, John J. 1981. "A prosodic theory of nonconcatenative morphology." *Linguistic Inquiry* 12(3): 373-418.

McCarthy, John J. 1982. "Prosodic templates, morphemic templates, and morphemic tiers." In Harry van der Hulst and Norval Smith, eds. *The Structure of Phonological Representations, Part 1*. Dordrecht: Foris. pp. 191-223.

McCarthy, John J., and Alan Prince. 1993a. Generalized alignment. *Yearbook of morphology 1993*, ed. by Geert Booij and Jaap van Marle, 79-153. Dordrecht: Kluwer.

McCarthy, John J., and Alan Prince. 1993b. Prosodic morphology I: constraint interaction and satisfaction. University of Massachusetts, Amherst, and Rutgers University ms.

McCarthy, John J., and Alan Prince. 1994a. The emergence of the unmarked: Optimality in prosodic morphology. *Proceedings of the 24th annual meeting of the Northern-Eastern Linguistic Society*, ed. by Merce Gonzales, 333-79.

McCarthy, John J., and Alan Prince. 1994b. Prosodic morphology: an overview. Paper presented at the Utrecht Workshop on Prosodic Morphology.

Marantz, Alec. 1982. "Re reduplication." *Linguistic Inquiry* 13(3): 435-82.

15. Word-Syntax

Morpheme-based	☑					**Word/Lexeme-based**
Formalist	☑					**Functionalist**
In grammar	☑					**In lexicon**
Phonological formalism					☑	**Syntactic formalism**
Incremental	☑					**Realizational**

The approach to morphology called Word Syntax has a special position in linguistic theory, especially in the area of GB-style syntax and its descendants. It owes much to the classic Item-and-Arrangement (IA) approach (Hockett (1954)). In Word Syntax, morphemes are the essential building blocks of words. Bound morphemes differ from free morphemes solely in that the bound morphemes subcategorize for a stem of a certain category to attach to. The name Word Syntax is an obvious choice for this approach, because one need only glance at an analysis to see the overt parallels being drawn between words and phrases. Lieber's (1981) dissertation is held up as an example of the Word Syntax movement in its crystallizing phase. Morphology from this perspective is first and foremost about the concatenation of discrete meaningful units, namely morphemes, and the binary-branching tree-structures constitute an account of how a morphologically complex expression comes to have the meaning and morphosyntactic features it does.

In Lieber (1981) and in Williams (1981), much attention is paid to the notion of headedness in morphologically complex words. Williams (1981:248) proposes that the rightmost morpheme at any level of morphological concatenation is the 'head' of the construction (his Righthand Head Rule, or RHR), i.e., that for any concatenation of two morphological elements, the element on the right determines the category and attributes of the resulting expression.

Lieber (1981) proposed the mechanism of feature percolation as the means of transmitting attributes from a constituent morpheme upward to a larger construct. Williams's RHR "works" for much of English derivational morphology and compounding, since English endocentric compounds are almost invariably right-headed, and since nearly all category-changing affixes in English are suffixes. One need not search too far to discover languages with systematic left-headed compounding (Italian, Gaelic, Vietnamese) and even within English there are a few category-changing prefixes ([*en*-[*noble*]$_A$]$_V$). Inflectional affixes in English are invariably suffixal, but part of the definition of inflection is that it cannot change the category of the word it applies to. The Spanish diminutive suffix *–ito*/*-ita* can attach to nouns, adjectives and adverbs, producing in each case a semantic change only, crucially being 'transparent to the category of the word it attaches to, quite unlike a head is predicted to behave (Di Sciullo and Williams 1987:26).

The RHR is clearly not adequate as a general principle of morphology, but perhaps a revision could redeem it? Lieber (1981) and Selkirk (1982) both reject the RHR as originally defined as simply being too strong. They both suggest alternatives to strict right-hand percolation, allowing for so-called "back-up percolation" in cases where the whole expression has attributed present in some non-head morpheme but not present in the head (Selkirk (1982:76):

Percolation (revised)

a. If a head has a feature specification [aF_i], a≠u[nspecified], its mother node must be specified [aF_i], and vice versa.

b. If a nonhead has a feature specification [bF_j], and the head has the feature specification [uF_j], then the mother node must have the feature [bF_j].

This allows nonhead features to be percolated to the construct, but only if the head has no non-null specification of its own to contribute for the feature in question. Prefixation is still a potential problem if multiple prefixes were to have conflicting specifications for the same feature. It may be that this situation never arises, especially if we assume strict binary branching, but there is nothing to rule it out in principle.

In Di Sciullo and Williams (1987:26), acknowledging some serious empirical problems for the RHR as originally defined, a relativized notion of head is put forward:

"The head$_F$ of a word is the rightmost element of the word marked for the feature F."

This permits a multiply affixed word to have several heads simultaneously, effectively allowing any morpheme to determine some categorial quality of the derivative. Prefixes cannot determine category, however, because the root is always to the right of them, and the root is always marked for at least grammatical category. Thus the facts like *ennoble* still are unexplained, and the predictiveness of the original hypothesis is severely weakened. As for left-headed compounds, Di Sciullo and Williams claim that such constructions in Romance languages aren't really compounds, but rather they are "phrases reanalyzed as words" (83, contra Selkirk 1982:21). The argumentation is less than conclusive, given the semantic idiosyncrasy of some of the expressions and the failure of agreement in at least some cases. The bottom line for Di Sciullo and Williams's proposed amendments to those of Lieber (1981), Williams (1981), and Selkirk (1982) is a weaker model overall and a smaller but remaining empirical problem.

Fabb (1988) proposes doing almost all word formation in the syntax (at least all productive affixation), with separate affixal nodes in the phrase marker, and concatenation via head movement. Di Sciullo and Williams (1987:87) disapprove of such intermingling of syntax and morphology as engendering a loss of generality in both morphological and syntactic rules. Developments in GB syntax converged with the idea of inflection in syntax, such that verb inflection (and sometime noun inflection as well) is performed (or, alternatively, 'checked') by the movement of lexical heads through a sequence of functional heads, each of which contains a morphosyntactic value appropriate to the clause in question, and often associated with overt inflectional morphology. Once head movement is complete, an inflected lexical head appears in S-structure as input to PF.

Whether the affixes are actually represented in the tree structure under the appropriate functional heads is a decision not without implications. The Lexicalist Hypothesis (Chomsky 1970) makes a qualitative distinction between syntax and the internal structure of words. Despite some formal similarities including apparent hierarchical relations among at least the derivational morphemes in a morphologically complex word, syntax does not have access to, and therefore cannot make reference to, any internal structure of the words which might appear in syntactic constructions. This point is recast in Selkirk (1982:2), "The category Word lies at the interface in syntactic representation of two varieties of structure, which must be defined by tow discrete sets of principles in the grammar." "Doing affixation in the GB syntax" as Fabb (1988) would have it, is clearly contrary to the Lexicalist position.

In 1992, Lieber re-entered the fray with an overtly syntactic approach to word formation, *Deconstructing Morphology*. Specifically in response to lexicalized phrases (which Di Sciullo and Williams (1987) looked to as a safety net against the falsification of the RHR), Lieber sees a need to intermingle principles of phrase-building and principles of word-building (21). Again it is claimed that all morphemes have lexical entries, and most, if not all, have syntactic categories of their own. Morphemes are thus X^0 elements to be inserted in syntactic tree structures. Allowing for unlimited recursion at the X^0 level, Lieber can concatenate any number of morphemes into a complex X^0 without untoward results in the X-bar syntax (37). The assimilation of morphology to syntax is fairly completed by the introduction of the notions of complement, specifier, and modifier morphemes, alongside the existing notion of head; Lieber assumes that parallel terms mean parallel behavior "above and below the word level" (39). She modifies some conventional parameter settings found in syntax and dubs them *Licensing Conditions* (38):

a. $X^n \rightarrow ...X^{(n-1,\,n)}...$, where recursion is allowed for n=0.
b. Licensing Conditions
  i. Heads are initial/final with respect to complements.
    - Theta-roles are assigned to left/right.
    - Case is assigned to left/right.
  ii. Heads are initial/final with respect to specifiers.
  iii. Heads are initial/final with respect to modifiers.
c. Pre- or post-head modifiers may be X^{max} or X^0.

With the above as general conditions holding of morphemes as well as words in this expanded view of syntax, the onus is on Lieber to demonstrate that full parallelism obtains. The cost of maintaining this assumption, however, is a series of *ad hoc* replies to empirical problems:

1. English synthetic compounds are left-headed because the construction is a holdover from Old English, when the parameter-settings were different (62-63);
2. Right-headed compounds in French (the only kind that matter, according to Di Sciullo and Williams (1987:83-86)), such as *radioactivité*, are dismissed as non-productive, learned, neo-classical vocabulary, with no import for the parameter-settings (66); and
3. Variable adposition patterning in Dutch is the result of treating the parameters as defaults rather than as true parameters (70-71).

The resulting correspondence between phrasal and word syntax is rough at best. The predictions which follow from Lieber's assumptions are quite strong, if we permit the specifier and complement morphemes, according to her unexpectedly brief presentation of the topic.

The Word Syntax approach to morphology has the formal advantage of making morphology similar or identical to the independently motivated syntactic component. The greater the insistence on assimilation, however, the more adjustments and riders there are to be included in the statement of syntactic rules and principles. Giving each morpheme a lexical entry, but at the same time suggesting that the lexicon is no more structured than a random collection of such entries (Lieber 1992:21)[11] makes one wonder what the lexicon is really good for, other than standing as a legitimizer for the putative equivalence of all morphemes, bound or free[12]. As was mentioned in the introduction to this section, the Word Syntax framework has had considerable influence on the treatment of morphology within the GB syntactic framework. If one is working in the GB/Minimalist framework, Word Syntax might be the most natural choice (but compare DM, above).

Di Sciullo, Anna Maria, and Edwin Williams. 1987. *On the Definition of Word* [Linguistic Inquiry Monograph 14]. Cambridge, MA: MIT Press.

Fabb, Nigel. 1988. Doing affixation in the GB syntax. *Morphology and Modularity* [Publications in Language Sciences 30], ed. by Martin Everaert, Arnold Evers, Riny Huybregts, and Mieke Trommelen, 129-45. Dordrecht: Foris.

Julien, Marit. 2007. On the relation between morphology and syntax. *The Oxford Handbook of Linguistic Interfaces*, ed. by Gillian Ramchand and Charles Reiss, 209-38. Oxford/New York: Oxford University Press.

11 This atom-oriented lexicon stands in contrast to Lieber (1981) and its decidedly more organized contents (complete with stems as well as roots). The agenda there was to move all morphology into the lexicon, and although the tree structures of Word Syntax may be taken more benignly as generalizations about lexical structures, those practitioners taking their cue from Fabb or the functional head movement (no pun intended) are taking a more literally syntactic view.

12 Somewhat ironically, the strongest form of Word Syntax implies that lexicalization is not real, since only single morphemes are inserted at terminal nodes, in keeping with proposals dating back at least to the *Sound Pattern of English* (Chomsky and Halle 1968).

Lieber, Rochelle. 1981. *On the Organization of the Lexicon*. Bloomington, IN: Indiana University Linguistics Club. [1980. MIT dissertation].

Lieber, Rochelle. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory*. Chicago: University of Chicago Press.

Selkirk, Elisabeth O. 1982. *The Syntax of Words* [Linguistic Inquiry Monograph 7]. Cambridge, MA: MIT Press.

Toman, Jindřich. 1998. Word Syntax. *Handbook of Morphology*, ed. by Andrew Spencer and Arnold M. Zwicky, 306-21. Oxford/Malden, MA: Blackwell.

Williams, Edwin. 1981. On the notions "lexically related" and "head of a word". *Linguistic Inquiry* 12(2).245-74.

APPENDIX A

A.1—Introduction: Scottish Gaelic Nouns (Stewart 2004)

I. *doras* (m.) 'door'

	SINGULAR	PLURAL
NOM.	*doras* /tɔrəs/	*dorais* /tɔrɨʃ/
GEN.	*dorais* /tɔrɨʃ/	*dhoras* /ɣɔrəs/
DAT.	*doras* /tɔrəs/	*dorais* /tɔrɨʃ/
VOC.	*a dhorais!* /əɣɔrɨʃ/	

II. *balach* (m.) 'boy, lad'

	SINGULAR	PLURAL
NOM.	*balach* /palax/	*balaich* /paleç/
GEN.	*balaich* /paleç/	*bhalach* /valax/
DAT.	*balach* /palax/	*balaich* /paleç/
VOC.	*a bhalaich!* /əvaleç/	

III. *sgoil* (f.) 'school'

	SINGULAR	PLURAL
NOM.	*sgoil* /skɔl/	*sgoiltean* /skɔltʃən/
GEN.	*sgoile* /skɔlə/	*sgoiltean* /skɔltʃən/
DAT.	*sgoil* /skɔl/	*sgoiltean* /skɔltʃən/
VOC.	*a sgoil!* /əskɔl/	

IV. *clach* (f.) 'stone'

	SINGULAR	PLURAL
NOM.	*clach* /kʰlax/	*clachan* /kʰlaxən/
GEN.	*cloiche* /kʰlɔjçə/	*chlach* /xlax/
DAT.	*cloich* /kʰlɔjç/	*clachan* /kʰlaxən/
VOC.	*a clach!* /əkʰlax/	

General facts of initial mutation (specifically "Lenition"):

p^h ~ f	t^h ~ h	s ~ h*	tʃ ~ ç	k^h ~ x
p ~ v	t ~ γ		dʒ ~ j	k ~ γ
m ~ ṽ	n**			
	l**			

*/s/ is immune to initial mutation before stops (including /m/), e.g., *sgoil*, above.

**/n, l/ are immune to initial mutation in most modern dialects.

A.2: A-Morphous Morphology's response

Inflectional rules take as input pairs {S, M} consisting of a lexically-specified stem and a (contextually appropriate) morphosyntactic representation (*MSR*). The stems in the stem set of a given lexeme are those not characterizable by (partial) suppletion, that is, alternating in ways that are lexically specific and not representative of systematically part of a lexical class. Since initial mutation is regular and productive in terms of its mapping between alternant pairs, it is preferable to capture that as an inflectional word-formation rule. For at least the *doras–balach* class (henceforth class Nα), i-Ablaut will similarly be (part of) a WFR.

<u>Stem sets</u>

Class Nα	Doras: {/tɔrəs/}
	Balach: {/palax/}
Class Nβ	Sgoil: {/skɔl/}
Class Nγ	Clach: {/klojç/ [gen/dat, sg.]; /klax/}

Since none of the forms in the set has multiple specifications for the same feature(s), there is no call for layering in the MSRs.

<u>WFRs</u> (all are +N)

(1) [{+Nom, +Dat}, +sg]
/X/ → /X/

(2) [+Gen, +sg]
/YVC/ → /YV [+high] C/ (Nα)
/X/ → /Xə/

Rule (1) states that the bare stem will be used in the nominative and dative singular. In the case of *clach*, the lexically specified [+Dat] stem will be selected, owing to its greater specificity, and will be used as-is for the dative. In (2), disjunctivity is to be invoked twice:

a. the more specific clause will apply to Nα nouns only, and the second clause will apply elsewhere, and
b. the lexically specified [+Gen] stem will be selected for *clach*.

(3) [+Gen, +pl]
/CY/ → /C'Y/

(4) [+pl]
/YVC/ → /YV [+high] C/ (Nα)
/X/ → /Xtʃən/ (Nβ)
/X/ → /Xən/ (Nγ)

The Elsewhere Condition is in play here, since rule (3) will precede and pre-empt rule (4). Within rule (4), where it does apply, the different clauses are indexed to the lexical class of the input stem, and thus apply disjunctively.

(5) [+Voc]
/CYVC/ → /əC'YV [+high] C/ (Nα)
/X/ → /əX/

In (5) the first clause precedes and pre-empts the second clause. C' is used to indicate the mutated alternant of the corresponding C in the input stem.

A.3: Articulated Morphology's response

It requires some formal ingenuity to represent non-concatenative, non-zero morphology in the AM framework. The following, however, is in keeping with the spirit of what AM rules do.

As for morphological objects in Scottish Gaelic, it seems clear that there are roots, different stems, and words. Case and Number are often marked jointly, and may also be marked in multiple ways on the same inflected word. Defining the morphological objects by means of content is problematic, therefore. Taking the root as the starting point, and since every rule must be information-increasing, the following rules are a significant subset of those required for the paradigms given.

(1) Singular in class N:
X X
[ ] → [N: sg]

(2) Nominative in class N:
X X
[ ] → [Case: nom]

(3) Dative in class N:
X X
[ ] → [Case: dat]

(4) Plural in class Nα:
[...VC] [...V [+high]C]
[ ] → [N: pl]

(5) Genitive singular in class Nα:
[...VC] [...V [+high] C]
[N: sg] → [N: sg, Case: gen]

(6) Genitive plural in classes Nα and Nγ:
[C...] [C'...]

[ ] → [N: pl, Case: gen]

(7) Genitive singular in classes Nβ and Nγ:

X		X*e*
[N: sg]	→	[N: sg, Case: gen]

(8) Vocative singular in class Nα:

[C...V C]		*a* [C'...V [+high] C]
[ ]	→	[N: sg, Case: voc]

(9) Vocative singular in classes Nβ and Nγ:

X		*a*X
[ ]	→	[N: sg, Case: voc]

(10) Plural in class Nβ: (11) Plural in class Nγ:

X		X*tean*	X		X*an*
[ ]	→	[N: pl]	[ ]	→	[N: pl]

In the above rules, C' is used to indicate the mutated alternant of the initial C in the input expression. Class Nβ almost motivates a distinct singular versus plural stem, but Nα and Nγ are not consistent with such a step. The Gen/Dat singular stem for *clach* would seem to be a lexical matter, rather than the stuff of rules.

A.4: Autolexical Syntax's response

Mutation and i-Ablaut are consigned them to the principles of Prosodic Phonology (McCarthy 1981, Marantz 1982), as was proposed in Sadock (1991:26). The remaining few "lexemes" have the following lexical representations:

	-e	*-tean*
Syntax	nil	nil
Semantics	nil	nil
Morphology	N[fem]\N[gen,sg]	N[Nβ]\N[pl]

	-an	*a-*
Syntax	nil	nil
Semantics	nil	nil
Morphology	N[Nγ]\N[pl,{nom,dat}]	N[voc,sg]/N

The Morphology describes appropriate insertion contexts, using Categorial Grammar formalism.

In Autolexical Syntax, stems are considered to be the head of inflected words. Inflections (Y) are introduced by the following general rule (X = N, for the present data set), and then placed with respect to the stem (X[–0]) depending on whether they are prefixes or suffixes:

X[–1] → X[–0], Y

The Case and Number properties would be assigned based on context, whereas declension class would be a lexical property of the noun. All four example lexemes are simple nouns (N[0]), and therefore semantically intransitive predicates (F[-1]) (Sadock 1991:31).

	doras	*balach*	*sgoil*	*clach*
Syntax	N[0]	N[0]	N[0]	N[0]
Semantics	F[-1]	F[-1]	F[-1]	F[-1]
	'door'	'boy'	'school'	'stone'
Morphology	N[-0]	N[-0]	N[-0]	N[-0]

The combination of the affixes and the stems give N[–1], i.e., inflected words in the morphology, once all appropriate inflections are introduced. These are N[0] elements in the syntax, and examples of such inflected words would be the following:

	doras	*bhalach*	*sgoile*	*cloich*
Syntax	N	N	N	N
	[1, nom, sg]	[1, gen, pl]	[1, gen, sg]	[1, dat, sg]
Morphology	N[-1]	N[-1]	N[-1]	N[-1]

Semantics are assumed to be unchanged under inflection.

A.5: Categorial Morphology's response

Whereas affixation is accounted for in Categorial Morphology by addition operations, non-concatenative morphology is effected by means of substitution operations (Hoeksema and Janda (1988)).

First the two-place operations, definable in terms of lexical entry triples on the morpholexically context-sensitive affixes.

-*tean* <Nstem\Nβ\, N, Suff>
-*an* <Nstem\Nγ\, N, Suff>
-*e* <Nstem\N*x*\, N, Suff> Where $x \in \{\beta, \gamma\}$

The Vocative prefix applies in all classes, and so does not require the subcategory specification in its input requirements.

a </Nstem, N, Pref>

These affixes will be added via a cancellation operation—left-cancellation for the suffixes, right cancellation for the prefix.

Initial mutation would have a lexical entry <Nbasic, Nmut, *f*mut>, and its effect, i.e., the operation *f*mut, should be treated with a rule of replacement.

*f*mut (C[–strident, –continuant, α spread glottis]X) = C[+continuant, –α voice]X

The i-Ablaut would parallel mutation to some degree, with and entry <Nbasic, Nablaut, *f*ablaut>, where application is limited to Nα (the class of *doras* and *balach*), and the operation defined as follows:

*f*ablaut (XVC) = XV[+high]C

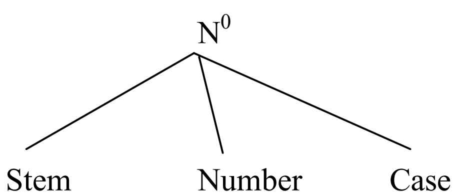
The alternation *a* ~ *oi* in *clach* seems to be separate from this, and so should probably be handled in the lexicon, rather than with a rule that would imply more general applicability. More data would make clear the (lack of) motivation for a separate synchronic ablauting rule.

(Note: Because of the multifunctionality of mutation and i-Ablaut, the entries given above contain purely formal second members, Nmut and Nablaut. Categorial Morphology would typically give more content-specific second members, such as N[+Nom] or the like, and so the above lexical entry formulations are rather more like schemata, containing a variable as the second member, thereby abbreviating (part or all of) several distinct morphological rules. The operations *f*mut and *f*ablaut, however, are defined over strings, and so are phrased appropriately without reference to input and output categories.)

These affixes and operations may be applied singly or jointly to bases, according to the rules of Categorial Grammar.

A.6: Distributed Morphology's response

In each case, Morphological Structure takes the terminal nodes of Surface Structure and creates morphosyntactic feature nodes (plus one for the stem). In order to consider larger structures involving agreement, a Gender node would be created as well.



From this point, morphological operations of Fission and/or Fusion will join or split nodes, depending on the nature of the morphemes to be inserted, e.g., are there multiple exponents (redundantly) marking the same category (fission), or are there morphemes which carry multiple feature specifications (fusion)?

Let's look at the various configurations needed for correct vocabulary insertion.

1. N^0

[Stem, Case, Num.]

doras, *balach*: [Nom, Sg.], [Dat., Sg.]
sgoil: [Nom., Sg.],[Dat., Sg.]
clach: [Nom., Sg.]

STEM IS USED 'AS-IS'.

2. N^0

[Stem] [Case, Num.]

doras, *balach*: [Gen.,Sg.],[Nom., Pl.], [Dat., Pl.]
sgoil: [Gen.,Sg.],[Nom.,Pl.], [Gen.,Pl.],[Dat.,Pl.]
clach: [Gen.,Sg.], [Dat., Sg.], [Nom., Pl.], [Dat., Pl.]

NULL OR OVERT SUFFIX, MAY TRIGGER I-ABLAUT IN STEM.

3. N^0

[Case, Num.] [Stem]

doras, *balach*: [Gen., Pl.]
sgoil: [Voc., Sg.]
clach: [Gen., Pl.][Voc., Sg.]

NULL OR OVERT PREFIX, MAY TRIGGER MUTATION IN STEM.

4. N^0

[Case, Num.] [Stem] [Case, Num.]

doras, *balach*: [Voc., Sg.]

OVERT PREFIX TRIGGERS MUTATION, NULL SUFFIX TRIGGERS I-ABLAUT IN STEM.

In this analysis, structures 1, 2, and 3 presuppose the operation of Fusion, whereas structure 4 requires Fusion, and the Fission of the fused node (these operations are crucially ordered, so as to minimize the number of morphological operations in the derivation).

The analysis above entails the following set of listed affixes:

Affix	**MP rules**	**Meaning**	**Restrictions**
Ø-X	[+mutating]	[+Gen., +Pl.]	Where X = *doras*, *balach*, *clach*...
ə-X	[+mutating]	[+Voc., +Sg.]	Where X = *doras*, *balach*, *clach*...
ə-X		[+Voc., +Sg.]	Where X = *sgoil*...
X-ə	[+i-Ablaut]	[+Gen., +Sg.]	Where X = *clach*,

			sgoil (vacuous)...
X-Ø	[+i-Ablaut]	[+Gen., +Sg.]	Where X = *doras*, *balach*...
X-Ø	[+i-Ablaut]	[+Nom., +Pl.]	Where X = *doras*, *balach*...
X-Ø	[+i-Ablaut]	[+Dat., +Pl.]	Where X = *doras*, *balach*...
X-ən		[+Nom., Pl.]	Where X = *clach*...
X-ən		[+Dat., +Pl.]	Where X = *clach*...
X-tʃən		[+Nom., +Pl.]	Where X = *sgoil*...
X- tʃən		[+Gen., +Pl.]	Where X = *sgoil*...
X- tʃən		[+Dat., +Pl.]	Where X = *sgoil*...

A further morphological operation of feature Deletion would allow a unified [+Pl.] morpheme in the case of *sgoil*, since Case is apparently not distinguished in the plural for that class. Alternatively, one might avoid Case-Number Fusion for the *sgoil* class and unify [+Pl.] that way, but at the cost of a special full set of (homophonous) null case markers. (Note: More data would show that initial <sg-> clusters are impervious to mutation, and so the [+Voc.] prefix can be unified as well.)

This analysis is a fairly conservative, in that a unitary stem is assumed for each "lexeme." It is for this reason that stem alternations are "projected" into the stem's phonological representation from without (cf. Pyatt (1997) for an extended DM analysis of Celtic Initial Mutation, largely consistent with the above methodology).

A.7: Lexeme-Morpheme Base Morphology's response

The analysis here needs to consider only I[nflectional]-derivation and Morphological Spelling, i.e., the realization of the inflectional categories Case, Number, and the inherent category of inflectional class (which may or may not correspond one-to-one with Gender). The grammatical functions for which the various inflected forms may be used are beside the point here.

The Separation Hypothesis permits the treatment of the relationship between inflectional categories and their exponents as a mapping. The evidence given supports treating I and II as instances of the same lexeme-class (call it Nα), and III and IV should provisionally be classes unto themselves (Nβ and Nγ, respectively).

Let us assume that the initial mutations are formally parallel (Note: more data would confirm this), despite some divergence in phonetic detail. All operations on the stem, whether affixations or alternations, are to be considered elements of Morphological Spelling. The lexeme contributes its phonetic representation as an input to MS, and depending on inflectional class, Case, and Number, different MS operations are selected.

Class	Case	Number	Mutation	i-Ablaut	-ə	-ən	- tʃən
Nα	Nom.	Sing.					
		Plur.		⇛			
	Gen.	Sing.		⇛			
		Plur.	⇛				
	Dat.	Sing.					
		Plur.		⇛			
	Voc.	Sing.	⇛	⇛			
Nβ	Nom.	Sing.					
		Plur.					⇛
	Gen.	Sing.			⇛		
		Plur.					⇛
	Dat.	Sing.					
		Plur.					⇛
	Voc.	Sing.					
Nγ	Nom.	Sing.					
		Plur.				⇛	
	Gen.	Sing.		⇛	⇛		
		Plur.	⇛				
	Dat.	Sing.					
		Plur.				⇛	
	Voc.	Sing.					

Two dimensions are unable to capture the complex mapping fully, but the matrix above does make clear the usefulness of a separation between inflectional categories and their exponents, in combination with lexical declension class.

A.8: Lexical Morphology and Phonology's response

Mutation and i-Ablaut in LM&P are level-one phenomena, despite their regularity and productivity, by virtue of the locus of their effects, i.e., the stem. It is difficult to say whether the Vocative prefix triggers initial mutation or not, since the mutation is motivated independently for the Genitive Plural. The plural suffixes do not interact with i-Ablaut (synchronically, anyway), and they do not pile up in the data here, so a precise level assignment for the suffixes is not possible here. To say that they must not apply before level one (the mutation and Ablaut) is not insightful.

The Blocking phenomenon has some interesting implications here, especially in the *sgoil* class, since neither initial mutation nor i-Ablaut is evident. It cannot reliably be determined whether the Plural suffix and the Genitive Singular suffix are applied in addition to level 1 inflection, or whether they apply as a back-up to the non-application of the level 1 inflection.

The identical Genitive Singular suffix is used in addition to a stem alternation in the case of *clach*, i.e., *cloiche*, and so this appears to simply be the Genitive Singular suffix used with Feminine nouns. If, on the other hand, the Plural marker in *sgoiltean* is a backup to initial mutation, the prediction would be that plural nouns which begin with /sp/, /st/, or /sk/ should mark the plural categorically with some affix, -*tean* or otherwise. To verify this prediction, it is necessary to go beyond the given data, and even then the facts are unclear. If the suffix is motivated by blocking, however, one is hard pressed to explain the suffix used in addition to mutation, in attested cases like *ghillean* 'of young men' (cf. *gille* 'young man'). The Elsewhere Condition and the related Blocking effect would not predict this multiple marking.

A.9: Natural Morphology's response:

In every case no form is less 'markered' (*merkiert*, a.k.a. 'featured') than the Nominative Singular, which is unmarked for case and number. In I, II, and III, the Dative is syncretic with the Nominative, and syncretism is considered to be bad semiotically in NM (it is not biunique). If you had to pick a form that was next in line in markedness to the Nominative, however, it would have to be the Dative, so the syncretism could be worse.

For the masculine nouns (I and II), the Plural is more markered than the corresponding singular, and that is in line with iconicity. Also in I and II, the Genitive is more markered than the Nom./Dat., but less markered than the Vocative, which is surely the most marked case of all in these paradigms.

In III, we observe neutralization of case within number, excepting the Genitive Singular. This is unusual in comparison to the other three examples, but chances are that *sgoil* may have been influenced by the cognate word in English English. Some morphological anomaly is less worrisome on that assumption. That the Genitive Plural is not distinguished formally from the other Plural forms is particularly unusual, however, given the other three examples.

As for IV, the syncretism between Nominative and Dative Singular is lost, which is good from a biuniqueness standpoint, but there is a new syncretism with the (marked) Vocative—very unusual on the markedness/iconicity dimension.

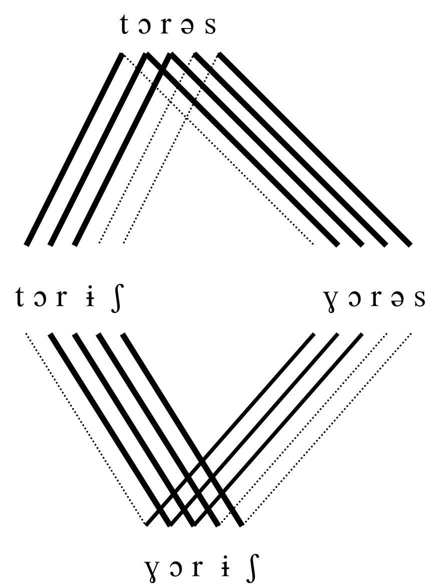
It is true, although perhaps merely by coincidence, that III exhibits the same Nom./Voc. syncretism as IV, so perhaps paradigm III is not as anomalous as it looks. The Genitive Singular in IV is just like the Dative Singular, but with a final /ə/, and the same is true in III, although again less strikingly than the facts in IV.

Although there are some affixes in use here, these paradigms rely to a remarkable degree on Modulator Featured symbolizing, the least optimal symbol type (other than no marker at all). The fact that there is at least one syncretic pair in each number column of each paradigm here would suggest that the case system is under pressure to collapse or to attract a new marker morpheme in one Case (more likely the Dative for I and II, on markedness grounds). It seems the Nom./Dat. distinction is being kept alive by patterns like IV. If IV is a(n unproductive) minority pattern in the language, the pressure to regularize forms like *cloich* to *clach* is quite high, and the Natural Morphology prediction is that Dative case will collapse in time, all else being equal.

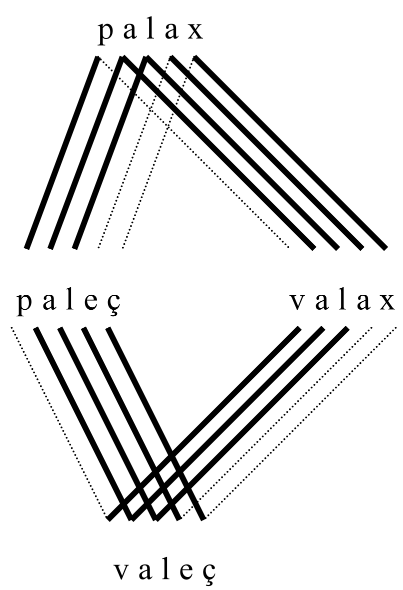
A.10: "Network Model"'s response

Since these are nouns, the relevance hierarchy doesn't really help out here. The first thing to do is to draw networks, and see how they compare:

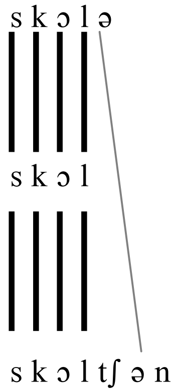
I.



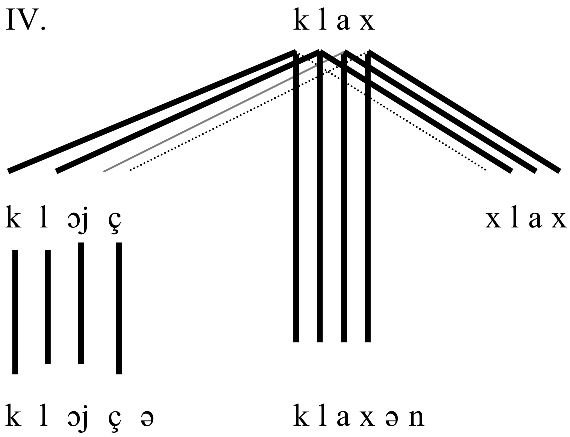
II.



III.



IV.



I and II show a nicely closed network, indicating that there is an element of regularity in these related forms. The fact that the corresponding forms also fill parallel grammatical functions is a sure sign of a paradigmatic pattern. The stronger versus weaker links are even in the same positions with respect to the phonemic sequences. This pattern is predicted to be stable and should be relatively productive.

The alternation of the initial consonant in I (changing place, manner, and voicing) is more distant in phonetic terms than in II (manner and voicing), which is still greater than in IV (manner only). The less the phonetic distance between alternants, the more recoverable the correspondence, and the easier will be lexical access. Class I, therefore, stands out in language independent terms, although if the alternation is productive, that may support the pattern's continued existence.

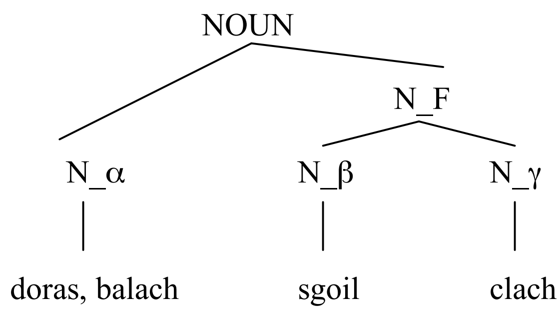
III shows a simpler pattern of identity of the stem across the board with suffixal inflection. The regularity here makes this an even more readily detectable morpheme than the patterns we observe in I and II, but the one-to-many form-meaning mappings undercuts the value of the stem's consistency.

As sets of related forms go, the pattern in IV is quite remarkable. There should be a lot of pressure on this paradigm to regularize at least the vowel quality of the stem. The lexical strength of the word for "stone," however, might be quite high for Gaelic speakers, given the frequency of occurrence of physical stones in the relevant parts of the world. That might explain the irregularity's ability to endure to this point.

It is also important to find out just how productive the vowel quality alternation is in Gaelic nouns more generally, since that may affect the degree to which the alternation may be considered an irregularity. In this limited data set, IV stands out. It would be premature to assume that this sample was representative of the language as a whole or that type and token frequencies can be reliably projected without more evidence.

A.11: Network Morphology's response

The situation here is remarkably similar to the Russian example discussed in the presentation of Network Morphology (above). The four paradigms under discussion here may be seen as belonging to two or three declension classes.



Since the phonology of the mutated and/or ablauted stem clearly depends on the phonology of the root, lexical items will be assumed to have up to four formally distinct yet relatable stems for use in the statement of particular morphological facts. There's more redundancy in the stem set at the phonological level, but this follows from a limitation in the formalism. There ought to be a way to capture the formal correspondences among stems with the First/Last/Rest convention (as used in Hoeksema and Janda (1988) and as used for argument structure in Evans and Gazdar (1995)). Brown (1999:216-17) offers a tentative hierarchical representation of morphophonological selection, but the system is not readily transferable to this case. This will not be pursued here.

```
Doras: < > == N_α
<infl_root> == tɔrəs
<mut_stem> == ɣɔrəs
<i_stem> == tɔrɨʃ
<mut_i_stem> == ɣɔrɨʃ.
```

```
Balach:< > == N_α
<infl_root>== palax
<mut_stem> == valax
<i_stem> == paleç
<mut_i_stem> == valeç.
```

```
Sgoil: < > == N_β
<infl_root> == skɔl
<stem> == “<basic_stem>”.
```

```
Clach: < > == N_γ
<infl_root> == klax
<mut_stem> == xlax
<stem sg> == klɔjç.
```

```
NOUN:  <basic_stem> == “<infl_root>”
       <mor dat> == “<mor nom>”
       <mor nom sg> == “<basic_stem>”
       <mor gen pl> == “<mut_stem>”.
```

```
N_α:        < > == NOUN   %      masculine nouns
            <mor nom pl> == "<i_stem>"
            <mor voc sg> == "a_<mut_i_stem>"
            <mor gen sg> == "<mor nom pl>".

N_F:        < > == NOUN   %      feminine nouns
            <mor gen sg> == "<mor dat sg>_e"
            <mor voc sg> == "a_<mor nom sg>".

N_β:        < > == N_F
            <mor pl> == "<mor nom sg>_tean".

N_γ:        < > == N_F
            <mor dat sg> == "<stem sg>"
            <mor nom pl> == "<mor nom sg>_an".
```

Inheritance principles together with the default/override relation and rules of referral will map the above lexical entries into the paradigms in question. In this way, generalizations between and across declensions are captured, and the fact that Nβ is more similar to Nγ than either is to Nα is captured without making sameness or difference a simple binary choice. Carstairs's (1987) notion of a *macro-paradigm* might therefore cover the relationship between Nβ and Nγ.

A.12: Paradigm Function Morphology's response

The given data show seven paradigm cells for Gaelic nouns. We are dealing, therefore with two morphosyntactic features, {CASE} and {NUM}. The former is an *n*-ary feature with four permissible values: *nom*, *gen*, *dat*, and *voc*. The latter is also *n*-ary, but since the feature has only two permissible values, it is effectively binary. There is only one co-occurrence restriction to mention here, and that is the (apparent) limitation of {CASE:voc} to extensions of {NUM:sg}. Thus the seven cells are defined (4 x 2 – 1 = 7).

Regular and productive stem-internal alternations are to be described as stem formation-rules, and since the formally most differentiated paradigms, *doras* and *balach* (class Nα), show four distinct but related stems, four stems are posited for the class N in general. Since initial mutation is unified from a conditioning perspective but not from a form perspective, it is misleading to render mutation as a quasi-phonological rule. The stated alternations as given below the data set are adequate for the present purpose. I-Ablaut can be simply formulated as a feature changing rule, but even this must be clearly recognized as a morphologically conditioned rule.

The alternation patterns, therefore, are assumed to be static relationships between alternants, Basic-C and Mutant-C for initial consonants, and Basic-V and Ablaut-V for stem-final vowels. Stem-formation rules will be as follows:

Where L is a masculine (=Nα) noun with root C_1YV_nC, each of (a)-(d) implies the other three:

(a) The Basic stem is identical to the root
(b) The Mutant stem has Mutant-C for C_1
(c) The Ablaut stem has Ablaut-V for V_n
(d) The Combo stem has Mutant-C for C_1 and

Ablaut-V for V_n

Where L is a feminine (=Nβ or Nγ) noun with root C_1Y, each of (a)-(c) implies the other two:

(a) The Basic stem is identical to the root
(b) The Mutant stem has Mutant-C for C_1
(c) Refer other stems to Basic stem

Lexically-specified stems such as *cloich* for *clach* override the application of more generally applicable stem formation and selection rules.

Given the limited data set, there is distributional evidence for exactly three rule blocks: a stem selection block (Block 0), a suffixing block (Block 1), and a prefixing block (Block 2). A general paradigm function for Gaelic nouns can be posited as follows:

Where σ is a complete set of morphosyntactic properties for lexemes of category N,

(i) $\text{PF}(\langle X,\sigma\rangle) =_{\text{def}} \text{Nar}_2(\text{Nar}_1(\text{Nar}_0(\langle X,\sigma\rangle)))$

The rule blocks are the following:

Block 0

(ii) $\text{RR}_{0,\,\{\text{CASE:voc}\},\,[\text{N}]}(\langle X,\sigma\rangle) =_{\text{def}} \langle Y,\sigma\rangle$, where Y is X's Combo stem
(iii) $\text{RR}_{0,\,\{\text{CASE:gen},\,\text{NUM:pl}\},\,[\text{N}]}(\langle X,\sigma\rangle) =_{\text{def}} \langle Y,\sigma\rangle$, where Y is X's Mutant stem
(iv) $\text{RR}_{0,\,\{\text{CASE:gen}\},\,[\text{N}]}(\langle X,\sigma\rangle) =_{\text{def}} \langle Y,\sigma\rangle$, where Y is X's Ablaut stem
(v) $\text{RR}_{0,\,\{\text{NUM:pl}\},\,[\text{N}]}(\langle X,\sigma\rangle) =_{\text{def}} \langle Y,\sigma\rangle$, where Y is X's Ablaut stem
(vi) $\text{RR}_{0,\,\{\,\},\,[\text{N}]}(\langle X,\sigma\rangle) =_{\text{def}} \langle Y,\sigma\rangle$, where Y is X's Basic stem

Block 1

(vii) $\text{RR}_{1,\,\{\text{NUM:pl}\},\,[\text{N}\beta]}(\langle X,\sigma\rangle) =_{\text{def}} \langle X\textit{tean},\sigma\rangle$
(viii) $\text{RR}_{1,\,\{\text{CASE:gen},\,\text{NUM:pl}\},\,[\text{N}\gamma]}(\langle X,\sigma\rangle) =_{\text{def}} \langle X,\sigma\rangle$
(ix) $\text{RR}_{1,\,\{\text{CASE:gen},\,\text{NUM:sg}\},\,[\text{N}\beta]}(\langle X,\sigma\rangle) =_{\text{def}} \langle X\textit{e},\sigma\rangle$
(x) $\text{RR}_{1,\,\{\text{NUM:pl}\},\,[\text{N}\gamma]}(\langle X,\sigma\rangle) =_{\text{def}} \langle X\textit{an},\sigma\rangle$

Block 2

(xi) $\text{RR}_{2,\,\{\text{CASE:voc}\},\,[\text{N}]}(\langle X,\sigma\rangle) =_{\text{def}} \langle \textit{a}X,\sigma\rangle$

According to PFM's paradigmatic interpretation of the Pāṇinian Principle, as represented in the formalization of the paradigm function (PF) above, the narrowest applicable rule in each block will apply in defining the evaluation of the PF for any given pair <X,σ> (the Pāṇinian Determinism Hypothesis). No rules in block 1 are applicable to lexemes of class Nα; inflection in that class is accomplished without suffixation. The distinct stem formation rules for masculine versus feminine noun lexemes allow the rules of stem selection to be stated generally across the category N. The following proofs exemplify the preceding analysis:

Where σ = {CASE:nom, NUM:pl},

PF($\langle$*clach*,$\sigma\rangle$)

= $\text{Nar}_2(\text{Nar}_1(\text{Nar}_0(\langle$*clach*$,\sigma\rangle)))$ [by (i)]

=$\text{RR}_{2,\{\,\},[\text{N}]}(\text{RR}_{1,\{\text{NUM:pl}\},[\text{N}\gamma]}(\text{RR}_{0,\{\text{NUM:pl}\},[\text{N}]}(\langle$*clach*$,\sigma\rangle)))$ [by Nar_n notation]

= $\langle$*clachan*,$\sigma\rangle$ [by IFD, (x), and (v)]

Where σ = {CASE:gen, NUM:sg},

PF($\langle$*sgoil*,$\sigma\rangle$)

= $\text{Nar}_2(\text{Nar}_1(\text{Nar}_0(\langle$*sgoil*$,\sigma\rangle)))$ [by (i)]

=$\text{RR}_{2,\{\,\},[\text{N}]}(\text{RR}_{1,\{\text{CASE:gen, NUM:sg}\},[\text{N}\beta]}(\text{RR}_{0,\{\text{CASE:gen}\},[\text{N}]}(\langle$*sgoil*$,\sigma\rangle)))$

[by Nar_n notation]

= $\langle$*sgoile*,$\sigma\rangle$ [by IFD, (ix), and (iv)]

Where σ = {CASE:voc, NUM:sg},

PF($\langle$*balach*,$\sigma\rangle$)

= $\text{Nar}_2(\text{Nar}_1(\text{Nar}_0(\langle$*balach*$,\sigma\rangle)))$ [by (i)]

= $\text{RR}_{2,\{\text{CASE:voc}\},[\text{N}]}(\text{RR}_{1,\{\,\},[\text{N}]}(\text{RR}_{0,\{\text{CASE:voc}\},[\text{N}]}(\langle$*balach*$,\sigma\rangle)))$

[by Nar_n notation]

= $\langle$*a bhalaich*,$\sigma\rangle$ [by (xi), IFD, and (ii)]

Recall that IDF—the *Identity Function Default*—serves, where no more specific rule is applicable within a rule block, to map the input to itself. Thus the block is evaluated, the form is definable, and no formal change to the input is effected, i.e., there are no zero-morphs involved in this analysis. Note that rule (viii) above is an identity function, but it is a separate stipulated override, not a default, partially realizing the properties {CASE:gen, NUM:pl} on lexemes of class Nγ.

Note also that (full or partial) syncretism in these paradigms is handled through the application of defaults, rather than through special rules of referral in the rule blocks. See Stump (1993a) for a discussion of criteria related to the decision 'to refer or not to refer'.

A.13: Prosodic Morphology's response

Assuming that part of the lexical entry for any root is a segmental tier, the mutation effects can be represented as features which are associated to the initial C position in the skeleton, adding or altering features so as to convert the initial C to its mutated counterpart. The same morpheme does not condition a uniform phonological effect on the initial C of the stem, so the Structural Description and Structural Change must be somewhat complex.

The morpheme contains at least the feature specification [+continuant], which overrides the lexical specification for the C_1 slot (vacuously where the stem is continuant initial). Since the stop contrast is one of aspiration rather than voicing, but the fricative contrast is one of voicing, [–α voice] can be a part of the morpheme, sensitive to the setting of [spread glottis] in the root. Since the mutation never results in a change from [+voice] to [–voice], an analysis in which [Voice] is a privative feature is also possible.

Since i-Ablaut seems to be assigned right to left, given its effect on C_n, morphemes triggering i-Ablaut can be formalized so as to attach to V_n of the stem, rather than to a V numbered left to right. Such morphemes will consist of a feature [+high], which will override the lexical specification for V_n's height. This could also be done as a spreading of palatality from C_n of the root, but palatalization of C next to front vowels is general enough in Scottish Gaelic that it needn't be handled in the morphology, separate from phonology.

Since mutation is a matter of changing specifications in roots, rather than filling empty slots in the C–V skeleton, the Prosodic Morphology analysis of mutation is different from Arabic interdigitation or spreading and prespecification in reduplication. This is a more powerful sort of operation than Prosodic Morphology was originally designed to handle.

Using the OT style (Prince and Smolensky 1993), however, the formalism is undaunted. Three constraints could be posited:

ALIGN (Mutation-L, Stem-L)
  A mutated segment must be at the left edge of a stem.
*hC[+stop]
  The sequence /h/ followed by a stop consonant is ill-formed.
PARSE
  An element in the underlying representation must appear in the surface form.

With the constraint ranking ALIGN (Mutation-L, Stem-L) >> *hC[+stop] >> PARSE, ALIGN keeps the mutation at the left edge of the stem. If the co-occurrence constraint were ranked higher than ALIGN, the mutation would be allowed to move in from the left edge just in case it would violate *hC[+stop].

Because underlying s-stop clusters do not license mutation, and because this co-occurrence constraint outranks PARSE, it is better to leave mutation unparsed than to force the /s/ to mutate before a stop.

Even though the mutated alternants are not phonetically parallel, and even though the conditioning for mutation in the data is completely morphological, this formulation within OT makes it seem as though it were driven primarily (if not purely) by segmental and prosodic phonology.

Since i-Ablaut is more restricted in its application than initial mutation, the restriction to the *doras–balach* class might have to be a condition on the Parse constraint, i.e.,

PARSE [Dat., Pl.]$_{\text{Class1}}$.

This mixes general morphological conditioning and particular lexical-class conditioning, but the OT formalism could handle it. The claim that constraints must be universal seems to be at odds with such an idiosyncratic constraint, but the usual counterargument in such cases is that in languages where there is no direct evidence for the constraint, it is assumed to be ranked very low.

As for the "normal affixation" cases, Prosodic Morphology doesn't differ fundamentally from a concatenation account, except that the morphemes are represented as belonging to distinct morpheme-tiers.

A.14: Word Syntax's response

Lieber (e.g., 1992:165-71) has dealt most directly with mutation and Umlaut in the Word Syntactic framework. Lieber's examples of mutation involve a complex affixation whereby an overt affix (a "mutation trigger") attaches to the stem at one point and an empty timing slot is attached adjacent to the segment to be mutated. On analogy with the Fula analysis in Lieber (1992:167-69), the empty timing slot attaches autosegmentally to the stem's initial segment, forming a geminate. The resulting initial geminate is assumed to meet the structural description of a phonological process of "lenition" which produces the observed mutation effects. The fact that no overt affix correlates with the mutation in Genitive Plural forms in Scottish Gaelic means simply that there is a zero affix meaning [+Gen, +Pl] which associates the empty timing slot in initial position. Perhaps both could be handled at once if we assume that the empty timing slot "is" the [+Gen, +Pl] affix, a prefix, although this move is an innovation here, not suggested in Lieber (1992) or elsewhere.

If we claim that the Vocative prefix *a* similarly contributes an empty timing slot just after it, this could add some indirect support for the empty Genitive Plural prefix. In classes Nα and Nγ (but not Nβ), a null [+Gen, +Pl] affix could explain the failure of additional [+Pl] marking, since that would be featurally redundant. If we assume further that the null [+Gen, +Pl] does not apply to Nβ instead of applying with no perceptible effect on the initial, this could explain the application of the [+Pl] suffix in *sgoiltean* [+Gen, +Pl]. (Note: See LM&P's response, however, for discouraging counterevidence from data beyond the set given in this Appendix.)

The analysis of Umlaut is similar to that of initial mutation, since Umlaut strictly speaking is triggered by a vowel in a following morpheme. Lieber (1992:170) appeals to a floating feature ([–Back], for German), which is part of the lexical entry of triggering suffixes. Stems, on this analysis, are underspecified, with only marked values present underlyingly. The floating feature, once associated to the last vowel in the stem, pre-empts the later association of the unmarked value ([+Back], for German). To accommodate the productive (e.g., *dorais* and *balaich*) Gaelic i-Ablaut facts, however, the triggering suffix must be null itself, but carrying a floating [+High], since we observe both /a/ and /ə/ raising (but not fronting) to /ɨ/.

Other affixes in the data contribute inflectional features to the stems they attach to by means of the unexceptional application of affixation and percolation.

APPENDIX B

B.1—Introduction: Georgian Verbs– Agreement Marker Disjunctivity

Georgian verb agreement has provoked much discussion in both morphological and syntactic theory. Co-occurrence facts have resisted principled explanation in just those cases where multiple arguments are present and (apparently) compete for control of agreement marking.

The facts relevant to underived transitive verbs in Georgian are the following:

(1)	1sg.	2sg.	3sg.	1pl.	2pl.	3pl.
“Subject”	*v-*	—	*-s*	*v-…-t*	*-t*	*-en*
“Dir.Obj.”	*m-*	*g-*	—	*gv-*	*g-…*-t	—

In combination, however, the facts are as follows (Stewart 2001, corrected from Cherchi 1999:42):

(2) **The present tense of *XEDAV*, ‘see’** (*shaded cells are reflexives, expressed periphrastically*)

Subj. \ DO	1sg.	2sg.	3sg.	1pl.	2pl.	3pl.
1sg.		*__g__xedav*	*__v__xedav*		*__g__xedav__t__*	*__v__xedav*
2sg.	*__m__xedav*		*xedav*	*__gv__xedav*		*xedav*
3sg.	*__m__xedav__s__*	*__g__xedav__s__*	*xedav__s__*	*__gv__xedav__s__*	*__g__xedav__t__*	*xedav__s__*
1pl.		*__g__xedav__t__*	*__v__xedav__t__*		*__g__xedav__t__*	*__v__xedav__t__*
2pl.	*__m__xedav__t__*		*xedav__t__*	*__gv__xedav__t__*		*xedav__t__*
3pl.	*__m__xedav__en__*	*__g__xedav__en__*	*xedav__en__*	*__gv__xedav__en__*	*__g__xedav__en__*	*xedav__en__*

To see which forms really require explanation, it is helpful to consider the “what-if” paradigm based on the above, but ignoring the apparent cases of disjunctive application/insertion. All else being equal, and assuming somewhat arbitrarily that subject markers would appear outside of object markers, one would expect the following affixes to appear (Ø stands as a place-holder; predicted but non-appearing affixes are given as capitals):

(3) An idealized paradigm for the present tense of *XEDAV*, ‘see’

Subj. \ DO	1sg.	2sg.	3sg.	1pl.	2pl.	3pl.
1sg.		*__V-g__xedav*†	*__v__xedav*		*__V-g__xedav__t__*†	*__v__xedav-__Ø__*
2sg.	*__Ø-m__xedav*		*__Ø__-xedav*	*__Ø-gv__xedav*		*__Ø__-xedav-__Ø__*
3sg.	*__m__xedav__s__*	*__g__xedav__s__*	*xedav__s__*	*__gv__xedav__s__*	*__g__xedav__t-S__*†	*xedav-__Ø-s__*
1pl.		*__V-g__xedav__t__*†	*__v__xedav__t__*		*__V-g__xedav-__T-t__*†	*__v__xedav__t__*
2pl.	*__Ø-m__xedav__t__*		*__Ø__-xedav__t__*	*__Ø-gv__xedav__t__*		*__Ø__-xedav-__Ø-t__*
3pl.	*__m__xedav__en__*	*__g__xedav__en__*	*xedav__en__*	*__gv__xedav__en__*	*__g__xedav-__T-en__*†	*xedav__en__*

Thus there are six forms (marked † above) out of 28 which are demonstrably not as expected. Every one of the six would otherwise have two consecutive overt prefixes or two consecutive overt suffixes. In the form **V-g-xedav-T-t*, it is obviously questionable which of two consecutive /t/ segments is deleted. Geminates are outlawed generally in Georgian, so the disjunctivity is a moot point in this case. The disjunctivity otherwise, however, is not a matter of phonotactic violations (Anderson 1992:87, fn. 13), but is rather entirely a matter of morphological distribution.

B.2: A-Morphous Morphology's response

Anderson (1984, 1986, 1992) has written extensively about Georgian agreement, and he considers it strong support for the positing of disjunctive rule blocks. Rules which apply disjunctively, by circular definition, belong to the same rule block. The rule which actually applies precedes the others in an ordered block, sequenced by the Elsewhere Condition if one realizes a proper subset of the features contained in the MSR of the others. If this subset relation does not hold, then the appeal is to extrinsic ordering, a brute force preferential application of the rule needed to match the surface facts. In the Georgian case, the 2nd person object prefix preempts the 1st person subject marker where both are applicable, e.g., *g-xedav*, and not *v-xedav*, *v-g-xedav*, or *g-v-xedav*. There is no attempt to motivate a principled precedence relation of *g-* over *v-*.

Anderson (1984) also casts the *–t* suffixes as a unified non-3rd person marker, and claims that the fact that a 1st person plural object is realized by *gv-* but not *–t* is the result of disjunctive ordering and the prefix's precedence in the block. Thus blocks are tied to MSRs, and not to position classes (cf. PFM). Because a subject marker and an object marker are keyed to different MSR layers, according to A-Morphous Morphology, the disjunctivity cannot even be explained by an MSR conflict. A-Morphous Morphology's tolerance of extrinsic ordering within rule blocks seriously compromises the predictive power of the theory in this area.

B.3: Articulated Morphology's response

Since representations in AM are informationally impoverished, a surface form may obtain its morphosyntactic specifications only through the application of rules. This means that for AM, apparent disjunctivity is actually allomorphy of affixes, i.e., in the ordinary case:

(1) 1st person subject

X		*vX*
[ ]	→	[P:1]

but just in case:

(2) 1st person subject

X		*X*
[P:2]	→	[P:1][P:2]

The rule format of AM allows the full details of the input to be part of a rule's domain. There is no limit on access to previously applied rules, in principle (cf. LM&P), and so this sort of broad contextual sensitivity is not a formal problem for AM. Whether this power is theoretically desirable and what its practical constraints are are separate but important issues, however.

B.4: Autolexical Syntax's response

As far as the syntax knows, so to speak, the forms given in the attested Georgian paradigm are fully specified verbs. Any problematic aspects are to be dealt with entirely within the morphological component. Autolexical Syntax is in a worse position to account for disjunctivity of apparently comparable inflectional affixes than AM is, however.

	v-	*g-*	*-xedav-*
Syntax	nil	nil	V[0]
Semantics	nil	nil	F[–2]
Morphology	V[–1]/V[–0]	V[–1]/V[–0]	V[–0]

Since *vxedav* and *gxedav* are both V[-1] forms, i.e., acceptable as fully inflected words, the autolexical specification of the prefixes puts them into competition. This gains the disjunctive application, but it does not explain the precedence of *g*- insertion over *v*-insertion where both are equally motivated in the sentence.

Although word order at the sentence level in Georgian is claimed to be free, the canonical order is S-O-V. Under the assumptions of Autolexical Syntax, since there is a default mapping between abstract syntactic structure and morphological structure, it is possible to derive the needed S-O-V prefix ordering "for free" from the syntax, with no need to appeal to separate linear precedence rules specific to the morphology. Thus it is possible to capture both the paradigmatic and syntagmatic aspects of the Georgian agreement prefix disjunction without extraordinary maneuvers (cf. Singer's (1999) dissertation, which relies heavily on Optimality Theory constraints on top of Autolexical theory in the analysis of every Georgian morphological phenomenon except the present question.)

B.5: Categorial Morphology's response

If both *v*- and *g*- take 'verb stems' as input, and a 'verb stem' as a formal unit crucially has no agreement markers already in place, then the competition between *v*- and *g*- is predicted.

v- </Vstem, V, Pref> *g*- </Vstem, V, Pref>

The dominance of {OB:2} over {SU:1} is not explained, however, and therefore must be stipulated. The analogous analysis is available for the suffixes—each one <Vroot\, Vstem, Suff>. This assumes, somewhat arbitrarily, that agreement suffixes are applied 'before' agreement prefixes. The order could be reversed, *mutatis mutandis*, with no ill effects, it would seem. It is again questionable whether the precedence relation could be captured in a natural way in this framework. (This solution is similar in most regards to that offered by Autolexical Syntax, above.)

B.6: Distributed Morphology's response

In Halle and Marantz (1993:117ff.) the Georgian agreement affixes are explicitly assumed to be clitics, rather than prefixes and suffixes *per se*, and so their morphosyntactic properties are fused into one proclitic, with the possibility of [+pl] fission, allowing the *–t* to be inserted at the right edge of the stem. The *v-*/*g-* issue is handled as a fusional clitic, but the competition of *–t* with *–s* and *-en* is ignored completely. Although the data are given in examples (2e-i, k, l and 4g, h), there is no discussion of *–s* and *–en* in the text, nor are they given as part of the clitic Vocabulary. It seems safe to assume that a more complete analysis would handle all three suffixes as part of the clitic cluster, subject to fission as the *–t* is, but there may come a point when the morphological operations would be fewer if a separate proclitic cluster and enclitic cluster were generated, fused, and then supplied with phonological features.

The clitic analysis allows DM a space apart from the rest of inflection to carry out clearly morphological operations without reference to the host. In a language like Georgian, which has little to no stem allomorphy conditioned by particular affixes, the clitic analysis is not obviously in error. In other languages, however, where more morphophonological operations accompany affixation, a comparable appeal to cliticization might be subject to the possibility of falsification. The fact that the markers do not show the distributional independence of clitics as opposed to affixes is a first indication that the choice of a clitic-based analysis is motivated by theoretical rather than empirical motivations.

B.7: Lexeme-Morpheme Base Morphology's response

Verbs receive their agreement specification in the syntax in LMBM and the abstractly inflected verb is submitted to the Morphological Spelling component for the (incremental) spell-out of features, operating "outward" from the phonological representation of the verb lexeme. The operation of the MS component in LMBM is described as a spelling mechanism which interprets inflectional features individually or in small groups (in the case of fusional exponents), executes the modification of the stem appropriate to the feature (set) in question—informed by the inflectional class of the lexeme at hand—and then immediately erasing the working read-only memory, beginning the process again with the next feature (set) as yet uninterpreted. This mechanism iterates exhaustively, and so disjunction between independently motivated affixes is not immediately predicted. Where *g-* precludes *v-* in a surface form, the most natural analysis is that the spelling mechanism has interpreted both arguments together and has spelled them as the canonical exponent of the object, presumably the first argument encountered in the set of inflectional features. The fact that the putative fused morpheme is phonologically identical to the 2nd person object marker, while not exactly portrayed as an accident, does not follow from anything else in the grammar. Georgian is otherwise quite agglutinative—why this particular formal economizing? This solution describes the fusion without really explaining it.

An alternative view, also permitted within LMBM, is the MS component's ability to selectively 'erase' features it finds in the output of inflection. Inflectional features are ordered by syntactic structure, but no internal sub-bracketing of the features is available to the spelling mechanism. In other words, the MS component can 'see' the full set of features that an inflected lexeme bears, and it can act on those features in groups of up to five at a time (a presumed constraint on working memory). From this perspective, it is certainly possible to conceive of a language specific spelling rule that said: "on encountering both {SU:1st} and {OB:2nd} in the same feature set, erase {SU:1st} (i.e., perform no modification to the stem) and spell {OB:2nd} as usual." The power of such a rule, and of this broad perspective for the spelling mechanism, is open to criticism, of course, but again it points up the ability of a theory which assumes the Separation Hypothesis to allow features to go unexpressed in the phonology, yet still be present in the representation, in a way that 'morpheme-as-sign' theories cannot.

B.8: Lexical Morphology and Phonology's response

The *v*-morpheme can attach to a verb root, but just not to a 2nd person object prefix. This could be taken to suggest that the 2nd person *g*- belongs to a later stratum than the *v*-, and so the presence of the *g*- precludes the further addition of the *v*- from a previous stratum. The question is, if *g*- belongs to a later stratum, why should it get to apply first in the first place?

On the different-strata analysis, we would be forced to say that *g*- <u>has</u> to apply first for some language-specific reason, perhaps that object markers must be inserted first. In this case, then, *g*- would apply and preclude the insertion of *v*-. This is, of course, begging the question.

An appeal to a template **subj | obj | root ...** is really just another way of saying the same thing, that moving out from the root, the object marker is inserted closer to the stem. We still have to assume the stratum explanation to get disjunction rather than simple S-O-V patterning in the morphology (cf. Autolexical Syntax's response, above). The weakening assumption of a 'loop' sometimes invoked in LM&P can be avoided in this account, however.

Since there is no apparent morphophonology to account for in these data, it is questionable whether any independent justification would be forthcoming for the different-strata account.

B.9: Natural Morphology's response:

It seems that there is a certain irreducible amount of un-sign-like behavior in the Georgian facts, whether the analysis involves zero-morphemes or syncretism between unitary and fused morphemes. Zero-marking of 3rd person is not exceptional on general markedness grounds, but for 1st person, this is less expected. The approach of Mayerthaler (1988:8ff.), however, allows for a more sophisticated picture of markedness calculation. Typical attributes of the speaker are to be taken as background in a discourse context, not requiring especially salient marking in contrast with non-speaker attributes, which are to be interpreted as 'figures' in the foreground.

With these two perspectives in place, then, the motivation for maintaining a marker for 2nd person even at the expense of a 1st or 3rd person marker is clear. Second person is more marked than 3rd, since there are indefinitely many 3rd person referents available in any given discourse situation. Second person is more marked than 1st, as well, since 2nd person is part of the non-speaker class. First and 3rd person do not conflict in the Georgian system, so no further hierarchical relationship is determinable. With an explanation for the dominance of 2nd person over 1st in hand, however, the motivation for pre-emption rather than closer linear proximity to the stem (i.e., *gxedav* rather than **vgxedav*) is not evident.

Natural Morphology, therefore, would seem to have the piece of the puzzle that more formalist theories are forced to simply stipulate. It is not, however, in a position to account for the disjunction, which would seem to be a purely formal matter (a perhaps arbitrarily limited amount of available 'real estate' in which agreement markers can appear).

B.10: "Network Model"'s response

The idea of a *schema* in the Network Model is thought to represent the connections which exist between words in the mental lexicon. Schemas can be defined phonologically, morphologically, syntactically, semantically, and in other ways as well. Prototypes or "best exemplars" are thought, therefore, to serve as an organizing principle for lexical categories. Based on the observed Georgian forms in (2), one could posit an abstract schema for inflected verbs:

Pref – Stem – Suff

This schema is instantiated in the inflection of transitive and intransitive verbs in Georgian, since some arguments are realized partially by a prefix and partially by a suffix, e.g., {SU:1st-pl} corresponds to a *v-* prefix and a *–t* suffix, all on its own. It could be assumed therefore, that the simple schema is used as a guide in determining "acceptable Georgian verbs," and hypothetical verbs which bear more than one agreement prefix or more than one agreement suffix will be judged as unacceptable. While this may work as a synchronic generalization, of course, it offers no insight into how such an arbitrary limitation could come into being. Iconicity would predict at least one marker per argument, but this does not always happen, e.g., *g-xedav* 'I see you'.

At the same time, patterns as regular and productive as this, i.e., agreement marking on verbs of all sorts, are predicted (albeit with some reservation in Bybee (1988)) to have a degree of independence from the words which instantiate them, since they are used freely in neologistic formations, etc. The extremely high frequency of the marking system, however, may yet have explanatory value if it is recalled that frequent forms more readily sustain idiosyncrasies, whereas rarer forms are subject to regularization (i.e., replacing zero-expression with something overt). There is no competing system of markers in Georgian, however, so perhaps at present there is no viable regularizing pressure outside of the observed pattern, and thus it is firmly and indefinitely entrenched, despite its (regular) quirks.

B.11: Network Morphology's response

In order to get the paradigmatic facts right, it seems that all subject-object combinations would have to be treated as units, sometimes realized as two bound morphs, sometimes one (and in the case of reflexives, partially periphrastically). When a combination of arguments is realized by a single morph, the marker would be a *de facto* fusional morpheme.

```
<mor su 1 ob 2> == g_ "<stem>"
```

An unfortunate side-effect of this analysis is that the homophony of the putative fusional morphemes with the canonical exponents of one of the arguments so combined is portrayed as an accident. The possibility of casting the above fact (in the technical sense) as a rule of referral is but a slight improvement:

```
<mor su 1 ob 2> == "<mor ob 2>"
```

for there is no *a priori* reason to expect that any correspondence would exist, let alone such a close one, between a fusional and a non-fusional exponent.

The rules of inference in the DATR format state facts about the realization of morphosyntactic features without reference to the broader context of other rules applying in a given form. There is no ready way to capture disjunctivity here without a notion of competition between applicable rules. Since Network Morphology does not formalize a notion of slots or position classes, there is no natural way of inducing competition, or of predicting a "winner," should such competition occur.

B.12: Paradigm Function Morphology's response

Georgian agreement is taken up in Stump (2001: ch. 3), partly in response to Anderson's (1986, 1992) analysis, and in particular because of the challenge the facts pose for the *Pāṇinian Determinism Hypothesis*: the assumption that for a given rule block, the narrowest applicable realization rule is always uniquely identifiable, and this rule applies to the exclusion of all competitors.

There are four agreement prefixes, *v*-, *m*-, *gv*-, and *g*-, and the realization rules introducing these markers are the following, respectively:

a. $\mathrm{RR}_{\text{pref, \{A\scriptsize GR\normalsize (su):\{P\scriptsize ER\normalsize :1\}\}, [V]}}(\langle X, \sigma\rangle) =_{\text{def}} \langle \mathit{v}X', \sigma\rangle$

b. $\mathrm{RR}_{\text{pref, \{AGR(ob):\{PER:1\}\}, [V]}}(\langle X, \sigma\rangle) =_{\text{def}} \langle \mathit{m}X', \sigma\rangle$

c. $\mathrm{RR}_{\text{pref, \{AGR(ob):\{PER:1, NUM:pl\}\}, [V]}}(\langle X, \sigma\rangle) =_{\text{def}} \langle \mathit{gv}X', \sigma\rangle$

d. $\mathrm{RR}_{\text{pref, \{AGR(ob):\{PER:2\}\}, [V]}}(\langle X, \sigma\rangle) =_{\text{def}} \langle \mathit{g}X', \sigma\rangle$

These rules embody several assumptions:

(1) they are all introduced by a single rule block ($\mathrm{RR}_{\text{pref}}$ is not qualitatively different from RR_2, or any arbitrarily indexed rule block), and so constitute a position class;

(2) there are distinct morphosyntactic features for subject and object agreement, identified diacritically, rather than structurally (cf. A-Morphous Morphology's layered MSRs); and

(3) {AGR} features are set-valued features, rather than atomic-valued features (i.e., {AGR} takes feature-value pairs as its value).

For evaluation purposes, it is important to compare the rules for narrowness and applicability. Rules (b), (c), and (d) are paradigmatically related, and so cannot co-occur for practical reasons. Rules (a), (b), and (c) all realize {PER:1}, and so, should they co-occur in a particular context, would be realized periphrastically as a reflexive construction, according to Georgian grammar. The only possible competition scenario, therefore, is between rules (a) and (d), the *v-/g-* conflict exactly.

Rules (a) and (d) are apparently equally narrow, and both are applicable in extensions of {AGR(su):{PER:1}, AGR(ob):{PER:2}}. Stump's response to this is to posit two modes of rule application, *expanded* and *unexpanded*. Rules generally apply in unexpanded mode, realizing a particular morphosyntactic property set. Certain rules, however, are defined as applying in expanded mode, "realizing EVERY WELL-FORMED EXTENSION of a particular property set" (Stump 2001:72). Rules applying in expanded mode are actually rule schemata, instantiated by each member of a class of rules applying in unexpanded mode. In the present case, the dominance and categorical applicability of rule (d) is assumed to be evidence of expanded application:

d'. $RR_{\text{pref}, \leftarrow\{\text{AGR(ob)}:\{\text{PER}:2\}\}\rightarrow, [\text{V}]}(\langle X, \sigma\rangle) =_{\text{def}} \langle gX', \sigma\rangle$

The arrows surrounding the second subscripted rule-index are the formal means of indicating expanded application. The effect of (d') in Georgian will be to realize every well-formed extension of {AGR(ob):{PER:2}} with the *g-* prefix in the prefix slot, i.e., every inflected form which realizes a 2nd person object will have a *g-* prefix, and never any other agreement prefix.

This approach is more restrictive than a theory which allows for the possibility of fully extrinsic rule ordering, because the constitution of rule blocks in PFM is fundamentally tied to distribution and position classes, whereas A-Morphous Morphology, for example, permits exponents which are realized in linearly distant positions to be part of the same rule block. PFM insists on localized competition. The rule schema approach also predicts that a schema cannot be preempted by another rule applying in expanded mode by definition. Schemata are therefore constrained, and can only be invoked where the "every well-formed extension" criterion is met.

The suffix *–t* which realizes only—but not all—extensions of {NUM:pl}, by contrast, cannot be handled with an expansion schema:

e. Where $\alpha \neq 3$,
$RR_{\text{suff}, \{\text{AGR(su)}:\{\text{PER}:\alpha, \text{NUM:pl}\}\}, [\text{V}]}(\langle X, \sigma\rangle) =_{\text{def}} \langle X't, \sigma\rangle$

f. $RR_{\text{suff}, \{\text{AGR(ob)}:\{\text{PER}:2, \text{NUM:pl}\}\}, [\text{V}]}(\langle X, \sigma\rangle) =_{\text{def}} \langle X't, \sigma\rangle$

g. $RR_{\text{suff}, \{\text{AGR(su)}:\{\text{PER}:3, \text{NUM:pl}\}\}, [\text{V}]}(\langle X, \sigma\rangle) =_{\text{def}} \langle X'en, \sigma\rangle$

h. $RR_{\text{suff}, \{\text{AGR(su)}:\{\text{PER}:3\}\}, [\text{V}]}(\langle X, \sigma\rangle) =_{\text{def}} \langle X's, \sigma\rangle$

Again, to evaluate narrowness and applicability, rules (e), (g), and (h) are in a paradigmatic relation and cannot conflict. Conflicts between (e) and (f) are resolved in two ways: where $\alpha = 1$, both subject and object are realized by a *–t* suffix, and so the resolution is vacuous (degemination or no); where $\alpha = 2$, the combination entails a periphrastic reflexive construction. The remaining conflicts are between (f) and each of (g) and (h): (f) trumps (h) by narrowness, but this is not so for the relation between (f) and (g), which are apparently equally narrow. The effect of (g) is never overridden, and so the criterion of "every well-formed extension" would seem to be met. Recasting (g) as an expansion schema:

g'. $\mathrm{RR}_{\mathrm{suff},\ \leftarrow\{\mathrm{A{\scriptsize GR}(su)}:\{\mathrm{P{\scriptsize ER}}:3,\ \mathrm{N{\scriptsize UM}}:\mathrm{pl}\}\}\rightarrow,\ [\mathrm{V}]}(\langle X, \sigma\rangle) =_{\mathrm{def}} \langle X'\mathit{en}, \sigma\rangle$

This analysis predicts very simply that every extension of {AGR(su):{PER:3, NUM:pl}} will show the *–en* suffix, and this is indeed the case. The phonetic realization of (e) and (f) are distinct in some dialects, and so the splitting of the *–t* suffixes is diachronically and dialectologically supported, and not simply a theoretical expedient.

The expansion schema as a theoretical construct preserves the Pāṇinian Determinism Hypothesis, but there is no formally based explanation for why only those rules defined as expansion schemata are so defined. That part of the explanation may well be extralinguistic.

Stewart (2001) focuses attention on lexeme class, rather than property set, as an alternative in the evaluation of the relative narrowness of competing rules in a block. Since the set of verb lexemes that may have two arguments is a subset of the set of verbs, a rule that is defined as applying to two-argument verbs is narrower than one defined as applying to the category of verbs as a whole. Thus it is the domain of applicability, not the range of properties realized, that is decisive in the case of Georgian, and the Pāṇinian Determinism Hypothesis may be upheld without an appeal to a second mode of application after all.

B.13: Prosodic Morphology's response

The special devices of prosodic morphology are actually superfluous here, since there is no copying or nonconcatenative operation to perform. Since there is no phonological motivation for the *g-*/*v-* disjunction, there is nothing to say here that would be different from what Word Syntax (see below) could offer.

B.14: Word Syntax's response

Overt and phonetically null affixes are no problem for Word Syntax, so long as the surface form is consistent, e.g., 2nd person singular subjects are consistently marked with a null prefix, 2nd person objects by (at least) a *g-* prefix. Since words are built up exhaustively from morphemes in Word Syntax, every element of content (grammatical or lexical/derivational) associated with the inflected word is attributable to at least one constituent morpheme. This entails that in cases like [1pl-subj, 2pl-obj] *gxedavt* there is some element in the morphological structure which contributes the [1pl-subj] specification. Since there is no such readily identifiable overt element, the simplest answer is to posit a phonetically null allomorph of the *v*- prefix (the –*t* suffix(es) may be a moot issue, given the language's general anti-gemination constraint). The problem is that the distribution of this allomorph is suspiciously specific, i.e., it "appears" just in case the *g*- prefix is also called for.

A language-specific filter could also be posited, reducing the morpheme sequence *v*-*g*- to *g*-. This filter could hold as a surface condition or as a prohibited structure, but either way the move is a mere stipulation with no explanatory value. Note further that both [gv] and [vg] are attested word initially in Georgian.

A third option is to claim that (as was proposed for Network Morphology, above) a class of fusional morphemes is used, introducing specifications for subject and object at the same time. This would mean that the markers observed in table (2) in the original data set are not combinations of two independent morphemes, but rather are affixes and circumfixes which fuse multiple argument specifications. In fact, every one of the markers in (2) is homophonous with an otherwise existing agreement marker appropriate to one of the arguments in question. Needless to say, this is not very satisfying, and verges on the willful omission of a generalization.

References

Akinlabi, Akinbiyi. 1996. Featural affixation. *Journal of Linguistics* 32.239-89.

Allen, Margaret R. 1978. *Morphological investigations*. University of Connecticut dissertation.

Anderson, Stephen R. 1969. *West Scandinavian Vowel Systems and the Ordering of Phonological Rules*. MIT dissertation.

Anderson, Stephen R. 1977. On the formal description of inflection. *Chicago Linguistic Society* 13.15-44.

Anderson, Stephen R. 1982. Where's morphology? *Linguistic Inquiry* 13(4).571-612.

Anderson, Stephen R. 1984. On representations in morphology: Case, agreement, and inversion in Georgian. *Natural Language and Linguistic Theory* 2.157-218.

Anderson, Stephen R. 1986. Disjunctive ordering in inflectional morphology. *Natural Language and Linguistic Theory* 4.1-31.

Anderson, Stephen R. 1988a. Inflection. *Theoretical Morphology*, ed. by Michael Hammond and Michael Noonan, 23-43. San Diego: Academic Press.

Anderson, Stephen R. 1988b. Morphological theory. *Linguistics: The Cambridge Survey* [vol. 1], ed. by Frederick J. Newmeyer, 146-91. Cambridge: Cambridge University Press.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Archangeli, Diana. 1987. An overview of the theory of Lexical Phonology and Morphology. *MIT Working Papers in Linguistics* 7.1-14.

Aronoff, Mark 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Aronoff, Mark 1994. *Morphology by itself*. Cambridge, MA: MIT Press.

Bach, Emmon. 1984. Some Generalizations of Categorial Grammars. *Varieties of Formal Semantics*, ed. by Fred Landman and Frank Veltman, 1-24. Dordrecht: Foris.

Beard, Robert. 1986. Neurological evidence for Lexeme–Morpheme Based Morphology. *Acta Linguistica Academiae Scientarium Hungarica* 36.1-22.

Beard, Robert. 1987. Morpheme order in a Lexeme/Morpheme-Based Morphology. *Lingua* 72.1-44.

Beard, Robert. 1988. On the separation of derivation from morphology: Toward a Lexeme/Morpheme-Based Morphology. *Quaderni di Semantica* 9(1).3-59.

Beard, Robert. 1995. *Lexeme–Morpheme Base Morphology*. Albany, NY: SUNY Press.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Booij, Geert. 1993. Against split morphology. *Yearbook of Morphology 1993*, ed. by Geert Booij and Jaap van Marle, 27-49. Dordrecht: Kluwer.

Booij, Geert, and Jerzy Rubach. 1985. Postcyclic versus postlexical rules in Lexical Phonology. *Vrije Universiteit* [Amsterdam] *Working Papers in Linguistics* 14.

Brown, Dunstan. 1998. Stem indexing and morphonological selection in the Russian verb: A Network Morphology account. *Models of Inflection* [Linguistische Arbeiten 388], ed. by Ray Fabri, Albert Ortmann, and Teresa Parodi, 196-224. Tübingen: Max Niemeyer Verlag.

Brown, Dunstan, Greville Corbett, Norman Fraser, Andrew Hippisley, and Alan Timberlake. 1996. Russian noun stress and Network Morphology. *Linguistics* 34.53-107.

Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*, [Typological Studies in Language, vol. 9]. Amsterdam/Philadelphia: John Benjamins.

Bybee, Joan L. 1988. Morphology as lexical organization. *Theoretical Morphology*, ed. by Michael Hammond and Michael Noonan, 119-41. San Diego: Academic Press.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5).425-55.Bybee, Joan L., and Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language* 59(2).251-70.

Carstairs, Andrew. 1983. Paradigm economy. *Journal of Linguistics* 19.115-25.

Carstairs. Andrew. 1987. *Allomorphy in inflexion*. London: Croom Helm.

Cherchi, Marcello. 1999. *Georgian* [Languages of the World: Materials, 147]. München: Lincom-Europa.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press,

Chomsky, Noam. 1970. Remarks on nominalization. *Readings in transformational grammar*, ed. by R. Jacobs and P. Rosenbaum, 184-221. Waltham, MA: Ginn and Co.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Corbett, Greville G. 1982. Gender in Russian: An account of gender specification and its relationship to declension. *Russian Linguistics* 6.197-232.

Corbett, Greville G., and Norman M. Fraser. 1993. Network Morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics* 29.113-42.

Di Sciullo, Anna Maria, and Edwin Williams. 1987. *On the Definition of Word* [Linguistic Inquiry Monograph 14]. Cambridge, MA: MIT Press.

Donegan, Patricia, and David Stampe. 1977. On the description of phonological hierarchies. *Chicago Linguistic Society* 13.35-38.

Donegan, Patricia J., and David Stampe. 1979. The study of Natural Phonology. *Current Approaches to Phonological Theory*, ed. by Daniel A. Dinnsen, 126-73. Bloomington: Indiana University Press.

Dowty, David R. 1979. *Word meaning and Montague semantics*. Dodrecht: Reidel.

Dressler, Wolfgang U. 1985a. Typological aspects of natural morphology. *Acta Linguistica Academiae Scientarium Hungarica* 35.51-70.

Dressler, Wolfgang U. 1985b. *Morphonology: The Dynamics of Derivation* [Linguistica Extranea, Studia 12]. Ann Arbor, MI: Karoma.

Dressler, Wolfgang U. 1985c. On the predictiveness of Natural Morphology. *Journal of Linguistics* 21.321-37.

Dressler, Wolfgang U. (ed.), Willi Mayerthaler, Oswald Panagl, and Wolfgang U. Wurzel. 1987. *Leitmotifs in Natural Morphology* [Studies in Language Companion Series, vol. 10]. Amsterdam/Philadelphia: John Benjamins.

Embick, David, and Rolf Noyer. 2007. Distributed Morphology and the syntax-morphology interface. *The Oxford Handbook of Linguistic Interfaces*, ed. by Gillian Ramchand and Charles Reiss, 289-324. Oxford/New York: Oxford University Press.

Evans, Roger and Gazdar, Gerald. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics* 22.167-216.

Fabb, Nigel. 1988. Doing affixation in the GB syntax. *Morphology and Modularity* [Publications in Language Sciences 30], ed. by Martin Everaert, Arnold Evers, Riny Huybregts, and Mieke Trommelen, 129-45. Dordrecht: Foris.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fraser, Norman M., and Greville G. Corbett. 1995. Gender, animacy, and declensional class assignment: A unified account for Russian. *Yearbook of Morphology 1994*, ed. by Geert Booij and Jaap van Marle, 123-50. Dordrecht: Kluwer.

Gazdar, G., Klein, E., Pullum, G., and Sag, I. 1985. *Generalized phrase structure grammar*. Oxford: Blackwell.

Goldsmith, John. 1976. An Overview of Autosegmental Phonology. *Linguistic Analysis* 2.23-68.

Halle, Morris. 1990. An approach to morphology. *Proceedings of the 20th annual meeting of the Northern-Eastern Linguistic Society*, 150-84.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. *The View from Building 20*, ed. by Kenneth Hale and Samuel Jay Keyser, 111-76. Cambridge, MA: MIT Press.

Halle, Morris and K. P. Mohanan. 1985. Segmental phonology of Modern English. *Linguistic Inquiry* 16(1).57-116.

Halle, Morris, and Jean-Roger Vergnaud. 1980. Three Dimensional Phonology. *Journal of Linguistic Research* 1(1).83-105.

Harris, Alice. 1981. *Georgian Syntax: A Study in Relational Grammar*. Cambridge: Cambridge University Press.

Harris, Alice. 1984. Inversion as a rule of Universal Grammar: Georgian evidence. *Studies in Relational Grammar 2*, ed. by David M. Perlmutter and Carol Rosen, 259-91. Chicago: University of Chicago Press.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10.210-34.

Hoeksema, Jack. 1985. *Categorial Morphology*. New York: Garland. [1984. Rijksuniversiteit Groningen dissertation].

Hoeksema, Jack, and Richard D. Janda. 1988. Implications of process-morphology for Categorial Grammar. *Categorial Grammars and Natural Language Structures*, ed. by Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler, 199-247. Dordrecht: Reidel.

Hooper, Joan B. 1976. *An Introduction to Natural Generative Phonology*. New York: Academic Press.

Hooper, Joan B. 1979. Substantive principles in Natural Generative Phonology. *Current Approaches to Phonological Theory*, ed. by Daniel A. Dinnsen, 106-25. Bloomington: Indiana University Press.

http://www.facstaff.bucknell.edu/rbeard

http://www.ling.upenn.edu/~rnoyer/dm/

http://www.surrey.ac.uk/LIS/SMG/surrey_morphology_group.htm

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51(3).639-71.

Janda, Richard D. 1986. *On the Motivation for an Evolutionary Typology of Sound-structural Rules*. UCLA dissertation.

Janda, Richard D. (n.d.) *On Disjunctivity in Morphology and Phonology*. Unpublished book ms.

Janda, Richard D., and Brian D. Joseph. 1986. One rule or many? Sanskrit reduplication as fragmented affixation. *ESCOL '85*.103-19.

Janda, Richard D., and Brian D. Joseph. 1992a. Meta-templates and the underlying (dis-)unity of Sanskrit reduplication. *ESCOL '91*.160-73.

Janda, Richard D., and Brian D. Joseph. 1992b. Pseudo-agglutinativity in Modern Greek verb-inflection and 'Elsewhere'. *Chicago Linguistic Society* 28.251-66.

Janda, Richard D., and Brian D. Joseph. 1999. The Modern Greek negator *μη*(*ν*)(-) as a Morphological Constellation. *Proceedings of the 3rd International Conference on Greek Linguistics*, ed. by Georgios Babiniótes and Amalia Mozer, 341-51.

Janda, Richard D., and María Sandoval. 1984. 'Elsewhere' in morphology. Bloomington, IN: Indiana University Linguistics Club.

Julien, Marit. 2007. On the relation between morphology and syntax. *The Oxford Handbook of Linguistic Interfaces*, ed. by Gillian Ramchand and Charles Reiss, 209-38. Oxford/New York: Oxford University Press.

Kathol, Andreas. 1999. Agreement and the syntax-morphology interface in HPSG. *Studies in Contemporary Phrase Structure Grammar*, ed. by Robert D. Levine and Georgia M. Green, 223-74. Cambridge: Cambridge University Press.

Kiparsky, Paul. 1973. 'Elsewhere' in phonology. *A Festschrift for Morris Halle*, ed. by Stephen R. Anderson and Paul Kiparsky, 93-106. New York: Holt, Rinehart, and Winston.

Kiparsky, Paul. 1982a. Lexical Morphology and Phonology. *Linguistics in the Morning Calm*, ed. by Linguistic Society of Korea, 3-90. Seoul: Hanshin.

Kiparsky, Paul. 1982b. From Cyclic to Lexical Phonology. *The Structure of Phonological Representations* [part 1], ed. by Harry van der Hulst and Norval Smith, 131-75. Dordrecht: Foris.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2.85-138.

Lamb, Sydney M., and Leonard E. Newell. 1966. *Outline of Stratificational Grammar*. Washington, DC: Georgetown University Press.

Lieber, Rochelle. 1981. *On the Organization of the Lexicon*. Bloomington, IN: Indiana University Linguistics Club. [1980. MIT dissertation].

Lieber, Rochelle. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory*. Chicago: University of Chicago Press.

McCarthy, John J. 1979. *Formal Problems in Semitic Phonology and Morphlogy*. MIT dissertation.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12(3).373-418.

McCarthy, John J. 1982. Prosodic templates, morphemic templates, and morphemic tiers. *The Structure of Phonological Representations* [part 1], ed. by Harry van der Hulst and Norval Smith, 191-223. Dordrecht: Foris.

McCarthy, John J., and Alan Prince. 1993a. Generalized alignment. *Yearbook of morphology 1993*, ed. by Geert Booij and Jaap van Marle, 79-153. Dordrecht: Kluwer.

McCarthy, John J., and Alan Prince. 1993b. Prosodic morphology I: constraint interaction and satisfaction. University of Massachusetts, Amherst, and Rutgers University ms.

McCarthy, John J., and Alan Prince. 1994a. The emergence of the unmarked: Optimality in prosodic morphology. *Proceedings of the 24th annual meeting of the Northern-Eastern Linguistic Society*, ed. by Merce Gonzales, 333-79.

McCarthy, John J., and Alan Prince. 1994b. Prosodic morphology: an overview. Paper presented at the Utrecht Workshop on Prosodic Morphology.

Marantz, Alec. 1982. Re reduplication. *Linguistic Inquiry* 13(3).435-82.

Matthews, P. H. 1972. *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.

Mayerthaler, Willi. 1981/1988. *Morphologische Natürlichkeit/ Morphological Naturalness* [Linguistica Extranea, Studia 17]. Ann Arbor, MI: Karoma.

Mohanan, K. P. 1986. *The Theory of Lexical Phonology*. Dordrecht: Reidel.

Montague, Richard. 1970. Universal grammar. *Theoria* 36.373-98.

Montague, Richard. 1973. The proper treatment of quantification in ordinary English. *Approaches to Natural Language*, ed. by J. Hintikka, J. Moravcsik, and P. Suppes, 222-42. Dordrecht: Reidel.

Peirce, Charles S. 1965. *Collected Papers* [ed. by C. Hartshorne and P. Weiss]. Cambridge, MA: Harvard University Press.

Perlmutter, David M. 1988. The Split Morphology Hypothesis: Evidence from Yiddish. *Theoretical Morphology*, ed. by Michael Hammond and Michael Noonan, 79-100. San Diego: Academic Press.

Pollard, Carl J., and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Prince, Alan, and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University.

Pullum, Geoffrey K., and Arnold M. Zwicky. 1991. A misconceived approach to morphology. *Proceedings of the West Coast Conference on Formal Linguistics*, ed. by Dawn Bates, 387-398.

Pyatt, Elizabeth J. 1997. *An Integrated Model of the Syntax and Phonology of Celtic Mutation.* Harvard University dissertation.

Riehemann, Susanne. 1998. Type-based derivational morphology. *Journal of Comparative Germanic Linguistics* 2.49-77.

Sadock, Jerrold. 1985. Autolexical syntax: A theory of noun incorporation and similar phenomena. *Natural Language and Linguistic Theory* 3.379-440.

Sadock, Jerrold. 1988. The Autolexical classification of lexemes. *Theoretical Morphology*, ed. by Michael Hammond and Michael Noonan, 271-90. San Diego: Academic Press.

Sadock, Jerrold. 1991. *Autolexical Syntax*. Chicago: University of Chicago Press.

Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace, and World.

Schmerling, Susan F. 1983. Montague morphophonemics. *Chicago Linguistic Society* 19(P).222-37.

Selkirk, Elisabeth O. 1982. *The Syntax of Words* [Linguistic Inquiry Monograph 7]. Cambridge, MA: MIT Press.

Siegel, Dorothy L. 1979. *Topics in English Morphology*. New York: Garland.

Singer, Kora. 1999. *Agreement: Cross-Linguistic Variation and Acquisition*. University of Chicago dissertation.

Spencer, Andrew. 1993. The morphology-syntax interface. [Review of Sadock 1991]. *Journal of Linguistics* 29.143-56.

Spencer, Andrew. 2004. Generalized Paradigm Function Morphology. URL: http://privatewww.essex.ac.uk/ ~spena/papers/GPFM.pdf

Steele, Susan. 1990. Pass it on! A combinatorics-based approach to feature-passing. *Chicago Linguistic Society* 26.407-21.

Steele, Susan. 1995. Towards a theory of morphological information. *Language* 71(2).260-309.

Stewart, Jr., Thomas W. 2000. Apparent congruence and elusive bootstraps: Assessing systemic similarity in language contact. Poster presented at the 9th International Morphology Meeting, Vienna.

Stewart, Jr., Thomas W. 2001. "Georgian verb agreemant without extrinsic ordering." *OSU Working Papers in Linguistics* 56.107-33.

Stewart, Jr., Thomas W. 2004. *Mutation as Morphology: Bases, Stems, and Shapes in Scottish Gaelic*. The Ohio State University dissertation.

Stewart, Jr., Thomas W., and Gregory T. Stump. 2007. Paradigm Function Morphology and the morphology-syntax interface. *The Oxford Handbook of Linguistic Interfaces*, ed. by Gillian Ramchand and Charles Reiss, 383-421. Oxford/New York: Oxford University Press.

Stump, Gregory T. 1990. Breton inflection and the Split Morphology Hypothesis. *The Syntax of the Modern Celtic Languages* [Syntax and Semantics 23], ed. by Randall Hendrick, 97-119. San Diego: Academic Press.

Stump, Gregory T. 1991. A paradigm-based theory of morphosemantic mismatches. *Language* 67(4).675-725.

Stump, Gregory T. 1992. On the theoretical status of position class restrictions on inflectional affixes. *Yearbook of Morphology 1991*, ed. by Geert Booij and Jaap van Marle, 211-41. Dordrecht: Kluwer.

Stump, Gregory T. 1993a. On rules of referral. *Language* 69(3).449-79.

Stump, Gregory T. 1993b. Position classes and morphological theory. *Yearbook of Morphology 1992*, ed. by Geert Booij and Jaap van Marle, 129-80. Dordrecht: Kluwer.

Stump, Gregory T. 1993c. Reconstituting morphology: The case of Bantu preprefixation. *Linguistic Analysis* 23(3-4).169-204.

Stump, Gregory T. 1995. The uniformity of head marking in inflectional morphology. *Yearbook of Morphology 1994*, ed. by Geert Booij and Jaap van Marle, 245-96. Dordrecht: Kluwer.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure* [Cambridge Studies in Linguistics 93]. Cambridge: Cambridge University Press.

Toman, Jindřich. 1998. Word Syntax. *Handbook of Morphology*, ed. by Andrew Spencer and Arnold M. Zwicky, 306-21. Oxford/Malden, MA: Blackwell.

Wheeler, Max W. 1993. On the hierarchy of naturalness principles in inflectional morphology. *Journal of Linguistics* 29(1).95-111.

Williams, Edwin. 1981. On the notions "lexically related" and "head of a word". *Linguistic Inquiry*, 12(2).245-74.

Wurzel, Wolfgang Ullrich. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer.

Zwicky, Arnold M. 1985. How to describe inflection. *Berkeley Linguistics Society* 11.372-86.

Zwicky, Arnold M. 1986. The general case: Basic form versus default form. *Berkeley Linguistics Society* 12.305-14.

Zwicky, Arnold M. 1987a. Morphological rules, operations, and operation types. *ESCOL '87*.318-34.

Zwicky, Arnold M. 1987b. Phonological and morphological rule interactions in highly modular grammars. *ESCOL '87*.523-32.

Zwicky, Arnold M. 1992. Some choices in the theory of morphology. *Formal Grammar: Theory and Implementation*, ed. by Robert D. Levine, 327-71. New York: Oxford University Press.