PROCEEDINGS OF THE FIRST MEDITERRANEAN CONFERENCE OF MORPHOLOGY

(Mytilene, Greece, Sept. 19-21 1997)

Edited by:

GEERT BOOIJ
Free University of Amsterdam

ANGELA RALLI
University of Patras

SERGIO SCALISE
University of Bologna

UNIVERSITY OF PATRAS

PROCEEDINGS OF THE FIRST MEDITERRANEAN CONFERENCE OF MORPHOLOGY

(Mytilene, Greece, Sept. 19-21 1997)

Edited by

GEERT BOOIJ
Free University of Amsterdam

ANGELA RALLI
University of Patras

SERGIO SCALISE
University of Bologna

UNIVERSITY OF PATRAS

PREFACE

It is increasingly clear that progress in the theory of grammar greatly depends on progress in morphology. Under this perspective, a number of crucial issues like allomorphy, compounding and inflection fall equally well in more than one grammatical components, namely, morphology, syntax and phonology, and have assumed the status of "hot topics" in the field of grammatical theory.

The aim of the First Mediterranean Conference of Morphology, held in Mytilene (island of Lesbos, Greece) in September 19-21, 1997, was to address these issues and bring together experts working on morphology in a variety of theoretical approaches. The conference was received very enthusiastically by morphologists all over the world and was attended by 84 linguists from 20 different countries. 15 papers were selected for presentation in three sessions referring to "Allomorphy", "Compounding" and "Inflection". 9 more papers were chosen as alternates to these 15, and 3 talks were given by invited speakers.

This volume brings together the majority of contributions at the conference. The papers are organized thematically into four sections. Section I contains the papers of the three invited speakers, M. Aronoff, A. Spencer and A. Anastassiadis-Symeonidis. Section II covers allomorphy; Section III consists of papers on compounding and Section IV presents contributions on inflection.

We would like to express our deep gratitude to the following sponsors of the conference: The Municipality of Mytilene, the Prefecture of Lesbos and the Ministry of the Aegean.

We are also grateful to the University of Patras, particularly to its Rector Prof. Stamatis Alahiotis, whose generous and most valuable support made the publicaton of this volume possible.

Finally, we would like to thank Anna Thornton and Takis Hadjipanayotis, for their precious help to the preparation of the volume, as well as all the contributors for their understanding and willingness to collaborate with the editorial requirements of this publication.

The Editors,

Geert Booij, *University of Amsterdam*
Angela Ralli, *University of Patras*
Sergio Scalise, *University of Bologna*

October 1998

CONTENTS

Section I
Plenary Papers

Mark Aronoff, SUNY Stony Brook

GENDER AGREEMENT AS MORPHOLOGY*

'Gender is the most puzzling of the grammatical categories.'
Greville Corbett, *Gender* (1991), p.1.

ABSTRACT

Gender, conceived narrowly as agreement class, is morphosyntactic. While the question of what can agree with what is purely syntactic and universal, the gender categories or features that the actual agreement markers instantiate are not, and the phonological form that agreement takes is purely morphological. Language-particular arbitrary and irregular gender interact with universal agreement through defaults. The final universal default type of agreement morphology is phonological copying, which is a reflection of the underlying universal syntactic mechanism of agreement: complete copying of the controller onto the target. This complete copying is normally disguised by the morphology of the language through its arbitrary system of genders. Actual phonological copying only emerges when it is forced to surface under unusual circumstances, through the rare phenomenon of alliterative agreement, when no gender is otherwise available. The interaction of language-particular and irregular gender systems with the universal mechanism of agreement thus sheds light on the nature of the agreement mechanism.

1. Introduction

I would like to explore the interaction between gender, conceived narrowly as agreement class, and agreement. I have argued elsewhere in some detail (Aronoff 1994) that gender is a morphosyntactic bridge between morphology and syntax: while the question of what can agree with what is purely syntactic, the gender categories or features that the actual agreement markers instantiate are not, and the phonological form that agreement takes is purely morphological. As I have shown in this previous work, although genders must be distinguished from inflectional classes, gender systems are always realized through purely morphological inflectional classes. Morphology, being a kind of phonological realization (Zwicky 1992), necessarily involves Saussurean arbitrariness in the mapping between form and meaning. Indeed, were it not for morphology, arbitrariness might be confined to individual lexical items. It is the fact that morphology intervenes between syntax and phonology that makes languages arbitrarily systematic. Because gender is always routed

* In this presentation, I rely heavily on the work of Lise Dobrin, with whom I have discussed this and related problems for several years, and Fiona McLaughlin, whose work I have only recently discovered. Thanks to Kinyalolo Kasangati for discussion. Thanks also to the organizers of the Mitilini conference for getting me to put this down on paper and to all the participants at the conference for a truly splendid few days.

through morphological categories and their realization, it follows that some element of arbitrariness will also figure in the realization of gender. To take a trivial example, there is no reason why the feminine gender in Spanish is normally realized through the *-a* class and the masculine through the *-o* class. In Hindi, masculine is realized through *-a* and feminine through *-i*, showing that the realization of a gender through any particular form is arbitrary. But gender itself, not merely its morphological realization, is arbitrary. The categorizations that gender systems follow vary quite widely across languages, no matter how regular they may be (though they are not often very regular), unlike those for its sister categories of person and number or even case. Some gender systems are sex-based, some shape-based, some rooted in animacy, and some based almost entirely on phonological form, which is by definition arbitrary. I assume the agreement mechanism, by contrast, to be universal, invariable, and obligatory. What is language-particular is the specific way in which agreement is realized through morphology.

This leads to the heart of the matter: how does language-particular arbitrary gender interact systematically with universal agreement? The first part of the answer is that, even though gender may be arbitrary and often irregular, the grammatical and hence obligatory nature of gender forces the individual languages that exhibit gender distinctions to be systematic in the realization of these distinctions, despite any irregularities. In the case of gender, this does not mean that every noun, what Corbett, in his survey of gender systems (Corbett 1991), calls the *controller*, must bear a gender marker (indeed there are languages in which gender is always covert and is never actually manifested phonologically on nouns themselves). Instead it means that agreement morphology on what Corbett calls the *target* must always be realized (except of course when the morphology provides no actual phonological agreement marker), even when the controller does not provide sufficient information.

The normal way in which systematicity emerges in the face of irregularity or lack of information is through defaults (Zwicky 1986, Fraser and Corbett 1997). I will assume that the final default or unmarked type of agreement morphology is copying and that this copying is a reflection of the underlying universal syntactic mechanism of agreement: complete copying of the controller onto the target.[1] This complete copying, though, is normally disguised by the morphology of the language through its arbitrary system of genders. Furthermore, complete copying violates a basic corollary of the lexicalist or lexical integrity hypothesis that has been elaborated by a number of researchers (Zwicky and Pullum 1988, Anderson 1992), which Zwicky and Pullum call 'the principle of phonology-free syntax' and according to which, in the words of Anderson (1992, 84), 'the syntax neither manipulates nor has access to the internal form of words'. If agreement morphology were to actually copy the

[1] Pollard and Sag (1994) argue that agreement is not done by copying, but rather by indexing. This view has the advantage of being able to unite pronoun/antecedent agreement with verb/argument agreement. For my purposes, whether agreement is done by copying or by indexing is not important. What matters is that there be some way for the target to be sensitive to all the information contained in the controller, whether through indexing or through copying. Pollard and Sag provide a number of arguments against the directionality of the relationship between controller and target, none of which I find persuasive. See Lapointe (1988) for discussion of directionality in agreement.

entire controller lexeme, then it would contravene this generalization, since the lexeme includes its phonological form(s). In this light, one can see gender as a partial solution to the conflict between the principle of phonology-free syntax and the copying nature of agreement: actual agreement morphology is not normally sensitive to the entire controller lexeme; instead it is sensitive to only the grammatical features of the lexeme, including a set of morphosyntactic features. In particular, agreement does not usually need to know the phonological form of the lexeme, so that the principle of phonology-free syntax is preserved in most instances. Actual phonological copying only emerges when it is forced to, through the rare phenomenon of *alliterative agreement*, because for some reason there is no gender available (Dobrin 1997). The interaction of language-particular and irregular gender systems with the universal mechanism of agreement thus sheds light on the nature of the agreement mechanism. Indeed, it is through the idiosyncrasy of individual language systems that we come to appreciate what is universal in language in this domain.

2. Ideology

There are no wholly natural languages or wholly natural grammars. The idea of natural languages or natural grammars is the historical residue of earlier stages in our thinking about human language. Natural language, not natural languages, lies at the base of linguistic theory. By *natural language* I mean Chomsky's U[niversal] G[rammar], or more precisely the species-specific cognitive propensity proposed by Lenneberg on which UG rests:

> The appearance of language may be thought to be due to an innately mapped-in program for behavior, the exact realization of the program being dependent upon the peculiarities of the (speech) environment. As long as the child is surrounded at all by a speaking environment, speech will develop in an automatic way, with a rigid developmental history, a highly specific mode for generalization behavior, and a relative dependence upon the maturational history of the child.
>
> (Lenneberg 1964, 600)

Actual human languages are not wholly natural objects but rather partly cultural objects rooted in the interaction of natural language with individual cultures and happenstance. Grammars (I-languages) are not wholly natural objects, unless we attempt to exclude from I-languages everything that is determined by culture and accident. Lenneberg's program does not include a theory of I-languages (what is acquired), but is a set of tools, the universal human capacities that contribute to the construction of I-languages, 'the specific innate abilities that make [language acquisition] possible' (Chomsky 1965, 27). Actually realized human languages are systematic and at least partly unnatural.

Assuming that human languages are not entirely natural, how do we confront the residual but systematic unnatural aspects of languages? One reductionist approach says that this systematic unnatural residue is of no interest. My own preference is to address the residue directly, on the working assumption that actual languages are the product of the interaction of nature and culture. I will attempt to show in this paper that such direct confrontation with residual but systematic aspects of individual languages can in fact lead to a better understanding of the universal mechanisms of the human language program.

3. Reductionism, structuralism and morphology

There are three major types of reductionism in linguistic theory. According to innatist

reductionism, languages can be reduced to innate principles and mechanisms that are specific to the language program, as opposed to other human propensities or behaviors — for example, Chomsky's 'perfect system', which is not realized in any actual language, but which resides in the unrealizable innate language faculty. According to social/cognitive reductionism, non-language-specific principles of cognition, communication, or social structure are the underlying explanatory causes of languages. Finally, one may assume both together and try to account for languages exhaustively in terms of the interaction of natural factors, some of which are specific to language and some of which are traceable to general cognitive and social principles. From any one of these viewpoints, there is no interesting residue in individual languages.

Within the structuralist tradition, the arbitrariness of the linguistic sign dictates that every actual language must be in part a conventional system, what Saussure calls 'a system of pure values' (1959, p. 111), and therefore no individual language can be understood entirely in terms of internal (innate) or external (social) common human factors. The fact that the sign is arbitrary means that there will be systematic unnatural (purely conventional) structural components or subsystems in every individual human language. It is important to realize that acknowledging the arbitrary aspect of individual languages does not commit one to claim, as some structuralists did, that language is entirely arbitrary or unnatural, or that languages may vary among themselves without limit. Instead, one is acknowledging that individual languages always arise through the interplay of universal and arbitrary factors and then allowing for some of the systematicity of individual languages to emerge from this interplay, resulting in unnatural but systematic aspects of languages.

The most likely candidates for partly unnatural subsystems of languages are those that are systematic but not universal and that are close to the arbitrary bond between the signifier and the signified that constitutes the linguistic sign. Inflectional morphology is not universal, but in languages where it exists, it is highly systematic, and morphology lies as close to the arbitrary sign bond as anything does. Certain aspects of inflection can be traced to natural or general social/cognitive functional factors, but full-blown inflectional systems are usually unnatural in part. Inflectional morphology is therefore a good candidate for the sort of system that I have in mind.

4. Gender

'Genders are classes of nouns reflected in the behavior of associated words.' (Hockett 1958, 231). More narrowly, genders are agreement classes:

Def: An agreement class is a set of nouns such that any two members of that set have the property that

whenever	(i)	they stand in the same morphosyntactic form
and	(ii)	they occur in the same agreement domain
and	(iii)	they have the same lexical item as agreement target
then	their targets have the same morphological realization.	

(Corbett 1991, 147; following Zaliznjak 1964)

Gender is an excellent candidate for an unnatural grammatical category, first because gender distinction within a language is far from universal and hence most likely not necessary. Gender flourishes in a few areas (Eurasia, Africa, New Guinea) and a few language families

(Indo-European, Afro-Asiatic, Nilo-Saharan, Niger-Congo, North Caucasian, Dravidian). Genders are also much more highly varied in their structure and motivation than the other categories usually involved in agreement: person, case, and number. These other categories are, by contrast, fairly restricted across languages. Person systems universally distinguish speaker, addressee, and other, with a few distinguishing further among others. Number is not always distinguished in the grammar, but when it is, it follows a simple implicational system: if the languages distinguishes any particular n by means of a special class, then it also distinguishes n-1 by means of a special class; if the languages distinguishes any particular n by means of a special class, then it distinguishes $n > 1$ (plural). Case is more variable than number or person, but it obeys a fairly strict hierarchy. But gender systems vary quite widely, not only in the number of genders that a language has, but also in the cognitive basis for these genders (Corbett 1991).

For any language, the set of genders partitions the set of nouns in the language. For most languages, we may say that there is only one gender, or no partitioning of the set of nouns. For languages where the set of nouns is partitioned into more than one gender, even if there is no well-defined criterion for this partitioning, the partitioning is exhaustive.

Gender has been much discussed from a reductionist point of view, with little success. Formal syntax provides no account of gender, though formal syntactic theories have made much progress in our understanding of agreement, which is related to but independent of gender. The mechanism of agreement has been reasonably well explored within a variety of formal theories (G/HPSG, LFG, GB, Minimalism). From a syntactic point of view, the problem of agreement is how and under what syntactic circumstances the agreement takes place, but there is no formal syntactic reason within any theory for a language to have genders, and indeed most actual languages do not.. Theories differ on the details of how agreement is accomplished, but in any theory, information about the controller lexeme must somehow get to the target. The simplest possible mechanism for this in any theory is copying some aspect of the controller onto the target. I will therefore make the minimal assumption: that agreement is always done in the syntax by copying the controller lexeme in its entirety onto the target and that this will be true in any theory. This says nothing whatsoever about the morphology of agreement, though.

Social reductionists have had more success with genders, though not with agreement. There is a reasonably successful literature within the Greenbergian typological tradition tying gender systems historically to numerical classifier systems of the sort found in East Asia and elsewhere, though there are few if any real examples of languages moving historically from one type to the other and few examples of intermediate types. One possible intermediate case is Yagua (Payne 1986). Classifier systems are rooted in general cognitive principles, as exemplified by the titles of some of Lakoff's works on the subject: 'Classifiers as a reflection of mind' (1986); *Women, fire, and dangerous things: What categories reveal about the mind* (1987). There is a large literature on functional motivations within individual gender systems, especially explanations of exceptionally classified nouns (Denny and Creider 1976; Lakoff 1986, 1987; Zubin and Köpcke 1984,86).

The major differences between genders and classifiers are rooted in the fact that gender is morphological and grammatical while classifiers are lexical items (Dixon 1982, 1986).

i. Classifiers are always morphologically free separate lexemes.

ii. Overt realizations of gender are always morphologically bound.
iii. Gender is obligatory and normally fixed for any given noun; there is little or no variation among speakers and change in the gender of individual nouns is very slow (e.g. Spanish *mano* is feminine because it was feminine in Latin, even though nouns in *-o* are overwhelmingly masculine).
iv. Classifiers are variable or optional and often exhibit wide stylistic and age-related variation (Erbaugh 1986). In any language with classifiers, there is usually a class of nouns that admits none.
v. There are often large numbers of classifiers (400 in Tzeltal) and they are quite fluid.
vi. The number of genders is small (2 – 15) and stable.
vii. Gender is always reflected in syntax (by definition of gender as agreement class).
viii. Classifiers are realized outside the classified noun's NP only when used anaphorically.

A major concern of research on genders has been their lack of morphosyntactic coherence, the fact that there are normally many diverse linguistic criteria for membership of a noun in a gender. I have argued elsewhere (Aronoff 1994) that what is central is the exhaustive partitioning: membership in the gender itself as realized and evidenced through agreement, not any criteria for membership. The semantic coherence of genders has often been exaggerated (Lakoff 1986, 1987). In fact, semantically incoherent gender systems (in which there are exceptions to the semantic categorization that seems to underlie the system) are usually stable over very long periods of time; witness the gender systems of the languages of the two most widely-studied gender-intensive language families, Niger-Congo and Indo-European, all of which remain irregular as far back as we can reconstruct them. We would not expect such stability if genders were in principle rooted in semantic classification. Semantically mixed gender systems are also fairly common. In one type, the individual genders of the language may be motivated differently. Classic examples of this type are found in parts of Papua New Guinea. Arapesh and Yimas are the best described (Foley 1986, 1991; Fortune 1942). In Arapesh, one gender, male human, is entirely semantic. The eleven remaining genders are entirely phonological, except possibly for one, which contains human females, and another, which is the default. In Yimas, according to Foley, there are four semantic genders (male human, female human, higher animal, and plant), six purely phonological genders, and one default. In the other type of mixed gender system, members of a single gender may be motivated by a variety of factors. Latin is a good example: animates are typically assigned gender by their sex, when sex is differentiated; most names of plants are feminine; most names of rivers and mountains are masculine; stems ending in *Cr* are usually masculine; third-declension nouns with stems ending in certain sequences (e.g. *el, al, os*) are neuter, as are indeclinables. Semantically coherent genders emerge in the breakdown of languages, as in young people's Dyirbal (Schmidt 1983), where the last stage is the most coherent. This has been taken as evidence of the true nature of the robust system, but it may also be understood in the opposite way: robust gender systems do not need to be semantically coherent.

Gender is obligatory inasmuch as every noun in a language must belong to one of its genders, but because gender systems are not usually coherent, it often happens for some subset

of nouns that there will be no regular method for assigning gender to its members.[2] This subset is therefore placed in the default class (there may also be ordered defaults). One type of default class is partially motivated. While the bulk of its members are assigned by a reasonably well-defined criteria, others are assigned by default. This is fairly common. A more striking type is the 'garbage can', which has no members assigned other than by default. Yimas gender v may be an example of this type. It is very heterogeneous and seems to be best defined to include all nouns that do not fall under the semantic or phonological criteria set by the other genders.

5. Default agreement

The combined factors of the obligatoriness of genders and their frequent lack of coherence shed light on another phenomenon: default agreement-target gender. If an agreement target must show some gender, because gender is obligatory, but there is no well-defined agreement controller (for whatever reason), then the target shows a default gender, simply because target gender is obligatory. I will give example from default target agreement in Arapesh, which I have discussed at some length in Aronoff 1994. This language has 13 genders. Adjectives, prowords of all sorts, and verbs always agree with their controlling noun phrases. When an element must bear an agreement marker and the proper gender of that marker cannot be determined for whatever reason, then the marker always bears the mark of the default gender. Default agreement can arise in several ways: the head noun of the controller may be null; there may be gender clash between coordinated noun phrases within the controller, the head noun may not fit into one of the genders for some reason; or the controller may be outside the gender system (first and second person do not agree in gender with targets). See Corbett 1991, §7.2, for many other examples of this sort of phenomenon.

Most languages have a particular gender or set of genders for default target agreement, but in a few languages, default agreement takes the form of what I will call *radical alliterative agreement*, in which the agreement target simply alliterates or rhymes with its controller, which is outside the gender system. Radical alliterative agreement must be distinguished from the apparently alliterative agreement found in Niger-Congo languages and many languages of Papua New Guinea. In these languages, a number of the genders have overt affixes that are identical to the corresponding gender agreement markers. Corbett cites the following example from Welmers:

1) ki-kapu ki-kubwa ki-moja ki-lianguka
7-basket 7-large 7-one 7-fell
'one large basket fell'

This noun falls into the 7 class in the singular, which has the prefix *ki-*. This prefix then appears on all the targets that agree with this controller. The major difference between the Niger-Congo type of apparently alliterative agreement and true radical alliterative agreement is that what appears to be copied in these Niger-Congo cases is not part of the

[2] Borrowed nouns, which do not belong entirely to the language, may sometimes lie outside or on the borders of the gender system. They often figure prominently in alliterative agreement.

noun radical, but a segmentable prefix. This can be seen from the fact that some nouns in these languages have no prefix, but still get 'alliterative' gender agreement and from the fact that some nouns have the 'wrong' prefix. Furthermore, in apparently alliterative systems, there is usually a fairly small fixed number of alliterative affixes. In true radical alliterative agreement, there is no limit on the number of possible number of alliterative agreement markers except that imposed by the phonology of the language. Historically, it is likely that apparently alliterative agreement always arises from radical alliterative agreement, but that is a separate issue. I will sketch some examples of radical alliterative agreement, in order to give the reader a taste of how it works.

In Godie (Marchese 1986,88), a Kru (Benue-Congo) language, agreement targets (definite, demonstrative, and adjective) agree with the final vowel of the controller noun in backness (front, central, or back). We may say that there is agreement, expressed in terms of the final vowel, but that there are no genders, since the actual form of agreement is always completely predictable.

2) mlɛ̀-ɛ ɔ nii nə, ɛmɛ plʉ̀ lʊ̀ɛ
animal-DEF he saw NON-FINAL it is elephant

3) nyʉ̀kpɔ kəd-ɔ nɔ nii mlɛ kəd-ɛ
man big this saw animal big (data from Dobrin 1996)

We see in both examples that the quality of the last vowel of the controller is reflected in that of the target definite marker, demonstrative, or adjective.

Default radical alliterative agreement appears clearly in Bainuk (Sauvageot 1967), an Atlantic (Niger-Congo) language, which has both prefixed and unprefixed nouns. Prefixed nouns fall into eleven genders (each containing a matched pair of singular and plural prefixes) and show agreement through the prefix, which appears on pronouns, demonstratives, and adjectives (as a prefix or suffix, depending on the category of the target). Unprefixed nouns, which are apparently loans, lie outside this gender system; they show no singular marker and a plural suffix (a harmonic nasalized vowel) instead of a plural prefix. The agreement targets of individual nouns of this unprefixed type take one of two sorts of gender agreement in both singular and plural, either a default prefix *a-* or a suffix copy of the first CV of the noun stem (Sauvageot can find no reason for which method a given noun 'chooses'). Some examples are given in the following Table:

Prefixed

si-dēn-o in-si	mu-dēn-o um-mu
'this canoe'	'these canoes'
si-dēn si-wuri	mu-dēn mu-wuri
'long canoe'	'long canoes'

Unprefixed with copying of CV

katāma-ŋɔ in-ka	katāma-ŋɔ-nɔ in-ka-ā
river-DEF this-AGR	river-DEF-PL this-AGR-PL
katāma ka-wayi	katāma-ā ka-wayi-e
river AGR-wide	river-PL AGR-wide-PL

Note that the unprefixed nouns trigger the same gender agreement affix in both singular and plural, which is very unusual in Niger-Congo languages, where pairs of singular and plural agreement affixes are hallmarks of the gender system. Note also the co-occurrence of this agreement affix with the plural agreement suffix, again very unusual in Niger-Congo. Finally, these same unprefixed nouns may appear with a prefix in the diminutive or augmentative. For example, the unprefixed noun *saha* 'sheep' may receive the diminutive prefix *ko-* or the augmentative prefix *da-*. In these cases it will 'enter' the normal gender system and show the regular plural prefix for the diminutive or augmentative and no alliterative agreement. Since regular nouns allow only one prefix, the fact that the stem does not lose its initial CV in the diminutive or augmentative shows that it is indeed unprefixed.

In Wolof (McLaughlin 1994, 97), an Atlantic (Niger-Congo) language, there are ten singular noun classes and two plural (one of which has only 4-5 members), although there is no class marker on nouns, only on targets. In some cases, class assignment is alliterative: the class of a noun is determined by its initial consonant: *galaas gi* 'the ice'; *miskit mi* 'the biscuit'; *waliis wi* 'the suitcase'. This is true of many earlier loan words, arguing that the default method of class assignment was alliterative at an earlier period. Recently, the *b-* class has become the default, but the earlier assignments survive (supporting, incidentally, the 'once-only' theory of Aronoff 1976). As the methods of assigning defaults succeed each other historically, the gender system becomes layered.

A final spoken-language example comes from Arapeshan (Dobrin 1997). In Fortune's original grammar (1942), there were some thirteen distinct phonologically-based genders, each with its own set of target agreement markers. These were largely but not entirely alliterative. Fortune shows extensive evidence of a default gender, both for controllers and for targets. In Fortune's time, there were no singular nouns ending in *s*, which is common as the final sound in plural markers. In recent years, however, singular nouns ending in *s* have been borrowed from English and Tok Pisin. These show no plural form and take *-s* as a singular target agreement affix. The following examples are all from Dobrin 1997.

4) ənan-is kes
PRO.vii-AGR suitcase 'his suitcase'

5) balus sa-bih
airplane AGR-came.down 'the airplane landed'

6) bas s-afiʔi a-ndaʔ pasim-as
bus AGR-came I-did flag.down-AGR 'When the bus came, I flagged it down'

Dobrin takes this as evidence that alliterative agreement is a default, running in the

background and only surfacing in those instances where there is no gender on the controller and the controller does not fall into the default category either.

In American Sign Language (and most other sign languages), subject and object agreement is deictic. Participants are given locations in the signing space and agreement is encoded in the initial and final locus of agreement verbs. This type of agreement can be interpreted as copying the locus of the subject and object onto the verb as prefix and suffix. This is the purest example of agreement as copying, since there is no evidence for distinct genders, and shows that sign languages may be closer to the 'natural' state than spoken languages in this regard.

Dobrin (1996, 1997) has pointed out that alliterative agreement is paradoxical. Although it violates the principle of separation of syntax and phonology, it seems prototypical: agreement as copying. But if it is prototypical then why is it true that any form of alliterative agreement is rare, and where it does occur, it has been argued to be late historically:

> M. Richardson: Monsieur le Professeur Guthrie, vos remarques concernant l'évolution apparemment plus récente des morphèmes *<<allitératifs>>* d'accord (cl. 7, 8, 11, 12, 13, 19: 16, 17,18) pourraient conduire à une nouvelle hypothèse, celle du greffage sur le système originel d'une méthode allitérative de classification nominale qui serait plus récente.
>
> M. M. Guthrie: Il semblerait au moins possible que les classes où l'on constate pour les morphèmes d'accord une allitération complète fussent plus tardives, puisqu'une innovation devrait avoir eu pour effet la répétition d'un element identique plus probablement que les types d'accord quelque peu irréguliers caractéristiques des classes 1, 2, 3, 4, 5, 6, 9, et 10.
>
> (Manessy 1967, 353)[3]

The answer is that alliterative agreement reflects the universal syntactic nature of agreement as copying, but that this universal does not normally surface, because of the principle of lexical integrity. Only in rare cases where the obligatoriness of agreement meets a gap in the gender system does the copying mechanism surface. Alliterative agreement is therefore a paradigm case of what has been called in optimality theory the emergence of the unmarked (McCarthy and Prince 1995): the unmarked case is a default that emerges just in case nothing else does. Alliterative agreement is gender agreement in the absence of a gender.

---

[3] Mr. I. Richardson: Professor Guthrie, your remarks concerning the apparently more recent evolution of the "alliterative" agreement morphemes (cl. 7, 8, 11, 12, 13, 19,; 16, 17, 18) could lead to a new hypothesis, that of the grafting onto the original system of an alliterative method of nominal classification which would be more recent.

Mr. M. Guthrie: It would seem to be at least possible that those classes where complete alliteration for the agreement morphemes is established were later, since an innovation would have had as its effect the repetition of an identical element more probably than the somewhat irregular types of agreement of types 1, 2, 3, 4, 5, 6, 9, and 10.

REFERENCES

Aronoff, Mark. 1976. Word formation in generative grammar. Cambridge, MA: MIT Press.

Aronoff, Mark. 1994. Morphology by itself: Stems and inflectional classes. Cambridge, MA: MIT Press.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Corbett, Greville. 1991. Gender. Cambridge: Cambridge University Press.

Fraser, Norman M, and Greville Corbett. 1997. Defaults in Arapesh. Lingua 103.25-57.

Craig, Colette, ed. 1986. Noun classes and categorization (Typological studies in language, 7). Amsterdam; John Benjamins.

Denny, J. Peter, and Chet A. Creider. 1986. The semantics of noun classes in Proto Bantu. In Craig, ed., 217-239.

De Saussure, Ferdinand. 1959. Course in general linguistics. Tr. By Wade Baskin. New York: The philosophical library.

Dixon, R. M. W. 1982. Where have all the adjectives gone? and other essays in semantics and syntax. Berlin: Mouton.

Dobrin, Lise. 1996. Literal alliterative concord. Chicago Linguistic Society 37.127-42.

Dobrin, Lise. 1997. The morphosyntactic reality of phonological form. Yearbook of Morphology 1997. 59-81.

Erbaugh, Mary. 1986. Taking stock: The development of Chinese noun classifiers historically and in young children. In Craig, ed., 399-436.

Foley, William A. 1986. The Papuan languages of New Guinea. Cambridge; Cambridge University Press.

Foley, William A. 1991. The Yimas language of New Guinea. Stanford: Stanford University Press.

Fortune, Reo. 1942. Arapesh (Publications of the American Ethnological Society, 19). New York: J. Augustin.

Hockett, Charles. 1958. A course in modern linguistics. New York: Macmillan.

Lakoff, George. 1986. Classifiers as a reflection of mind. In Craig, ed., 13-51.

Lakoff, George. 1987. Women, fire and dangerous things: What categories reveal about the mind. Chicago: University of Chicago Press.

Lapointe, Steven G. 1988. Toward a unified theory of agreement. In Michael Barlow and Charles Ferguson, eds., Agreement in natural language, 67-87. Stanford: CSLI.

Lenneberg, Eric. 1964. The capacity for language acquisition. In Jerry A. Fodor and Jerrold J. Katz, eds., The structure of language: Readings in the philosophy of language, 579-603.

Manessy, Gabriel. 1967. La classification nominale dans les langues Negro-Africaines. Paris: CNRS.

Marchese, Lynell. 1986. The pronominal system of Godie. In Ursula Weisemann, ed., Pronominal systems. Tuebingen: Gunter Narr.

Marchese, Lynell. 1988. Noun classes and agreement in Kru: A historical approach. In Michael Barlow and Charles A. Ferguson, eds., Agreement in natural language: Approaches, theories, descriptions.

McCarthy, John J., and Alan S. Prince. 1995. Faithfulness and reduplicative identity. In Papers in optimality theory, Jill N. Beckman, Laura Walsh, and Susan Urbanczyk, eds., 249-384. (University of Massachusetts occasional papers in linguistics 18.) Amherst, MA:

Graduate Linguistics Students Association.
Mclaughlin, Fiona. 1994. Inflection and phonological form in Wolof. In Frances Ingemann, ed., 1994 Mid-America Linguistics Conference papers, vol. 2. 436-44. Lawrence, KS: The University of Kansas department of linguistics.
McLaughlin, Fiona. 1997. Noun classification in Wolof: When affixes are not renewed. Studies in African Linguistics 26.
Payne, Doris L. 1986. Noun classification in Yagua. In Craig, ed., 113-131.
Sauvageot, Serge. 1967. Notes sur la classification nominale en Bainouk. In Manessy, ed. 225-236.
Schmidt, Annette. 1983. Young people's Dyirbal: An example of language death from Australia. MA thesis, Australian National University, Canberra.
Zaliznjak, A.A. 1964. K voprosu o grammatičeskix kategorijax roda i oduševlennosti v sovremennom russkom jazyke. Voprosi Jazykoznanija 4.25-40.
Zubin, David, and Klaus Michael Kopcke. 1984. Affect classification in the German gender system. Lingua 63.41-96.
Zwicky, Arnold. 1986. The general case: Basic form versus default form. Berkeley Linguistics Society 12.305-14.
Zwicky, Arnold. 1992. Some choices in the theory of morphology. In Robert D. Levine, ed., Formal Grammar: Theory and Implementation, 327-71. Oxford: Oxford University Press.
Zwicky, Arnold, and Geoffrey Pullum. 1988. The syntax-phonology interface. In Frederick J. Newmeyer, ed., The Cambridge survey, vol 1, Linguistic theory: Foundations. 255-80. Cambridge: Cambridge University Press.

Andrew Spencer, University of Essex
spena@essex.ac.uk

RELATIONAL ADJECTIVES AND THE REDUNDANCY OF LEXICAL CATEGORIES

Abstract

In this paper I argue that the familiar lexical category labels, N, V, A, P or equivalent feature systems (e.g. [±N, ±V]) are redundant in a theory which admits a level of argument structure. I modify Zwarts' (1992) conception of a-structure by arguing that major class members always include a 'referential role': <R> for nouns, <E> ('eventuality') for verbs and 'A' ('attributive') for adjectives The <A> r-role is coindexed with the <R> role of the noun modified. Reference to categorial information can be read off the a-structure representations without the need for purely syntactic category features. 'Transpositions', in which just the syntactic category is shifted, are operations over a-structures. I illustrate this system in detail with respect to relational adjectives. I first develop a (constructional) semantics for compound nouns (N N) in which the modifier receives a new <A> role, with demotion of the original <R> role: *atom* <A: R> *bomb* <R>. Relational adjectives have the same a-structure representation with the same semantic interpretation, but lexically specified: *atomic* <A: R> *bomb* <R>.

1. Introduction[1]

In this paper I assume that the representation of a predicate includes a level of a(gument)-structure (cf Williams 1981). In the theory of theta-discharge advanced by Higginbotham (1985), it is positions in a-structure which are bound, identified with or marked by thematic elements such as verbs and adjectives. Higginbotham (1985) proposes that the PAS of a verb includes a position corresponding to the notion of 'event' and that this position is accessible to modification by adverbials and the tense operator. Williams' (1981) original model included a 'referential' role for nouns, which is coindexed with the thematic (semantic) roles of verbs when the verb

[1] I am grateful to the Economic and Social Research Council for supporting the research reported here, Project no. R000236115. Thanks to David Adger for alerting me to Zwarts 1992. This paper was also read to the meeting of the Associação Portuguesa de Lingüística, University of Lisbon, 1 October 1997and will appear in the Proceedings. Thanks to Keith Brown for detailed discussion of earlier versions and also to participants in seminars at the University of Essex, the Workshop on Inflection, 8th Morphology Conference, Vienna, 6th February, 1996, the Research Centre for English and Applied Linguistics, Cambridge, the LAGB Spring 1997 meeting (Edinburgh), the ESRC-funded 'Challenges for Inflection Description' (18 April 1997), particularly Roger Evans and Gerald Gazdar, and Lexical Functional Grammar '97, University of California, San Diego, particularly Joan Bresnan, Phil LeSourd, Joan Maling, Nigel Vincent.

discharges its semantic role onto a complement or subject. Adjectives are generally assumed to be one- or two-place predicates which have semantic roles akin to those of verbs, but without the event role. These assumptions are summarized in the amalgam shown in (1), which I shall treat as a 'traditional' view on PAS (ignoring prepositions):

(1) 'Traditional' PAS representations for transitive V, N, A:

kick	<E, Ag, Th>
tree	<R>
tall	<Th>
afraid(-of)	<Exp, Th>

There is considerable redundancy between PAS representations and lexical syntactic category membership: <E> = Verb, <R> = Noun, <Bare Theta role(s)> = Adjective. Note that this is more than a rehearsal of the 'notional' parts of speech tradition: with an intermediate level of argument structure as in (2) we open up the possibility that syntactic category membership might become redundant even in more complex cases of 'mixed' categories' such as deverbal nominalizations, or in denominal adjectives. That is particularly true of theories which make use of mapping principles governing argument realization (the Theta Criterion, Function-Argument Biuniqueness etc.), in which insertion into the syntax of a lexical item of the wrong category would cause the derivation to crash simply because of failure of argument selection.

The thesis to be defended here is that, given a level of a-structure, lexical syntactic category features such as [±N, ±V] or their equivalent are entirely superfluous, their place taken by the . Lexical categories can be defined in terms of their 'r(eferential)'-roles. At the same time, many of the properties which are often attributed to major category features are better thought of as properties of the functional categories or functional features (f-features) which accompany major parts of speech, such as determination, tense, agreement features of various sorts, and so on. These are assigned to lexical items on the basis of their a-structures by universal principles modulated by language particular codicils.

This perspective throws light on the problem of distinguishing inflection and derivation. One rather serious problem is the existence of inflectional morphology which changes syntactic category such as the verbal participle, which in many languages clearly behaves like part of the verb paradigm (and shows, for instance, tense and/or aspect distinctions as well as retaining the argument structure of the verb), while on the other hand it inflects like an adjective. Likewise, gerunds, infinitives and deverbal nominalizations ('action nominals'), which in many languages inflect rather like nouns (in taking case endings, for instance) pose another well-known problem (cf Haspelmath 1996). Less obviously problematical, but no less troublesome in some languages are noun-to-adjective transpositions, or relational adjectives. These transpositions are summarized in (2):

(2)

N	→	Adj	relational adjectives
V	→	Adj	participles
V	→	Adv? N?	gerunds
V, A	→	N	action nominals (incl. infinitives)

Consider for instance, deverbal action nominals. As is clear from typological surveys such as Koptjevskaja-Tamm (1993), the action nominal may retain a number of a-structure properties from the original verb (such as licensing subject- and object-like satellites), may assign the same quirky case to its object as the original verb (as when nominalizations of Russian transitive verbs with Instrumental case marked objects continue to assign the Instrumental to their complements), may retain tense marking (Turkish, Quechua) or aspect (Polish) and so on. For this reason nominalizations are often called 'mixed categories' (Lefebvre and Muysken 1988). Where the nominal simply names the event denoted by the verb, to what extent are we justified in saying that the nominal is the result of derivational as opposed to inflectional morphology? In other words, in *the shooting of the lions by the hunter* why can *shooting* not be a word form of the lexeme *shoot*? This problem is particularly acute in a language like German in which the commonest and most productive action nominal is the infinitive form of the verb (NB!) used as a noun, i.e. bearing nominal features of determination and case and being modified by adjectives.

I shall argue here that these problems largely evaporate if we admit that there are no syntactic lexical categories. Category-changing inflection is a species of PAS alternation. A deverbal action nominal will be a verb whose event role, 'E', has been 'demoted' and supplanted by a nominal 'R' role, indicating that its denotation is the name of an event, rather than the event itself: *shooting* <R: E, Ag, Th>. Language-particular principles then determine whether the <R> role or the <E> role is responsible for licensing arguments, tense/aspect features and so on.

In this paper I shall concentrate primarily on attributive modifiers and explore the relationship between N + N compounding in English and relational adjectives[2]. I begin with a survey of Zwarts' (1992) exploration of the homologies between syntactic structures and semantic structures.

2. Zwarts' model

Zwarts (1992) proposes a theory of lexical categories in which there is considerable redundancy between semantic and syntactic representation. He assumes a standard type-theoretic semantics together with a level of a-structure. A-structure representations are headed. He proposes that the four major categories of N V A P have a *referential argument position*, or 'r-role' as shown in (3):

[2] A treatment of nominalizations is given in Spencer (1998).

(3)	Noun	<R>	R = referent
	Verb	<E: Ag(x), Th(y), ...>	E = event
	Adjective	<G: Th(x)>	G= degree
	Preposition	<S: Th(x), Ground(y)>	S = space

'R' and 'E' have their familiar interpretation. The referential role of an adjective is denoted by 'G' (for 'gradable') and stands for a degree. The idea is that a canonical adjective phrase specifies the degree to which a property or attribute holds of an entity (here expressed as the Theme of that adjective)[3].

The referential roles help distinguish 'sorts' of predicates and in this sense are entirely different in function from the thematic roles. Thus, while an adjective, intransitive verb, intransitive preposition and common noun might all be analysed as one-place predicates of the type $\langle e, t \rangle$ they can be distinguished by their referential arguments. The other function of the referential arguments is to serve as the locus of 'theta discharge' (Higginbotham 1985). It is the r-role which is bound by determiners, tense operators, degree modifiers and so on. It is also the position which is coindexed with argument positions of predicators. Thus, in *the man sleeps* or *hit the man*, a theta role in the argument structure array of the verb *sleep* or *hit* is coindexed with the R role of *the man*. Finally, it is the r-roles which are coindexed by theta identification in modification. Thus, in *the tall man*, the G position of *tall* and the R position of *man* are coindexed to indicate the fact that *tall* modifies *man*.

This would give us a one-one correspondence between syntactic category labels and r-roles, so the syntactic categories seem to be redundant. However, Zwarts argues that proper nouns, stative verbs and non-gradable adjectives differ from their canonical counterparts in lacking a referential argument. Given this, the only thing which will distinguish, say, an intransitive stative verb such as *live* (as in *Jesus lives*) from a non-gradable monadic adjective such as *alive* will be the syntactic category features. This gives us the subclasses shown in (4):

(4)	common noun	<R>	proper noun	<>
	eventive verb	<E: ...>	stative verb	<...>
	gradable adjective	<G: Th>	non-gradable adjective	<Th>

I will briefly consider in turn each of the three categories for which this proposal is made.

In Spencer (to appear) I show that there are problems with this interpretation. For instance, if proper nouns lack the r-role, then we no longer have a uniform account of theta discharge to nominal arguments: *John* in *hit John* has to be theta-marked by a different mechanism from that which marks *the man* in *hit the man*. Similarly, if stative verbs lack an E-role then the theta binding by the Tense operator has to be given a gratuitously disjunctive definition (as in Zwarts 1992:131).

[3] I ignore prepositions here.

For adjectives, Zwarts argues for a distinction between those that are gradable, such as *tall*, *red*, *pretty* and those that are not. The latter include simple binary adjectives such as *dead* or *married* but also denominal relational adjectives such as *adjectival*, *atomic*. Zwarts, however, draws a finer distinction between measure adjectives such as *tall*, *old*, *rich* and non-measure adjectives such as *pretty*, *healthy*, *lazy*. The measure adjectives can take some kind of measure phrase (*two meters tall*) while the non-measure adjectives, while gradable (*very pretty*), don't denote properties which can be expressed as sets of degrees along a scale. The measure adjectives have in their argument structure a 'G' referential role which is bound by degree expressions. All other adjectives denote simple properties and lack the referential role in their a-structure. In order to express the fact that non-measure gradable adjectives like *pretty* can still receive degree modification (*very pretty*) Zwarts assumes type shifting. The type of simple properties will be e_P, corresponding to an argument structure with just a Theme role, <Th(x)>, while the type of measure adjectives such as *tall* is $\langle e_D, t \rangle$, where e_D is the type of degrees, with a-structure <G: Th(x)>. Thus, by shifting from *pretty* $\langle e_P \rangle$ to *pretty* $\langle e_D, t \rangle$ we obtain an argument structure <G: Th(x)> for *pretty* and this maps a property to the set of degrees that realize that property. (Type shifting also accounts for cases in which proper nouns are modified, e.g *the young Einstein*).

What remains unclear is why non-measure adjectives such as *pretty* fail to take measure phrases when they undergo type shifting. The representation for *tall* will be something like (5):

(5) $\exists d[\mathbf{tall}'(x, d) \& d > dA]$

where *dA* refers to some 'average' or 'standard' degree of tallness (p. 138 ex. (2c)). But this means that the difference between *tall* and *pretty* is essentially in the LCS representation, not in the a-structure, since both *tall* and *pretty* can be given an a-structure of the form <G: Th(x)>. Again, the facts of determination tell us about semantic incompatibility rather than a morphosyntactic failure of theta discharge. In fact, it is not obvious that *pretty* is a non-measure adjective, witness (6):

(6) Anna is twice as pretty as Bella

One of the differences between measure and non-measure adjectives is supposed to be that non-measure adjectives permit the entailment (7):

(7) x is more adj than y $\Rightarrow$ x is adj

Thus, if Anna is prettier than Bella, then Anna has to be pretty in some absolute sense. This is not true, however, of Anna is taller than Bella, since both could be very short. But this is a fact about syncategorematicity which is independent of measurability. For instance, not all syncategorematic adjectives like *tall* are necessarily measurable. Thus, *good* is the classic example of a syncategorematic adjective but it is impossible to

measure goodness. Likewise, there are measure adjectives which are not syncategorematic and in which entailment (7) therefore holds, as in (8):

(8) Your account is five pounds overdrawn ⇒ Your account is overdrawn

Thus, gradability is a matter for LCS representations (or perhaps encyclopaedic knowledge) and not an a-structure property.

We now turn to the nature of modification. Zwarts offers a fairly uncontroversial interpretation in (9) (p. 63):

(9) "A lexical head L is *modified* by a phrase XP iff:
a. L governs XP, and
b. the prominent argument of XP is coindexed with the referential argument of L."

The important part of this definition is (9b). The term 'prominent argument' refers to the first thematic argument in the theta array. For an intransitive adjective or preposition this will be the sole Theme argument, and for a transitive adjective or preposition, this will also normally be the Theme argument. An example with an intransitive adjective is (10):

(10) a. tall woman
b. tall<G: Th_i> woman<R_i>
c. λx[**tall′**<G: Th(x)> & **woman′**(x)]

Note that Zwarts' 'G' argument plays no role whatever in theta discharge here.

3. A revised theory of argument structure for adjectivals

In this section I shall begin with a consideration of the way in which a noun modifies another noun in a root compound and compare this with the modification of a noun by a relational adjective. This will motivate a new r-role, 'A', for adjectives which expresses their canonical function as attributive modifiers, replacing Zwarts' 'G' role.

3. 1 Compounds and relational adjectives

It seems to be widely accepted that compounds such as *atom bomb* are interpreted pragmatically (Downing 1977). The simplest way to account for such meanings is to assume that the compound construction itself was associated with an unspecified predicate, ρ, which asserts some pragmatically defined relationship between the denotata of the two nouns (cf Spencer, 1995), as in (11):

(11) λρλx[[**bomb′**(x)] & ρ(λw[w=x], λy[**atom′**(y)])]

In other words, an atom bomb is a bomb such that there is some relationship between the property of being an atom and the property of being that bomb. The semantic interpretation provides the modifier with the representation shown in (12):

(12) λPλρλz[P(z) & ρ(λw[w=z], λy[**atom′**(y)])]

The constructional meaning of a compound noun is given explicitly in (13):

(13) N_1<R_1> in the construction [N_1<R_i>, N_2<R_i>]
corresponds to λPλρλz[P(z) & ρ(λw[w=z], λy[**noun′**(y)])]
where **noun$_1$′** is the denotation of N_1

This means that the representation for *atom bomb* will be (14), which after λ-conversion collapses to (11):

(14) λPλρλz[P(z) & ρ(λw[w=z], λy[**atom′**(y)])](λx[**bomb′**(x)])

If proper nouns also have <R> referential role, this works equally for them:

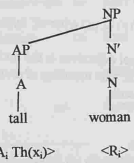
(15) a. London fog
b. London<R_i> fog<R_i>
c. λPλρλz[P(z) & ρ(λw[w=z], λy[**london′**(y)])](λx[**fog′**(x)])

The representation yielded by (13) is read off syntactic structure. We do not create a separate 'adjectival' lexeme every time we use a noun as modifier in a compound. Indeed, both the LCS and the PAS of the noun remain unaltered.

3.2 Attributive adjectives:

Zwarts 'G' r-role fails to bring out the principal function of adjectives, that of attributive modification. Let us therefore take all adjectives to have an ***attribute*** referential role, A, coindexed with the prominent argument. When modification occurs within a nominal phrase, the r-role, <A>, of the attribute is theta identified with the <R> of the modified noun, indirectly establishing a coindexation between the prominent argument of the adjective and the noun's referent, as shown in (16):

(16) a.



<A_i Th(x_i)> <R_i>

b. λx[**tall'**(x) & **woman'**(x)]

The default interpretation for (16) is given in (17):

(17) adj<A_i: x_i> translates as $\lambda Q \lambda x$[**adj'**(x) & Q(x)]

Applied to *woman* (translation λz[**woman'**(z)]) an adjective such as *tall* will give (18):

(18) $\lambda Q \lambda x$[**tall'**(x) & Q(x)](λz[**woman'**(z)]) ⇒
λx[**tall'**(x) & λz[**woman'**(z)](x)] ⇒
λx[**tall'**(x) & **woman'**(x)]

The account so far handles 'ordinary' qualitative adjectives such as *tall*, *pretty*, as well as non-gradable adjectives such as *married*. It will also handle derived adjectives such as *milky*, *girlish*, *cat-like*, *readable*, and so on. The relationship between, say, *cat-like* and *cat* is a matter of LCS and not PAS. Is this also true of relational adjectives such as *atomic*? That is, could we say that the relationship between the relational adjective *atomic* and the noun *atom* results from an operation over the LCS representation of the noun? This would mean, for instance, that *atomic* has some predicate, say, REL in its LCS meaning 'related to', giving [REL[ATOM]], just as *milky* means (very roughly) [LIKE[MILK]]. However, an element such as REL itself wouldn't really contribute anything to the LCS of the adjective. To call something an *atomic bomb* is to claim some relationship between the property of being that bomb and the property of being an atom, rather than attributing 'atomicity' to *bomb*. But this is exactly the pragmatically determined relation ρ used to define the constructional meaning of compounds. Hence, the relational adjective should be derived directly from the noun at the level of a-structure, in such a way that the noun acquires an attributive r-role <A> which then coindexes with the base noun's r-role <R>, as shown in (19):

(19) atomic: atom<A_i: R_i>

This can now be interpreted in the same way as the modifier in a compound noun, as in (20):

(20) noun<A_i: R_i> translates as $\lambda P \lambda \rho \lambda z$[P(z) & ρ($\lambda$w[w=z], λy[**atom'**(y)])]

In other words, the interpretation of relational adjectives is the lexical equivalent of the pragmatically defined relation in compounds. The meaning of *atomic bomb* is now derived as in (21), essentially as for *atom bomb*:

(21) a. atom<A_i: R_i> bomb<R_j> ⇒
b. $\lambda P \lambda \rho \lambda z$[P(z) & ρ($\lambda$w[w=z], λy[**atom'**(y)])](λx[**bomb'**(x)]) ⇒
c. $\lambda \rho \lambda z$[**bomb'**(z) & ρ(λw[w=z], λy[**atom'**(y)])

The basic interpretation of *atomic* is identical to that of the noun from which it derives, hence, (21b,c) make reference to the property λy[**atom'**(y)] and not the property λy[**atomic'**(y)]. The adjectival morphology is nothing more than a reflection of the changed a-structure of the noun, and not the bearer of a semantic constant, such as the *-like* of *cat-like* or the *-y* of *milky*. In this sense, then, the derivation of a relational adjective creates a distinct form of a nominal lexeme rather than creating a distinct adjectival lexeme.

Finally, how do we account for the fact that modifying nouns in compounds can (sometimes) be modified by adjectives, just like ordinary nouns (e.g. *red brick house*, *American history teacher = teacher of American history*)[4]? First, we form the phrase *red brick*. This is headed by a noun, though one which is modified by an adjective: *red*<A_i: Th(x_i)>(*brick*<R_j>). Then, the compound N interpretation rule converts the noun into a relational adjective to give (22):

(22) [*red*<A_i: Th(x_i)>(*brick*<A_j: R_i>)](*house* <R_j>)

This process is rare if the phrase is not listed (cf. **expensive brick house* in the sense *house made from expensive bricks*). This account of relational adjectives provides us with an unexpected solution to an intriguing problem. An expression such as *East German economy* illustrates a well-discussed kind morphosemantic structural mismatch: *East German* is clearly an adjectival form (essentially a relational adjective) derived from *East Germany*. But a part of what *East* is supposed to modify is lacking:

(23) a. [$_N$East Germany] b. [$_A$[$_N$East German-∅]??]

This is only a problem, however, if we persist in regarding the relational adjective as a new lexeme formed by derivational process. If we consider *German* (at least in (23b)) to be simply a form of the lexeme *Germany* then we can offer the analysis in (24), corresponding to (25):

(24) EAST<A_i> GERMANY<R_j> ⇒ EAST<A_i> GERMANY<A_j: R_i> (ECONOMY<R_j>)
East Germany ⇒ East German (economy)

(25) GERMANY<R_j> ⇒ GERMANY<A_j: R_j> (ECONOMY<R_j>)
Germany ⇒ German (economy)

[4] I am grateful to Phil LeSourd for discussion of this point.

The forms in small capitals in (25) are names of lexemes irrespect of their a-structures, while the word forms are given in lower case. The morphosemantic mismatch then disappears as an artefact of a wrong analysis (just as the past tense form *sang* doesn't represent a morphosemantic mismatch simply because it has no past tense suffix).

References

Downing, Pamela (1977). On the creation and use of English nominal compounds. *Language* 55: 810-42.

Haspelmath, Martin 1996. Word-class-changing inflection and morphological theory. In: G. Booij and J. van Marle (eds.) *Yearbook of Morphology 1995*, Dordrecht: Kluwer Academic Publishers, 43-66.

Higginbotham, James 1985. On semantics. *Linguistic Inquiry* 16, 547-621.

Jackendoff, Ray, S 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Koptjevskaja-Tamm, Maria 1993. *Nominalizations*. London: Routledge.

Lefebvre, Claire and Muysken, Pieter 1988. *Mixed Categories*. Dordrecht: Kluwer Academic Publishers.

Spencer, Andrew 1995. Noun incorporation in English. Talk given to the Workshop on Noun Incorporation at the Inaugural Meeting of the Association of Linguistic Typology, Filologia eta Geografia-Historica Fakultatea, Euskal Herriko Unibersitatea, Vitoria-Gasteiz, 8 September 1995.

Spencer, Andrew (to appear). The redundancy of lexical categories. In Andrew Radford (ed.): Studies presented to Prof. Keith Brown on the occasion of his retirement. Essex Research Reports in Linguistics.

Williams, Edwin 1981. Argument structure and morphology. *The Linguistic Review* 1: 18-114.

Zwarts, Joost 1992. *X'-Syntax - X'-Semantics: On the Interpretation of Functional and Lexical Heads*. Utrecht: Onderzoeksinstituut voor Taal en Spraak.

Anna Anastassiadis-Symeonidis
Aristotle University of Thessaloniki

ON MODERN GREEK DENOMINAL ADJECTIVES*

ABSTRACT

The Modern Greek adjectival system is described by means of D. Corbin's model and illustrated with the analysis of four suffixes.

1. Introduction

This contribution aims at presenting the general principles of the Modern Greek (MG) denominal adjectival system according to Danielle Corbin's model (University Lille3, France)[1]. This model explicitly formalises an ordered set of rules and principles that synchronically characterise the constructed part of the lexicon of a language, allowing us to distinguish between accidental and systematic gaps. Thus it is a model whose aim is to describe the speakers's lexical competence, by means of formal grammar and as a result to interpret not only the constructed words already existing but all the words that it would be to construct (possible words), since their creation will depend on the same rules. In contrast to many other theories and models, the notion of the attested form here acquires only relative importance as far as the lexicon is concerned, since it is only the concept of the possible word that is significant at the level of lexical competence.

As a result the description of the lexical system does not correspond to the observable lexicon. In other words the course of abstraction adopted for the remaining levels of linguistic analysis has to be followed by the study and interpretation of the lexicon as well. A form is not considered to be regular or irregular according to the surface structure which is dominated by heterogeneity nor is a form considered to be non-existent if it is not directly observable.

In this paper, we will first examine Corbin's model, i.e. the principles of analysis, namely associativity and stratification, the Word Formation Rules (WFR), the stratification of the meaning of the constructed words, the modules of the lexical component, and secondly the MG denominal adjectival system.

* Many thanks are given to Danielle Corbin for reading this paper and for her invaluable suggestions.

[1] For more details see Corbin (1987 and 1991) and Anastassiadis-Symeonidis (1992).

2. Corbin's Model

2. 1. Principles of analysis

Derivation morphology constitutes an autonomous grammatical component which is internally stratified into four subcomponents (principle of stratification). The meaning of a constructed word is structured at the same time as its morphological structure (principle of associativity). The choice of the principle of associativity as a way of presenting the relationship between the form and meaning of the constructed words involves, on the one hand, an accurate definition of the notion of a constructed word and, on the other, the possibility of rectifying a distortion between the form and the meaning that may appear in the observable data. The derivational base is not the word, but rather the lexical morpheme. It is all too clear that because MG is an inflectional language, the lexical morpheme is accompanied on its right by the inflectional morpheme, when it is actualised in discourse.

The constructed word is the construction resulting from the application of a WFR of a morphological structure which does not necessarily coincide with the observable form and meaning of the word, not necessarily identified. A word, in order to be considered constructed has to satisfy three conditions:

a) Each of the constituent elements of the morphological underlying structure belongs to a grammatical category and is related to one meaning in a reproducible way.

b) The predictable meaning is compositional in relation to the underlying morphological structure.

c) Possible distortions have to be interpreted with special regular mechanisms[2].

2. 2. The Word Formation Rule

The WFR is the application of one of the means (for example a suffix) of morphological paradigm on to its base (= a lexical morpheme belonging intrinsically to a grammatical category) and four components are included :

a) A structural operation that imposes a categorial relationship between the constructed word and its base, ex. The relational WFR constructing adjectives from nouns.

b) A semantic operation that constructs the basic meaning of all the words constructed by the relational WFR, eg. the meaning of the adjectives resulting from the application of the above mentioned rule is "of/concerning the Noun".

[2] The distorsions between the underlying structure and the surface structure, the form and the meaning of constructed words may be settled because the surface structure of a constructed word may derive from the application of rules that correct an underlying structure which cannot be presentable in the way it is. These distorsions may be caused by the truncation of a segment or the presence of a class marker.

c) A morphological paradigm which includes all the means of construction.

d) Various restrictions eg. the restrictions concerning the compatibility between the suffix and its base.

2. 3. The meaning of constructed words

The semantic interpretation of a constructed word includes (i) the predictable meaning and (ii) the conventional meaning.

(i) The predictable meaning is not identical either to the definitions given by traditional lexicography –this is attested meaning– or to its paraphrase given by a native speaker, but rather it is a combination of three parameters: (a) the meaning which is constructed by the WFR and is shared by all the products of the same WFR. (b) The meaning that is specified by the morphological means used constructing adjectives from a noun base. (c) The meaning inherited from the base and given to the constructed word.

(ii) The conventional meaning of the attested constructed words results from the adaptation of the predictable meaning to the extra-linguistic reality[3].

2. 4. The lexical component

The lexical component concerns the "internal syntax" of constructed words, that is, it includes all the necessary operations so that the form, the morphological structure, their semantic interpretation as well as the principles that govern the combination of all these in the possible and attested constructed words of a language are interpreted. The lexical component is stratified into four (sub)components :

a) The **base component.** It includes (i) the list of base entries each of which is accompanied by its properties (phonological representation, grammatical category, distinctive elements, ex. degree of affix availability), as well as (ii) rules.

b) The **derivational component.** It includes a list of WFRs and it is there that the potentially constructed words of the language are produced, each of which is accompanied by all its predictable properties. The WFRs are applied to base entries that belong to the major categories (autonomous or non autonomous). The WFRs are complex operations that simultaneously construct a) the morphological structure and b) the predictable meaning of constructed words. The products of the derivational component may appear in syntactic structures, except for those in which the rules of post-derivational component apply.

c)The **post-derivational component.** It includes operations that correct the forms produced according to the WFR and cannot appear in this way in the surface structure. These operations are mechanisms concerning form that can function in a recurrent way

[3] The acquisition of the conventional meaning of constructed words consists in 'unlearning', in this way limiting the possibilities offered by the WFRs.

and allow the transition from the predictable level to the observable level in case of distortion. These mechanisms are : (i) the rules of morphological truncation and (ii) the rules of class marking. With these mechanisms, the number of exceptions in the lexicon is reduced significantly.

(i) The rules of truncation allow us to explain the absence of segments in the surface structure, the presence of which is predicted by a morphological structure in accordance with their interpretation, e.g. τένις 'tennis' -ίστας 'ist' -->τενίστας 'tennis-player', αν-ενημέρωτος -->ανημέρωτος 'non-informed'.

(ii) The class marker is a suffix-like ending without having any semantic role. Its task is to give to the word, where it is present, the suitable form as far as its grammatical category or its reference class are concerned and to allow us to explain the presence of suffix-like ending not necessary for semantic interpretation, in the surface structure. Cf. e.g. άκαρδος 'heartless' and ανεδαφικός 'groundless', products of the same WFR that are differentiated as far as the presence of the class merker -ικός in ανεδαφικός is concerned. The class marker is not accidentally selected, but it is copied according to the copy principle from the suffix that bears the corresponding suffixed, but not prefixed, adjective, e.g. διεθνικός 'international' where the class marker -ικός is copied on the suffix –ικ(ός)[4] of εθνικός 'national'. The copy principle concerns a surface process with a wide range of applications entitled to limit the selection of the ending segment. The constructed lexicon includes the products of the derivational component and the post-derivational component.

Finally d) the **conventional component.** Its task is to interpret the conventional lexicon, that is the unpredictable sector of the constructed lexicon.

3. The Modern Greek Denominal Adjectival System

Herein, we will attempt to apply the theoretical framework presented above to four suffixes of the MG language. We will be presenting them in the following order : 1) -΄ιν(ος), which is connected with materials/substances, 2) –ιάτικ(ος) which is connected with time. The third and the fourth suffixes, -ίστικ(ος) and –(ι)άρ(ης) respectively, are connected with negative connotations.

3. 1. The suffix -΄ιν(ος)

In order to determine the range of this suffix, we should emphasize the importance of the posistion of the stress on the word in MG. In this way, we can distinguish between a) the suffix -΄ιν(ος) e.g. πέτρινος 'made of stone', b) the suffix -ιν(ός) eg. βραδινός 'evening' and c) the segment -ίν(ος) eg. θεατρίνος (ham actor).

In adjectives formed using the suffix -΄ιν(ος), the noun base is, first of all, interpreted extensionally, eg. : in πέτριvο γεφύρι 'bridge made of stone'. But it can be

[4] The inflexional morpheme is placed in brackets, which is a standard convention in this model.

interpreted intensionally, eg. : in πέτρινη καρδιά 'heart of stone'.

Let us, therefore, compare πέτρινο γεφύρι 'bridge of stone' with πέτρινη καρδιά 'heart of stone'. The intensional interpretation of the noun base is considered to be derived from its extensional interpretation. If a bridge is made of stone, a human heart can not possibly be, apart from perhaps having one or more properties of stone, in this case the hardness of that material.

The selection of one or other interpretation of the noun base is realised in relation mainly to the modified noun. Consequently, the suffix -΄ιν(ος) allows both an intensional and an extensional interpretation of the noun base.

In an attempt to determine the type of base to which the suffix -΄ιν(ος) can be applied, we can ascertain that the noun base preferably denotes material, eg. the nouns πέτρα 'stone', μάρμαρο 'marble', μπρούντζος 'bronze', πηλός 'clay' etc. and therefore belongs to uncountable nouns. This particular suffix cannot be applied to a noun base which does not denote material. However, should the primary meaning of a noun be something other than material, the application of the suffix -΄ιν(ος) to that base produces the selection of properties which are connected still with the meaning of 'material', eg. βροχή 'rain' (a meteorological phenomenon), but in the combination βρόχινο νερό 'rainwater', the word βροχή 'rain' functions as a noun which denotes material. Consequently the semantic relationship imposes restrictions on the selection of the noun base and of the modified noun.

We are certain that in the case of the extensional interpretation of the noun base, the adjectives are not modified in terms of degree *πολύ πέτρινο γεφύρι *'a very stone bridge'. Conversely, the adjectives are modified when the noun base has an intensional interpretation, eg. *πολύ πέτρινη καρδιά *'a very stone heart'. Consequently, the syntactic characteristics of the denominal adjectives are connected to their predictable meaning.

Finally, MG nowadays has a small number of adjectives which denote colour and which are formed with the suffix -΄ινος : πράσινος 'green', κίτρινος 'yellow', κόκκινος 'red' etc. We consider that the noun bases of these adjectives (πράσο 'leek', κίτρο 'citron', κόκκος 'tinctorial grain') can take an intensional interpretation, since a feature which can be perceived by sight –i.e. the colour– is selected.

3. 2. The suffix -ιάτικ(ος)

From a semantic point of view, we can ascertain that approximately 85% of the corpus consists of bases which are temporal nouns. However it is necessary to differentiate between two levels : the referential level, where the temporal axis is organised in calendar time, and the semantic level. The temporal nouns of the corpus are divided into two groups according to either the characteristic of consecutivity or that of non-consecutivity of the occurrences of the referents of a class. In the first group belong the names of units used for measuring time, eg. χρόνος 'year', μήνας 'month', εβδομάδα 'week', and in the second group their reanalyses which include two sub-

groups a) the special denominations of the units eg. Δευτέρα 'Monday', Ιανουάριος 'January', b) the denominations of the internal structure of units e.g. πρωΐ 'morning', άνοιξη 'spring'. In the following diagram, we can observe that at a linguistic level, the periods of time are not of equal importance. To be more specific, the framework of the 24 hours has a great importance for mankind, because of the social use of time (Taylor 1989).

Apart from these temporal nouns, the base of an adjective with the suffix –ιάτικ(ος) can be chosen from the list of names of the important holidays or spiritual celebrations. For these reasons we consider these words to be temporal nouns, eg. Πρωτομαγιά 'May Day', Πρωταπριλιά 'April Fool's Day', Πρωτοχρονιά 'New Year's Day', Πάσχα 'Easter', Χριστούγεννα 'Christmas', Άγιος Βασίλειος 'Santa Clauss', αποκριά 'carnival'. Finally, out of a total of approximately fourty bases of adjectives ending in -ιάτικος, only 6 do not have a clear semantic relationship with the meaning of time. These nouns are παιδ(ί) 'boy', γιορτ(ή) 'celebration', σκόλ(η) 'feast day, φεγγάρ(ι) 'moon',,γαμπρ(ός) 'groom', νύφ(η) 'bride'. We must first examine the meaning of the bases, since we can see a difference between the meaning of the two derived adjectives βραδιάτικος and βραδινός with the same noun base but a different suffix. If we make use of the distinctions between expert categories and 'folk' categories (Taylor 1989 : 72), we can assume that temporal nouns, such as καλοκαίρι 'summer', ημέρα 'day', Ιανουάριος 'January' have got the expert's definitions, meaning definitions from physicists and which are derived as a result of their classifications of the basis of necessary and adequate conditions. However, these nouns have also got folk definitions as a result of folk categories which are created around the prototypical elements which are created by the way in which people comprehend things around them and influence them (Taylor 1989 : 72).

In this way we can characterise the weather during the halcyon days in January as being 'spring' weather, although we well know that we should talk about winter if we want to be specific. While, according to the expert definition the winter is the time of the year which begins on the first of December and lasts until the 28th /29th of February, according to the folk definition we can call winter any time of the year which is judged as being extremely cold. Consequently, the concept 'winter' is dependent on the knowledge we already have about the winter. This knowledge is the basis of the representation of the winter in our mind and explains why winter, according to the folk definition, may differ from the 'winter' of the expert definition. It is, therefore, possible to view information about the semantic character of words as a net of cultural, conventional and up to a point idealised knowledge which is contained in a 'model' of ideas and practice.

For us, the suffix –ιάτικ(ος) is undoubtedly a morphological means used by MG to create adjectives which modify a noun with their prototypical or stereotypical features (Geeraerts 1985:30), which are derived from the folk definition of the referent of their base. In that way, we say ανοιξιάτικος καιρός 'spring weather' but εαρινή ισημερία 'spring equinox'. Consequently, different suffixes, when added to the same base,

assume different features from this base. For this reason, the words βραδιάτικος and βραδινός can not be synonyms.

I will now go on to make an assumption about the words in the remaining 15% of the corpus which, as we said before, do not seem to be temporal nouns. Provided that the suffix -ιάτικ(ος) is applied to bases which indicate time, we can assume that the application of -ιάτικ(ος) to these nouns converts them into temporal nouns, in the sense that the suffix -ιάτικ(ος) takes a meaning from these nouns which enables us to refer to time. In this way, for example 'ένα γαμπριάτικο κοστούμι' is the suit which carries the features associated with being a groom, and within the time indicated by the base. The same is true for 'νυφιάτικο τραγούδι' which refers to the bride within the time period which the wedding lasts ; 'παιδιάτικα καμώματα' which refers to childhood, γιορτιάτικος and σκολιάτικος, where the nouns celebration and leisure are able to place an event in time. Finally φεγγαριάτικος where the base φεγγάρι does not refer only to the natural satellite of the planet Earth but also to the moon's phases which are a result of its movement and which is believed to have an effect on the mental balance of human beings (eg. έχει τα φεγγάρια του 'he's in a really bad mood'). So, the word φεγγαριάτικος means 'mentally unbalanced' like a lunatic.

3. 2. 1. The Formation of the adverb

We can ascertain that the adverbs ending in -ιάτικα are formed from the base of adjectives ending in -ιάτικος, whose base is a noun relating to time. Consequently, the derived adverb 'inherits' a time related meaning from the derived adjective. Therefore, even though we admit that δευτεριάτικα and τη Δευτέρα are time adjuncts in utterances such as δευτεριάτικα άρχισε τη γκρίνια and τη Δευτέρα άρχισε τη γκρίνια 'On Monday he started complaining', it cannot be accepted that the two utterances are synonymous. This difference in meaning between time adverbials of this type and the corresponding construction with the definite article + temporal noun is reproducible, eg :

He ate meat during Easter week	?Μεγαλοβδομαδιάτικα νήστεψε
Μεγαλοβδομαδιάτικα έφαγε κρέας	?He fasted during Easter week
Τη Μεγάλη Βδομάδα έφαγε κρέας	
He woke me up, at lunchtime	?Κοιμήθηκα μεσημεριάτικα
Με ξύπνησε μεσημεριάτικα	?I slept at lunchtime
Με ξύπνησε το μεσημέρι	
He wore an overcoat in the summer	?Καλοκαιριάτικα φορούσε κοντομάνικο
Καλοκαιριάτικα φορούσε παλτό	?He wore an short-sleeved shirt in the summer
Το καλοκαίρι φορούσε παλτό	

From the above examples it becomes clear, I believe, that a) in each group the first two examples are not synonymous with each other, b) the third example is unnatural. I believe that the speaker who selects such an adverb intends to imply along with the temporal determination of the whole utterance, his objection to the representation as for what happens stereotypically during the time period to which the adverb refers. The speaker also wishes to indicate his expectations which are not satisfied as far as the cultural knowledge and the beliefs he has about that particular time period are concerned. Consequently, when the speaker's expectations are justified in the utterance, these adverbs have no place in the utterance.

3. 3. The suffix -ίστικ(ος)

Between the noun base and the modified noun there is a relationship of belonging to and, more specifically, we perceive the referential class which is indicated by the modified noun in a stereotypical way within the field of associations which has the referent of the noun base, eg. αγορίστικο παιχνίδι 'game for boys', which characterises boys, δωμάτιο κοριτσίστικο 'a room suited to girls', e.g. pink in colour, nicely decorated, etc.

We observe that the features, which have a relationship with the social norm and which are taken on by the noun base, lead to the concept of sterotypy. More specifically, as far as the bases of the adjectives ending in –ίστικος are concerned, preference is shown for nouns which include in their intension the semantic feature [+ animate], which orientates their reference towards the indication of persons or animals which are connected with certain types of behaviour, eg : καουμποίστικο καπέλο 'cowboy hat', μαϊμουδίστικα καμώματα 'monkey-like behaviour', παπαγαλίστικο διάβασμα 'rote learning'. Naturally, the position of the base can also be taken by a noun [-animate] provided that it refers to special human behaviour, eg : κουκλίστικη εμφάνιση 'pretty appearance', καρναβαλίστικη ατμόσφαιρα 'carnival atmosphere'. The suffix -ίστικ(ος) selects as its base, nouns whose referent symbolises a type of human behaviour. Following on from all the above, it is easy to see that a proper noun, a noun relating to a person or a place can be selected as a noun base. The referent of a noun related to a human name which plays the role of the base of an adjective ending in -ίστικ(ος) does not concern the original referent, meaning the person himself, e.g. Don Juan, although some features of this person are transferred to somebody else, so δονζουανίστικη συμπεριφορά of the person X is behaviour which has the specific characteristics of Don Juan, such as being charming towards women, and unstable in love, etc. Of course, one presupposition both in the choice of a (human) name as a base and in the understanding of the meaning of the relevant derived adjective is that both speakers have the same knowledge of the world. Finally, certain place names can appear as symbols of certain features, e.g. χολιγουντιανίστικος which refers to the mythical wealth of V.I.P.'s, and particularly of actors, who live in Hollywood.

Thanks to the adjective, the referential class which is indicated by the modified nouns becomes a type/model, since it shares certain features of the referent of the noun base, e.g. in contrast with the νομικό επιχείρημα 'legal argument', which is in accordance with legal science, the phrase νομικίστικο επιχείρημα is an argument which shares certain features with the former one, which are related to the behaviour of the person who uses them, in other words jurists, in their arrogance and quibbling. These stereotypical features are associated with the social norm, and consequently also have a pejorative meaning. The adjectives ending in -ίστικος, since they include a personal evaluation, appear in speech which is characterised by subjectivity and help to underline the ideological-cultural stance of the speaker and his pejorative attitude towards the referent of the modified noun, rather than identifying that referent. Contained with the adjectives ending in -ίστικος with a noun base [+human], a picture of the social norm is clearly projected which Modern Greeks have, regarding the general behaviour of a girl, a boy, a jurist, a philologist, a teacher, a priest, a greengrocer, a soldier etc., meaning discriminations between the sexes and between various professions. The adjectives which end in -ίστικος are traces of the strong presence of the speaker in his speech, and that's why they are not compatible with passages which demand objectivity, as in, for example, scientific texts. Conversely, they appear in the language of argument, eg : in opposing political speech and in informal register. It is worth noting that the noun bases of an adjective ending in -ίστικος can be marked as for the register, e.g. πουτάνα 'whore', but it can also be unmarked e.g. φιλόλογος 'philologist', παπάς 'priest', δάσκαλος 'teacher'. Whatever meaning it has however, the application of the suffix -ίστικ(ος) to these nouns has, as a result, the selection of those characteristics which show the stereotypical views of the speaker in relation to the social norm of behaviour for the referent of the noun base and, therefore, the speaker's evaluation of the referent of the noun base, and consequently his negative view of the modified noun.

3. 4. The suffix -(ι)άρ(ης)

With the aim of determining which adjectives are constructed using the suffix -(ι)άρ(ης), we can apply the principle of associativity to the corpus. Based on this principle, we exclude the word μπεκιάρης ('bachelor' < Turkish word 'bekâr'). Similar examples are the words Γενάρης (January), Φλεβάρης (February), which do not constitute constructed words. Moreover the adjectives ξεμαλλιάρης 'somebody with scruffy hair', ξεδοντιάρης 'toothless person' are not suffixed words, since they are adjectives formed with the prefix ξε- + the noun base μαλλί 'hair', δόντι(α) 'tooth/teeth'. The element -ιάρης, which is applied on the right, is analysed as a class marker which has the role of putting these adjectives into the referential class of adjectives which permanently give a negative feature to the modified noun which deviates from the norm in a way which becomes directly perceptible by the senses.

Another aspect of definig the range of this field is concerned with the relationship

between -άρης and -ιάρης . In MG Grammar the suffix appears in allomorphic variation , -άρης and -ιάρης . This unification is probably based on etymology, since the latin *-ari(u)s* led to -άριος, which in turn led to -άρης and -ιάρης, in which the -ι was considered part of the suffix, while it was the thematic vowel of the noun base (Hatzidakis 1905 : 421, 1907 : 527, 582, Petrounias 1991).

Therefore the question is posed whether it is a suffix with two allomorphs or two different suffixes. The answer which will be suggested is supported by semantic criteria. With the first approach, two semantic categories can be distinguished : apart from adjectives ending in -άρης derived from numbers e.g. δεκαπεντάρης 'fifteen year old', (a) adjectives ending in -άρης which at the surface structure appear as 'nouns denoting profession', e.g. αγελαδάρης 'cowboy', βαρκάρης 'boatman', (b) adjectives ending in -ιάρης used to describe people e.g. κοκαλιάρης 'skinny person', κιτρινιάρης 'sickly-looking person'. According to this classification, we could support that there are two suffixes in MG, -άρης and -ιάρης (see Christofidou 1990 : 73), if there were not few examples of 'profession' nouns ending in -ιάρης , without the -ι can be considered part of the base καρβουνιάρης 'charcoal seller', μαϊμουδιάρης 'someone who has performing monkeys', also few adjectives which describe people end in -άρης e.g. πεισματάρης 'stubborn'. Consequently, we consider that there is only one suffix. Nevertheless, the fact is clear that in MG (a) only the form -ιάρης is available to create adjectives which describe people, and (b) the form -άρης is not longer available for the formation of 'profession' nouns. Furthermore, the case of the suffix -(ι)άρ(ης) is interesting as it poses the problem of defining the limits of the morphemes, which is connected with the allomorphy but also with the origins of the morpheme, a topic which presents great interest for research.

Another problem relates to the grammatical category of the base to which the suffix -(ι)άρ(ης) is applied. According to the descriptions up until now, the base belongs to the category (a) of the noun e.g. γ(κ)ρινιάρης 'complainer', κοκαλιάρης 'skinny person', (b) of the adjective e.g. in αγαθιάρης 'good and naive person', κιτρινιάρης 'yellow person', or (c) of the verb e.g. κλαψιάρης 'somebody who cries', λησμονιάρης 'forgetful person'. In an attempt to unify the category of the base, we make the following proposal : the base of adjectives ending in –(ι)άρης belong to the category of the noun. In fact the majority of derivatives ending in -(ι)άρης fit onto the noun base without any problem. The derivatives from the third category can be deverbal nouns. As far as the grammatical category of words ending in –ιάρης is concerned, which always refer to people, and for that reason only have masculine or feminine forms, we suggest the following : denominal adjectives ending in -ιάρης, can be converted into nouns using the process of conversion[5]. The suffix -ικ(ός) can then be applied to the

[5] Conversion is a kind of derivational process which is very extensively used in MG and involves change of grammatical category and meaning without the addition of derivational means (suffix prefix) but only a inflexional one. It can take place before or after the application of a WFR of suffixation e.g. $\psi\iota\lambda\acute{o}\varsigma_A$ 'thin' → $\psi\iota\lambda\acute{\alpha}_N$ 'change'. Sometimes, the application of a conversion WFR

noun form to produce the adjective, as follows :

$[[[[\sigma\pi\upsilon\rho\acute{\iota}]_N\ (\iota\acute{\alpha}\rho(\eta\varsigma))_{suf}\ ]_A\ ]_N\ (\iota\kappa(\circ\varsigma))_{suf}]_A$ σπυρί 'spot'

As far as the semantic analysis of these adjectives is concerned, the suffix –(ι)άρ(ης) forms adjectives which establish a permanent relationship between the noun base and the noun which modifies the adjective in -ιάρης. The noun base is interpreted extensionally: ένας ψωριάρης άνθρωπος 'a scabby person' is a person who has ψώρα 'scabies'. If we examine, in particular, the noun base from a semantic point of view, we will see that it refers to (a) illnesses or symptoms of illnesses, human infections by their common name e.g. βλογιά 'smollpox', πανούκλα 'plague', χτικιό 'consumption', ψώρα 'scab', ψείρα 'louse', βήχας 'cough', σπυρί 'spot', but also the hyperonym αρρώστια 'illness', (b) to external imperfections on the body e.g. φακίδες 'freckles', or to one's character, which is expressed by behaviour which is considered socially unacceptable, e.g. γ(κ)ρίνια 'moaning', ζήλια 'jealousy', ξεχασιά 'forgetfulness', παράπονο 'complaining', (c) to entities or external characteristics which downgrade the appearance as different from the social norm, e.g. κουρέλι 'rags', λίγδα 'grease', βρόμα 'dirtiness'.

We believe that the negative connotation, or at least that which is below a socially acceptable norm, which accompanies the meaning of these nouns forms the connection between the 'professions' ending in -(ι)άρης and the adjectives ending in -(ι)άρης and which characterise the professions mentioned. These professions lack social prestige and are considered to be at the lowest levels of the social hierarchy of professions, e.g. μαϊμουδιάρης 'someone who has performing monkeys', σκουπιδιάρης 'dustbin man', βαρκάρης 'boatman' etc. However, can an adjective ending in -(ι)άρης still have negative connotations even though its noun base does not ? In this particular case, it is interesting to note the adjective κιτρινιάρης 'a yellow person'. The base of this adjective is the noun κίτρινο 'yellow', which is derived from the adjective κίτρινος 'yellow', the referential meaning of which is not negative, e.g. κίτρινο λουλούδι 'yellow flower'. However, the word κιτρινιάρης when used to refer to a member of the white race implies a deviation from the healthy colour as it characterises a sick person, which means we can assume that the suffix –(ι)άρ(ης) selects the negative side of the colour yellow in relation to the norm. Because of the negative connotations which are associated with words which are formed from the suffix –(ι)άρ(ης), many nicknames and, consequently, surnames are formed from this particular suffix. Furthermore, if we examine the base of the adjectives ending in –(ι)άρης from a lexicological point of view, we will find that 1) it belongs to [–learned] nouns cf. χτικιό 'consumption' in contrast with φθίση 'phthisis'/φυματίωση 'tuberculosis', αρρώστια 'illness' in contrast with the medical terms ασθένεια/νόσος 'disease', or that 2) in the case that there are doublets, the [–learned] type is selected, eg. διακονία 'diaconate' - διακονιά 'begging'

imposes an obligatory alternation of the stem vowel, providing that the stem is of Ancient Greek origin e.g. ανατέλλω 'rise (of the sun)' - ανατολή 'rising of the sun', βρέχω 'to rain' - βροχή 'rain', στρέφω 'to rotate' - στροφή 'rotation', τρέφω 'to feed' - τροφή 'food'.

→ διακονιάρης 'beggar' (Setatos 1969). Therefore, we expect these adjectives to be used in informal communication. Thus, the suffix –(ι)άρ(ης) forms adjectives which permanently attach a negative characteristic to the modified noun, so that it deviates from the norm in a way which can be perceived clearly by the senses.

4. Conclusion

As has been discussed in this contribution, the originality of the model consists of the following points: 1) Associativity, 2) Stratification, 3) Overgeneration, a model property according to which the derivational component generates all the constructed words of a language independently of whether these words are attested or not. The application of the theoretical principles of this model has allowed us to explain : A) How it is possible to have derived adjectives which consist of the same noun base but with different suffixes: each suffix takes different characteristics from the noun base. B) Why some suffixes do not apply to some nouns: in order to apply a suffix to a noun base, there should be compatibility between them. C) Why all the adjectives which could be formed with a suffix are not mentioned (possible words).

A detailed description of the MG derivational system, which would be of special interest not only to theoretical linguistics but also to psycholinguistics, to applied and computational linguistics, is not available. The most important feature of this model is that it contributes to a better understanding of the word formation mechanism.

REFERENCES

Anastassiadis-Symeonidis A. (1992) "Modern Greek Derivation" in *Studies in Greek Linguistics,* Thessaloniki, Aristotle University of Thessaloniki, pp. 505-526.

Christofidou A. (1990) "Literary neologism and productivity" in *Studies in Greek Linguistics,* Thessaloniki, Kyriakidis, pp. 477-492.

Corbin D. (1987) *Morphologie dérivationnelle et structuration du lexique,* Niemeyer, 1991[2] Villeneuve d'Ascq, Presses Universitaires de Lille, 2 vol.

Corbin D. (1991) "Introduction-La formation des mots : structures et interprétations", *Lexique* 10, pp. 7-30.

Geeraerts D. (1985) "Les données stéréotypiques, prototypiques et encyclopédiques dans les dictionnaires" in *Cahiers de Lexicologie* 46/1, pp. 27-43.

Hatzidakis G. (1905, 1907) *Mediaeval and Modern Greek,* Athens, 2 vol.

Petrounias E. (1991) "The suffix –αρία" in *Studies in Greek Linguistics,* Thessaloniki, Kyriakidis, pp. 53-69.

Setatos M. (1969) *The Modern Greek doublets,* Thessaloniki.

Tavlaridou E. (1996) *Etude du suffixe* –istik(os) *en grec moderne,* Master's Thesis at the University of Lille III (France).

Taylor J. R. (1989) *Linguistic Categorization-Prototypes in Linguistics Theory,* New York, Oxford University Press.

Section II
Allomorphy

Luigi Burzio
Johns Hopkins University
burzio@jhu.edu

Italian Participial Morphology and Correspondence Theory*

Abstract. It is argued that word-formation does not utilize a unique base, but rather that multiple lexical relations can be at work simultaneously. In Italian, a number of formations that are clearly based on the past participle, also exhibit direct links with the infinitive, upon which the participle itself is based. This multiplicity of relations can be adequately expressed within the 'parallel' organization of Optimality Theory, in which multiple constraints can all apply simultaneously. I propose in particular that 'surface-to-surface' faithfulness constraints are in general at work across all instances of the same morpheme ('Anti-allomorphy'). Such multiplicity of surface-to-surface dependencies is shown to hold for both stems and affixes.

1. Introduction. In this article I analyze Italian past-participial morphology from the general perspective of Optimality Theory ('OT'; Prince and Smolensky, 1993), and the more specific perspective of Burzio (1996, 1997, 1998a), in which lexical organization rests on the three sets of constraints in (1).

(1)			
	a.	Input-Output Faithfulness	(IO-F)
	b.	Phonology	(PHON)
	c.	Output-Output Faithfulness	(OO-F)

IO-F (1a) is understood in the standard sense of work in OT, as is (1b), which refers to the structural constraints of the Phonology at large. OO-F (1c) expresses the role of surface-to-surface association among lexical items, as in McCarthy (1995); Benua (1995, 1997), Burzio (1997, 1998a), and Burzio (1994a, b, 1996), where it appears under the name of (Metrical) Consistency, or Anti-Allomorphy.[1] OO-F constraints require identity of the target items with other items with which it is in 'correspondence' (McCarthy, 1995, Benua, 1995, 1997). I take such correspondents to be items with which the target independently shares sound and meaning, so that, for instance, *parent* is a correspondent for *parent-al*, though there is the complex question, not addressed here, of whether correspondence should be symmetric, with *parent-al* also serving as a correspondent to *parent*. As in Burzio (1996, 1998a), I take OO-F as sufficient to express word-to-word relations, hence dispensing with the traditional notion of 'underlying representation' (UR).

* Material closely related to the present work is presented in Burzio (1998b).

[1] OO-F constraints have been argued to derive the effects of the phonological cycle, as well as some of the cases of 'Non-derived Environment Blocking'. See Burzio (1998a, and reff.).

From this point of view, there will be no reason to expect that word-formation will employ a unique 'base' as in a derivational system. Rather, since there is no limit to the number of constraints that can apply simultaneously in OT, we expect that morphologically complex words could be calculated from multiple bases or 'correspondents' by simultaneous application of multiple sets of OO-F constraints. Indeed, I will argue in this article that, in Italian, certain formations that superficially appear to be 'de-participial' are in fact based on both the past participle and the infinitive simultaneously, as some of the data in (2) already indicate.

(2)	***Gloss(Infin)***	**Infinitive**	**Participle**	***-ore* Noun**	
a.	*adapt*	adatt-are	adatt-**at**-o	adatt-**at**-ore	(≈ partic.)
b.	*compress*	comprim-ere	compres-**s**-o	compres-**s**-ore	(≈partic.)
c.	*win*	vinc-ere	vin-**t**-o	vinc-**it**-ore	(≈both)
d.	*ascend*	ascend-ere	asce-**s**-o	ascen-**s**-ore	(≈ both)
e.	*aggress*	aggred-ire	aggred-it-o	aggres-**s**-ore	(≈ neither)

In (a), the agentive noun in *-ore*, cognate to English *-er* is transparently related to the participle as is the one in (b). The one in (c), however, has material from both the participle (the *t*), and the infinitive (the *c*). Similarly, the one in (d) has the *s* from the participle and the *n* from the infinitive, while the one in (e) is dissimilar from both infinitive and participle, and seems explainable only in terms of its correspondence with other *-ore* nouns, like the one in (b).

2. Metrically-conditioned syncopes. Past participles in Italian exhibit the two types of outcome illustrated for the three different conjugations in (3).

(3)	**Infinitive**	**Participle: non-syncopated**	**Participle: syncopated**
I.	**-áre**	associ-át-o, gener-át-o	
IIa.	**-ére**	cad-út-o, sap-út-o	
IIb.	**-ĕre**	vend-út-o, ricev-út-o, ced-út-o	vín-T-o, rí-S-o, discús-S-o
III.	**-íre**	scolp-íto, inib-ít-o	

With rare exceptions, participles in conjugations I, III are formed by adding to the infinitival stem the sequence - $\acute{V}$*t*-, where $\acute{V}$ is a stressed 'thematic' vowel. This morpheme is then followed by gender/ number inflection (*-o*: MASC.SG). The same is true for the variant (a) of conjugation II, that bears the stressed infinitival affix *-ére*. In contrast, in variant (b) of that conjugation, in which stress falls on the stem rather than the inflection, participles vary between the usual -$\acute{V}$*t*- and two syncopated outcomes: *-t-*, and *-s-*. As argued in DiFabio (1990), such syncopes can be accounted for in terms of Metrical Consistency, a form of OO-F. By removing the suffixal vowel and hence its associated stress, the main stress can fall on the stem, consistently with the infinitive, as

in *víncere/ vínto*. This account explains why participial syncope *only* occurs in the stem-stressed conjugation IIb. The oscillation between syncope and no syncope in that conjugation can be attributed to competition between metrical and segmental OO-F. In the non-syncopating cases, the affix *-út-* matches other such forms in the lexicon, hence satisfying OO-F relative to those forms. By the same token, one could take the same to be true of *-t-* or *-s-*, which would thus also satisfy OO-F. However, I take *-út-* to be the 'primary' allomorph, because more general, and hence the ranking in (4) to hold.

(4) OO-F/ **-út-** (primary allomorph) >> OO-F/ **-t-, -s-** (suppletives)

Hence the assumption is that it is more important to be faithful to the segmentism of *-út-* than to that of *-t-* or *-s-*. Now, although Italian allows antepenultimate stress, as in hypothetical **vínc-ut-o*, stem stress under such circumstances is precluded by the conclusion, drawn in Burzio (1998a), that outer affixes obey a higher-ranked OO-F than stems because they represent the 'head' of the word, determining its categorial status, as also shown by English vowel shortening, as for instance in *satyr-ize*, where the stem falls prey to shortening, (compared *satyre*), but the affix does not. I state this as in (5).

(5) OO-F/ affixal head >> OO-F/ stem (cf. satYr-**I:**ze)

This can be construed as a case of 'positional' faithfulness, in the sense of Beckman (1996), i.e. a case in which the rank of faithfulness constraints is modulated by the type of 'position' they affect. Returning to (3), each sequence *-Vt-* is the affixal head referred to by (5), and hence relatively immune to re-stressing. Given this, the oscillation of (3IIb) will now follow from taking metrical and segmental OO-F to be unranked with respect to one-another, as in (6).

(6)

	vínc- -út-o -t-	SYL	METRICAL OO- -út-	METRICAL OO- stem	SEGMENTAL OO-F -ut- >>-t-	SEGMENTAL OO-F stem
a. (☞)	vinc-út-o			*		
b.	vínc-ut-o		*			
c.	vínc-t-o	*			*	
d. ☞	vín-t-o				*	*

In (6), 'SYL' stands for all conditions on syllable structure, including those that exclude a complex coda *nc* in Italian. Given the ranking indeterminacy expressed by the vertical dotted line, candidates (a) and (d) will be co-optimal, while (b) will lose to (a) by the greater resilience of affixes stated in (5), and (c) will lose to (d) due to SYL. The choice

between (a) and (d) will thus be made lexically (by INPUT-OUTPUT FAITHFULNESS), in this case in favor of the syncopated candidate *vínto*, but in other cases in favor of the other candidate, as in *vendúto*. The choice between syncopated affixes *-s-* and *-t-* must also be viewed as lexical, to the extent that it is not fully predictable, as shown by minimal pairs like *assúm-ere/ assún-T-o* ('assume/-d'), versus *opprím-ere/ oppréss-S-o* ('oppress/-ed'); *pórg-ere/ pór-T-o* ('hand/-ed'), versus *emérg-ere/ emér-S-o* ('emerge/-d'). Note that the syncopated affixes *-s-*, *-t-* do not violate metrical OO-F despite their lack of stress compared with *-út-*. The reason is that metrical faithfulness is taken here to concern the association of stress with some segmental material, here the vowel *u*. Simultaneous disappearance of that vowel and its associated stress satisfies, rather than violate, metrical OO-F, while violating segmental OO-F. Note too that stem syncope must be prevented despite the ranking in (5) that would seem to favor it over affixal syncope. Intuitively, this must be to avoid unrecoverable distortions of the stem, as in hypothetically syncopated **v-úto*, in place of either *vin-t-o* or *vend-ut-o*. I put aside a formal account of this.

3. Syncope in derivatives. A number of formations from non-syncopated participles exhibit the variation illustrated in the (b-e) pairs in (7). These formations involve the suffix *-óre* of (2) above and suffixes *-ióne, -ívo, -úra, -ório/ -óio*, all of which have transparent English cognates (*-óio*, a variant of *-ório* '-ory').

(7) **Derivatives from non-syncopated participles:**

		Gloss(Infin)	**Infinitive**	**Participle**	**Derivatives**
-áre	a	*generate*	gener-áre	gener-át-o	gener-AT-óre
-ére	b	*detain*	deten-ére	deten-út-o	deten-T-óre
	b'	*contain*	conten-ére	conten-út-o	conten-IT-óre
-ĕre	c	*exceed*	eccéd-ere	eccéd-út-o	eccés-S-ívo
	c'	*sell*	vénd-ere	vend-út-o	vend-IT-óre
-íre	d	*sculpt*	scolp-íre	scolp-ít-o	scul-T-óre
	d'	*abolish*	abol-íre	abol-ít-o	abol-IT-óre
	e	*scan*	scand-íre	scand-ít-o	scan-S-ióne
	e'	*prepare*	imband-íre	imband-ít-o	imband-IT-óre

All cases in the rightmost column in (7) utilize a participial affix, given in upper-case. However, the first member of each pair in (b-e) features a syncopated affix despite the fact that the participle itself does not. This second case of syncope, unlike the first, now concerns all conjugations except the one in *-áre* of (7a), which I put aside for the moment. Yet the earlier account will extend to these new cases as well, as can be seen in (8).

(8)

aggred-ít -s-	-óre	SYL	METRICAL OO-F -óre	METRICAL OO-F aggred-ít-	SEGMENTAL OO-F -ore	SEGMENTAL OO-F aggred-it- >> aggred-s-
a. (☞)	àggred-it-óre			*		
b.	àggred-ít-ore		*			
c.	àggred-s-óre	*				*
d. ☞	àggres-s-óre					* *

In (8), for each type of faithfulness, metrical and segmental, again unranked with respect to one-another, we rank the one for the outer affix, here *-óre*, above that of the stem, here *aggred-ít-*, in accordance with (5) above. Unlike participial affixes, neither *-óre* nor the other derivational suffixes in question have syncopated suppletives –a lexical matter. But the participial suffixes themselves are relevant here as well. With unstressed *-ore* of (8b) excluded by the inalterability of outer affixes sanctioned by (5) above, and stress clashes excluded by undominated metrical constraints, the participial affix will have to be unstressed, as in any of (a), (c), (d), but excluding (b). Now the candidate in (a) will satisfy segmental OO-F to the primary participial allomorph *-ít-*, but violate metrical OO-F by featuring an unstressed *i*. Note that *-ít-* is no longer under the scope of (5) here, since it is not the head of the word. In contrast, the syncopated candidates in (c, d) will violate segmental OO-F, but satisfy metrical OO-F --the *raison d'être* of all syncopated suppletives. Finally, candidate (c) will lose to (d) due to SYL, which excludes this kind of coda in Italian. Hence candidates (a, d) will be co-optimal, and the alternations of (7d/d'; e/e') will reduce to the one of (3IIb) above. The alternations in (7b/b', c/c') are rather similar. The syncopated variant satisfies metrical OO-F as before, while the non-syncopated variant satisfies segmental OO-F, though only with respect to participial suffix *-ít-* of the *-íre* conjugation. OO-F to *-ít-* of the -ére/-ëre conjugation is in fact violated. This effect is the 'Anti-allomorphy' of Burzio (1996), and the 'Lexical conservatism' of Steriade (1997), describable as the attraction by items within the same general paradigm. So, when forced to deviate from the form *-ŭt-*, the items in (7b',c') adopt the form *-ít-*, independently existing with the items in (7e'), rather than creating a new unstressed allomorph **-ut-*, which is thus avoided altogether in the language. This is in fact quite parallel to the borrowing of syncopated affixes *-t-*, *-s-* by the items in (7b, c, d, e). These affixes exist in participles only in the *-ĕre* conjugation, as we saw in (3) and are 'imported' from that conjugation here. The various patterns of affixal consistency will be summarized in (13) below. The syncope occurring in the derivatives in (7) is thus to avoid re-stressing segmental material in the stem, and thus just like the syncope occurring in the participles in (3). The latter occurred only with *-ĕre* verbs because only those have stressed infinitival stems. The

syncope of the derivatives is more general because all conjugations have stressed participial stems. The only conjugation that is altogether immune to syncope is the first, in *-áre*. Its immunity to segmental allomorphy is more general, however, as shown by the comparisons in (9).

(9)				
	I.	manco/man**ch**iamo	[k]	*'I-am-missing/we-are-missing'*
		lan**c**io/lan**c**iamo	[č]	*'I-launch/we-launch'*
		volo/ voliamo	[l], [o]	*'I-fly/we-fly'*
		tag**li**o/tag**li**amo	[λ]	*'I-cut/we-cut'*
		suono/**suo**niamo	[wo]	*'I-sound/we-sound'*
	IIa.	vin**c**o/vin**c**iamo	[k/č]	*'I-win/we-win'*
	IIb.	**vuo**le/**vogli**amo	[o/wo], l/λ]	*'he-wants/we-want'*
	III.	fu**gg**o/fu**gg**iamo	[gg/ǧǧ]	*'I-flee/we-flee'*

As shown in (9), conjugations IIa,b, and III each exhibit segmental alternations of various sorts beside syncope, specifically palatalizations and diphthongizations, while conjugation I does not, each stem maintaining an invariant form. To account for this (and similar resistance to allomorphy by this conjugation in other Romance languages), it seems necessary to single out this conjugation in terms of a higher-ranked (segmental) OO-F, a fact attributed in Burzio (1997) to the much larger size of this conjugation compared with the others.

4. Syncope revoked. Participial derivatives can deviate from their participial bases not only in being syncopated unlike their participles, as shown in (7) above, but also in the opposite way, as shown in (10).

(10)	***Gloss(Infin)***	**Infinitive**	**Participle**	**Derivatives**
a.	*convince*	convínc-ere	convín-T-o	còn**v**in-Z-ióne
a'.	*win*	vínc-ere	vín-T-o	vinc-IT-óre
b.	*add*	aggiúng-ere	aggiún-T-o	àggiun-Z-ióne
b'.	*collect*	raccógli-ere	raccól-T-o	racc**ò**gl-IT-óre
c.	*disperse*	dispérd-ere	dispér-S-o	dìsper-S-ióne
c'.	*lose*	pérd-ere	pér-S-o	pèrd-IZ-ióne

In (10), the first member of each pair of derivatives maintains the segmentism of its participle, which is syncopated. This is putting aside the assibilation that affects *t* before *i* in hiatus, yielding t^S (orth. *z*). The second member, however, does not maintain that segmentism, utilizing instead a non-syncopated participial affix *-it-* (*/-it^S-*). Therefore, in participle-derivative pairs, syncope gives rise to four patterns, by being present in either participle or derivative, neither one, or both. The variation in (10) can again be reduced to the familiar tension between metrical and segmental faithfulness, as shown

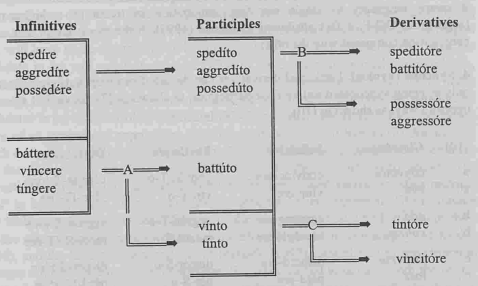
in (11).

(11)

vínt- -óre	METRICAL OO-F		SEGMENTAL OO-F	
	affix	stem	affix	stem
a. (☞) vint-óre		*		
b. vínt-ore	*			
c. ☞ vinC-IT-óre				*

Candidate (b) in (11) is excluded for the same reasons as the one in (8) above: outermost affixes do not re-stress, as mandated by (5) above. Candidate (a) violates metrical OO-F to the participle by de-stressing the stem (no stress-clashes), while candidate (c) satisfies metrical OO-F thanks to the inserted syllable, which, however, causes a violation of segmental OO-F. Hence candidates (a, c) are co-optimal as in previous cases and the choice is made again lexically, whence the variation in (10). I turn to the nature of the inserted material shortly. The diagram in (12) recapitulates the infinitive-participle and participle-derivative relations we have so far seen.

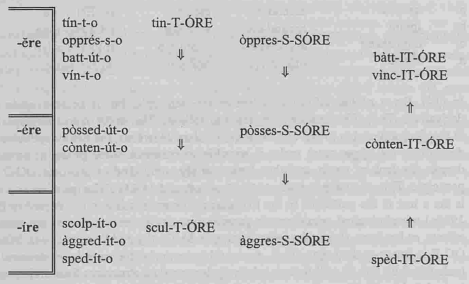
(12)



In (12), the infinitives in the lower box are in *-ĕre*, and their participles bifurcate at point A into syncopated and not. The infinitives in the upper box are from other conjugations, and their participles do not syncopate, as we saw. In going from non-syncopated participles (upper box) to their derivatives, there is a bifurcation at point B, into -again - syncopated and not. Similarly, in going from syncopated participles to their derivatives, there is also a bifurcation into syncopated and not, at point C. I have argued that the three bifurcations in (12) are all of the same nature: the upper branch maintains the segmentism of the stem, while the lower one avoids re-stressing the stem.

Let us now turn to the segmental material inserted in cases like (11c), given in upper-case. The sequence *-it-* is clearly the participial affix of other cases, indigenous to the third conjugation, but adopted in its unstressed variant by other conjugations as well, thus limiting metrical allomorphy to *-ŕ-*, *-ít-*, and avoiding unstressed *-ut-* as noted. (Recall that unstressed *-it-* becomes possible when further derivational affixes follow, thus demoting it from 'head'). The overall distributional pattern of participial affixes is as reconstructed in (13).

(13)



The *-ĕre* conjugation has syncopated participles like *tinto* and *oppresso*, which give syncopated derivatives like *tintore* and *oppressore* by segmental faithfulness. The other two conjugations only have non-syncopated participles but these can still give syncopated derivatives by metrical faithfulness. When the segmentism of these cases thus breaks away from that of their participle, it falls in with the affixal segmentism *-t-ore*, *-s-ore* that independently exists, as indicated by the downward arrows. At the same time, other derivatives from *-ĭto* participles are segmentally consistent with those participles rather than being syncopated, yielding affixal sequences *-it-ore*, *-iz-ione*, *-it-ivo* etc. As indicated by the upward arrows in (13), these sequences are utilized by derivatives of participles in *-út-* from the *-ére* and *-ĕre* conjugations, as an alternative to the syncopes, thus altogether avoiding unstressed *-ut-*. Derivatives of syncopated participles like *vinto*, whose segmentism breaks away from the participle for metrical reasons, also find this independently available segmentism well-suited, whence *vinc-it-ore*, etc. Hence affixal correspondence/ faithfulness is pervasive: whenever affixal material is driven into allomorphy, recourse is had to independently existing patterns, even if this crosses boundaries between the conjugations, otherwise segregated systems by definition.

This leaves us with the *c* of *vinC-it-ore*, present in the derivative in (10a') despite its absence in the participle. The source for it is obviously the infinitive *vinC-ere*, revealing that both participle and infinitive simultaneously serve as bases for the derivatives.

Similar considerations hold for (10b') *raccoGL-it-ore* (*gl* = {λ}), and (10c') *perD-iz-ione*. Although the participle is the primary base, when the derivative strays from participial segmentism under compulsion from metrical OO-F, the infinitival segmentism comes in as next best, revealing the multiple correspondence.

Links with the infinitive are revealed as well by cases like (14).

(14)	***Gloss****(Inf.)*	**Infinitive**	**Participle**	**Derivatives**
a.	*ascend*	ascénd-ere	ascé-s-o	àsCEN-s-óre
b.	*ignite*	accénd-ere	accé-s-o	àcCEN-s-ióne
c.	*apprehend*	apprénd-ere	appré-s-o	àpPREN-s-ívo
d.	*defend*	diténd-ere	dité-s-o	diFEN-s-óre

In (14), the *n* of the infinitive is lost in the participle, but is present again in the derivatives. This effect can be interpreted as follows. The stress of the participle is lost in the derivatives due to the ban on stress clashes. However, the heavy syllables of the derivatives (in upper-case) make it possible to maintain some prosodic prominence despite the lack of stress, and thus better satisfy some kind of prosodic OO-F to the participle. This of course is at the expense of segmental OO-F. The question now is why is the *n* lost in the participle in the first place. The answer is that (as argued in Burzio, 1998a) morphological operations in general give rise to 'emergence of the unmarked' effects, as has been shown for reduplication in McCarthy and Prince (1994). Whenever OO-F is lower-ranked than IO-F, as would seem to be the case here, derived words (calculated by OO-F) will feature relatively less marked structures than underived ones (calculated by IO-F), another case in point being English vowel length, which falls prey to markedness only in derived environments, e.g. *vitamin/ divinity* (see Burzio, 1998a). Hence, in the formation of the participles in (14), the closed syllable turns into a less marked open one, a change that does not affect the preservation of the stress, since Italian allows stress on open penultimates. However, in the formation of the derivative, participial stress is lost, and prosodic prominence can only be maintained by making the syllable heavy again.[2] The point is that, when extra material is needed, it is the infinitive, rather than some general epenthetic process that supplies it, both in (11) and in (14), revealing the *double* correspondence of the derivatives, with both the participle and the infinitive.

In sum, derivatives whose participial base is syncopated vary between maintaining the syncopated form of the participle and revoking the syncope by inserting extra material. Such variation reduces to the usual tension between metrical and segmental OO-F, resolved lexically. The nature of the material inserted reveals a cross-derivational relationship to other derivatives, and a trans-derivational relationship to the infinitive.

5. Conclusion. Participial morphology in Italian varies between a non-syncopated stressed form -*V́t-* and two syncopated forms *-t-*, *-s-*. The pattern of variation reveals a complex network of relations among lexical items which defies both traditional work based on sequential derivations, as well as work that utilizes OO-F within OT more conservatively than in the present work, by attributing to morphologically complex words unique bases. What the above evidence suggests is that words can in principle influence each-other's sound structures whenever they are independently similar in content and regardless of whether the similarity concerns stem or affixes. The question that this raises, important but beyond the scope of this article, is of course how to determine from principle the weight that each relation holds (i.e. the rank of OO-F that it imposes) in the calculation of sound structure.

Beside the relations discussed above, other similar ones appear to exist as well. One is a relation between participles and preterits. Like the participles, preterits of *-ĕre* verbs vary between syncopated and not, though they only syncopate in *-s-*, as in *vĭnc-ere/ vĭn-s-i* 'to-win/ I-won', versus non-syncopating *vénd-ere/ vend-é(tt)-i* 'to-sell/ I-sold'. As with the participle, syncope in the preterit maintains the stem stress of the infinitive, and is thus amenable to the familiar analysis. Interestingly, while the variation in the preterit is in itself unpredictable, like that of the participle, the correlation within participle-preterit pairs is near-perfect, as shown in (15), revealing OO-F at work within such pairs.

(15)	***Gloss****(Inf.)*	**Infinitive**	**Participle**	**Preterit**
a.	*hide*	nascónd-ere	nascós-T-o	nascó-S-i
a'.	*sell*	vénd-ere	vend-út-o	vend-é(tt)-i
b.	*write*	scrív-ere	scrít-T-o	scrís-S-i
b'.	*receive*	ricév-ere	ricev-út-o	ricev-é(tt)-i
c.	*laugh*	ríd-ere	rí-S-o	rí-S-i
c'.	*yield*	céd-ere	ced-út-o	ced-é(tt)-i
d.	*put*	métt-ere	més-S-o	mí-S-i
d'.	*beat*	bátt-ere	batt-út-o	batt-é(tt)-i
e.	*discuss*	discút-ere	discús-S-o	discús-S-i.
e'.	*repeat*	ripét-ere	ripet-út-o	ripet-é(tt)-I

[2] Since in (14) the derivatives violate segmental faithfulness to the participle for the sake of some prosodic faithfulness, and since we know this in general to be an even trade-off, we will expect the opposite outcome as well. The pattern in (i) confirms this expectation.

(i)	**Infinitive**	**Participle**	**Derivative**
	diffónd-ere '*diffuse*'	diffú-s-o	diffu-s-óre

In (i), and similarly with other verbs based on *fond-ere* 'fuse', the derivative is segmentally, rather than prosodically, faithful to the participle.

f.	*oppress*	opprim-ere	opprés-S-o	opprés-S-i.
f'.	*press*	prém-ere	prem-út-o	prem-é(tt)-i

Another relation appears to exist among derivatives of the same participle. We have seen that derivatives from non-syncopated participles may or may not syncopate. However, if one does, all do, as shown by the contrast in (16).

(16)	***Gloss(Inf.)***	**Partic.**	**Derivatives**
a.	*send*	spedito	sped-it-ore, sped-iz-ione, sped-it-ivo
b.	*aggress*	aggredito	aggres-S-ore, aggres-S-ione, aggres-S-ivo

Furthermore, we have seen that syncope varies between *-t-* and *-s-*, rather unpredictably. However, the choice is completely consistent across derivatives of the same participle, as shown by (17) and (16b).

(17)	***Gloss(Inf.)***	**Partic.**	**Derivatives**
a.	*adhere*	aderito	ade-S-ore, ade-S-ione, ade-S-ivo
b.	*assert*	asserito	asser-T-ore, asser-Z-ione, asser-T-ivo

These facts reveal the presence of OO-F across derivatives. The overall network of lexical links thus identified is summarized in (18).

(18) **Network of correspondences:**

a. **Derivatives of the same participle are cross-linked**: (16), (17)
b. **They are linked to both participle and infinitive**: (10)-(11), (14).
c. **Participle and preterit of the same verb are cross-linked**: (15).
d. **Derivatives of different verbs are cross linked**: (12).

References

Beckman, J. (1996) *Positional Faithfulness*, PhD Dissertation, University of Massachusetts, Amherst.

Benua, L. (1995) "Identity Effects in Morphological Truncation," in Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk, eds. *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, Amherst: GLSA.

Benua, L. (1997) *Transderivational Identity: Phonological Relations between Words*, PhD Dissertation, U. Mass. Amherst.

Burzio, L. (1994a) *Principles of English Stress*, Cambridge University Press.

Burzio, L. (1994b) "Metrical Consistency," in E.S. Ristad, ed., *Language Computations*, American Mathematical Society, Providence RI.

Burzio, L. (1996) 'Surface Constraints versus Underlying Representation,' in Durand, Jacques & Bernard Laks, eds. *Current Trends in Phonology: Models and Methods*. European Studies Research Institute, University of Salford Publications (123-141).

Burzio, L. (1997) 'Strength in Numbers', in V.Miglio and B. Morén, eds. *University of Maryland Working Papers in Linguistics 5 (1997): Selected phonology papers from H-O-T 97*, 27-52.

Burzio, L. (1998a) 'Cycles, Non-Derived-Environment Blocking, and Correspondence,' ms. Johns Hopkins University. To appear in Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer, eds. *Optimality Theory: Syntax, Phonology, and Acquisition*. Oxford University Press.

Burzio, L. (1998b) 'Multiple Correspondence', *Lingua 103*, 000-000

DiFabio, E. (1990) *The Morphology of the Verbal Infix /-isk-/ in Italian and in Romance*." Doctoral Dissertation, Harvard.

McCarthy, J. (1995) "Extensions of Faithfulness: Rotuman Revisited," ms. U.Mass, Amherst.

McCarthy, J. and A. Prince (1994) "The Emergence of the Unmarked: Optimality in Prosodic Morphology," in Mercè Gonzàlez, ed. *Proceedings of the North East Linguistic Society 24*. Amherst, MA: Graduate Linguistic Student Association. 333-379.

Prince, A. and P. Smolensky (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*, ms., Rutgers University, New Brunswick, and University of Colorado, Boulder. To appear, MIT Press.

Steriade, D. (1997) 'Lexical conservatism' in SICOL, 1997: *Linguistics in the Morning Calm*, Hanshin, Seoul, Korea.

Lynne J. Cahill and Gerald Gazdar
University of Sussex
Email: {Lynne.Cahill,Gerald.Gazdar}@cogs.susx.ac.uk

ALLOMORPHY IN PolyLex[1]

Abstract

The PolyLex project aims to produce a hierarchical multilingual lexicon for Dutch, English and German, in which information common to more than one language is inherited from a shared component. The PolyLex work done to date has concentrated on the morphology and morphophonology of the three languages. In this paper we present the morphological framework used in PolyLex with examples of the ways in which allomorphic variation is handled.

1. Introduction

Our general approach to inflectional morphology[2] falls within the tradition that treats paradigms (inflectional classes, declensions, conjugations, etc.) as analytically central[3] rather than epiphenomenal or of secondary status[4]. The central notion is the lexeme, not the word or the morpheme. Words exist, but only as *realizations* of (morphosyntactic specifications of) lexemes – hence Stump's use of the term *realizational* to characterize this tradition. Morphemes also exist, but only as second class citizens. The appearance of a morpheme is just one among several ways that morphosyntactic information gets expressed in the realization of a lexeme as a word (cf. Wurzel 1990, 208-209). And we share Zwicky's view that "all realization rules are treated as expressing **defaults**, which are automatically overridden by more specific rules (and these in turn by still more specific rules, and so on)" (1985, 372).

As regards current work, our approach is closely related to Corbett & Fraser's Network Morphology[5] and the most recent version of Stump's Paradigm Function Morphology (forthcoming). In our approach, unlike those of Stump and Corbett *et al.*, abstract inflectional rules are typically stated in terms of phonological units, most commonly the syllable and the segment (as in Cahill 1990a, 1990b, 1993). Gibbon and his collaborators in the ILEX (Integrated Lexicon with EXceptions) project at Bielefeld[6] have pioneered the use of default inheritance hierarchies for the representation of lexical phonology and morphophonology. Our work is thus also indebted to theirs.

2. The DATR language

The PolyLex lexicons are implemented in the lexical knowledge representation language DATR (Evans & Gazdar 1996)[7]. DATR is a rather spartan nonmonotonic language for defining inheritance networks with path-value equations. The development of DATR was guided by a number of concerns which we summarise here. The objective was to design a language which (i) has an explicit theory of inference, (ii) has an explicit declarative semantics, (iii) can be readily and efficiently implemented, (iv) has the necessary expressive power to encode the lexical information presupposed by work in the unification grammar tradition, and (v) can express all the evident generalizations and subgeneralizations about such entries. In keeping with its intendedly minimalist character, it lacks many of the constructs embodied either in general purpose AI knowledge representation languages or in contemporary grammar formalisms. The language is nonetheless sufficiently expressive to represent concisely the structure of lexical information at a variety of domains of language description.

It should be stressed that DATR itself is no more than a very general **language** for lexical description and therefore does not commit or restrict the linguist using it to any particular linguistic framework, theory or formalism, nor is it restricted in the class of natural languages that it can be used to describe. Clearly, it is well suited to lexical frameworks that embrace or are consistent with inheritance and non-monotonicity through networks of nodes, but these are not requirements. DATR can be (and has been) used to implement differing theoretical approaches (including ILEX, HPSG, Word Grammar, LTAG, Finite State Morphology, Network Morphology, Paradigm Function Morphology), and is perhaps best thought of as a programming language which can be used to implement and test linguistic theories. Indeed, it would not be entirely misleading to think of DATR as a kind of assembly language for constructing (or reconstructing) higher level theories of lexical representation. Unlike most other formal languages proposed for lexical knowledge representation, DATR is also not restricted in the domains of linguistic description to which it can sensibly be applied. It is designed to be equally applicable at phonological, orthographic, morphological, syntactic and semantic domains of description. But it is not intended to replace existing approaches to those domains. DATR cannot be (sensibly) used without a prior decision as to the theoretical frameworks in which the description is to be conducted; there is thus no 'default' framework for describing, say, morphological facts in DATR.

In DATR, information is organised as a network of **nodes**, where a node is essentially just a collection of related information. In the context of lexical description, a node might correspond to a phoneme, a syllable, a morpheme, a word, a lexeme, etc., or a class of such items. For example, for German, we might have

[1] This work was supported by ESRC research grant *Multilingual lexical knowledge representation*, number **R000235724**, to Gazdar & Cahill.

[2] Described in more detail in `http://www.cogs.susx.ac.uk/lab/nlp/polylex/polylex.html` Cahill and Gazdar (1997, forthcoming).

[3] As in the work of Matthews (1972), van Marle (1985), Zwicky (1985, 1990), Carstairs (1987), and Stump (e.g., 1992; 1993a; 1993b; 1993c; 1995).

[4] Thus, for example, inflectional class is a secondary notion for Wurzel (1990, 204): for him it is the citation form that determines the inflectional class, not the converse.

[5] See Brown *et al.* (1996), Brown & Hippisley (1994), and Fraser & Corbett (1995; in press) for work in this framework.

[6] See Bleiching (1992; 1994), Bleiching et al. (1996), Gibbon (1990; 1992), Gibbon & Bleiching (1991), Reinhard (1990) and Reinhard & Gibbon (1991) for examples of this work.

[7] See also `http://www.cogs.sussex.ac.uk/lab/nlp/datr/datr.html`

a node describing an abstract `Word`, a node for the class of nouns, a node for the subclass of nouns that mark plurals with *-s*, a node for the particular noun lexeme `Klub` ('club') and still more for the individual words that are instances of this lexeme *Klub*, *Klub-s*. Each node has associated with it a set of equations that define partial functions from **paths** to **values** where paths and values are both sequences of **atoms** (which are primitive objects). Atoms in paths are sometimes referred to as **attributes**. The syntax and terminology of DATR, like its name and its minimalist philosophy, owes more than a little to that of the unification grammar language PATR (Shieber 1986).

3. Phonology

Our interest in phonology in the PolyLex project is restricted to those aspects of phonological structure that are relevant to the description of inflection in the languages considered. Those aspects include syllable structure but do not include any structure above the level of the syllable, such as metrical structure.

We also restrict ourselves to a segmental representation of the phonology. Our phonological segment inventory is taken from CELEX (Baayen et al., 1995) and uses the SAMPA machine-readable phonetic alphabet (Wells, 1987). As one of us has shown in earlier work (Cahill 1993), the step from representing structures with segments to representing the same structures with full feature sets at each point in the tree is relatively simple. We have not taken that step here because it would not add anything to most of the present analysis but it would make our DATR code much harder to read. However, a featural encoding would permit a more elegant treatment of phonological alternations such as final consonant devoicing and morphophonological alternations such as vowel lengthening and umlaut.

As in Cahill (1990b) and Bleiching (1992), we define syllabic structures by means of simple context-free phrase structure rules:

syllable → *onset rhyme*
rhyme → *peak coda*
coda → *body tail*
disyllable → *syllable syllable*
trisyllable → *syllable syllable syllable*

A syllable consists of an onset and a rhyme; a rhyme consists of a peak and a coda; and a coda consists of a body and a tail[8]. A disyllable consists of two syllables, and a trisyllable of three. We can express these in DATR as follows[9]:

```
Syllable:
    <phn $yll form>  == "<phn $yll onset>" "<phn $yll rhyme>"
    <phn $yll rhyme> == "<phn $yll peak>"  "<phn $yll coda>"
    <phn $yll coda>  == "<phn $yll body>"  "<phn $yll tail>"
    <phn root>  == <phn syl1>
    <> == Null.

Disyllable:
    <> == Syllable
    <phn root>  == <phn syl2> <phn syl1>.

Trisyllable:
    <> == Syllable
    <phn root>  == <phn syl3> <phn syl2> <phn syl1>.
```

This rule schema makes crucial use of a variable `$yll` that ranges over attributes (`syl1, syl2, ..`) that denote syllable positions. Note also that the maximally unspecified path (`<>`) at the `Syllable` node is defined by reference to `Null` which always returns the empty sequence as its value. An `<onset>`, `<peak>` or `<coda>` which is left undefined at lower levels of the hierarchy will, as a consequence, end up as null.

The definitions of di- and trisyllables number the syllables from the right. This is a language-specific aspect of our analysis and reflects the fact that Dutch, English and German morphology all primarily involve suffixation. Reference to final syllables is thus more frequent than reference to the initial syllables and it is technically convenient to have a constant identifier (`syl1` here) for final syllables.

Given this set of axioms for syllabic structure, we can now use them to help define particular concrete (poly)syllables. Here, for example, is a possible definition for the monosyllabic *-es* suffix, realized phonologically as `/@s/`.

```
Suffix_es:
    <> == Syllable
    <phn syl1 peak> == @
    <phn syl1 coda> == s.
```

Likewise, a disyllabic word root such as the German *Tutor* can be specified in terms of the individual components of its two syllables[10]:

```
Tutor:
    <> == Noun_L
    <phn root form>  == Disyllable
    <phn syl2 onset> == t
    <phn syl2 peak>  == u:
    <phn syl1 onset> == t
    <phn syl1 peak>  == O
    <phn syl1 tail>  == r.
```

[8] The tail of a coda is its final segment and the body consists of any remaining consonants in the coda. This simplifies reference to final consonants of roots.

[9] We have simplified and/or modified the DATR code from the actual PolyLex lexicons whenever this has looked likely to enhance the readability of the present paper and assist us in making the points at issue. We have also spared our readers the many pedantic footnotes that would be required to document every case of such code editing.

[10] Default information for a lexeme node like this comes from the declensional class node, in this case, `Noun_L`.

From these node definitions, taken together with the axioms for syllable structure given above, we can now infer that:

```
Suffix_es:
    <phn root form> = @ s.

Tutor:
    <phn root form> = t u: t @ r.
```

4. The representation of allomorphy

Within this framework, there are two principal methods for representing allomorphy: (i) the use of path extensions on the left hand side of equations and (ii) the use of conditional statements on the right hand side of equations. These two methods can also be combined. In discussing the applicability of these two approaches, we make a distinction between the variant and inherent properties of a class of lexemes: nouns, for example, have gender as an inherent morphosyntactic property whilst case and number are variant morphosyntactic properties. The variance or inherence of a property is relative to the class of lexemes involved, thus adjectives, for example, have gender as a variant property, not an inherent one[11].

4.1. Path extensions

When querying the form of a word, a query path is invoked that is partly composed of attributes representing the particular values of the variant morphosyntactic properties of the lexeme involved. So, to find the form of the genitive singular of a noun, for instance, the query path would be `<mor word sing gen>`. The morphological word is defined, by default, as a root followed by a (possibly null) suffix. The `Word` node, from which all word class nodes and ultimately all words inherit by default, thus appears as follows:

```
Word:
    <> == Syllable
    <mor word> == "<phn root form>" "<mor suffix>".
```

Given this definition, the query path `<mor word sing gen>` leads to the phonological form query (`<phn root form>`) having the variant morphosyntactic attributes appended, so the query path for the root is `<phn root form sing gen>`. This allows us to define realizations which are contingent on variant morphosyntactic properties by specifying the relevant attributes in appropriate path equations as follows:

```
Noun_L:
    ...
    <phn syl1 peak plur> == Lengthen:<"<phn syl1 peak>">.
```

[11] Note that we are making the distinction with respect to classes of lexemes, not individual lexemes. There is a sense in which the noun lexeme *trousers* is inherently plural, but that sense is not to the point here.

which says that if the feature `plur` is present in the query path then the peak is realized by application of the `Lengthen` function.

Several examples of this kind of allomorphy can be found in the three PolyLex languages. In one class of Dutch nouns the stem vowel in the plural form is always `/e:/`, regardless of what vowel the singular form has, e.g. *stad/steden, lid/leden*. This is captured in PolyLex in the following manner:

```
Noun_e:
    ...
    <phn syl1 peak plur> == e:.
```

English nouns which have a final voicing alternation, such as *wife/wives, house/houses* can be accounted for in a similar way, the realization of their final coda being dependent on whether the form is singular:

```
Noun_D:
    ...
    <phn syl1 coda sing> == Devoice:<"<phn syl1 coda>">.
```

This is just a restricted application of final consonant devoicing, something which applies more generally in German and Dutch.

German umlaut is the classic example of this type of alternation, and is interesting in the present context because of the fact that the relevant morphosyntactic property differs in nouns and verbs. In German nouns which belong to one of the declensional classes which undergoes umlaut, the umlaut function applies only in the plural forms:

```
Noun_U:
    ...
    <phn syll peak plur> == Umlaut:<"<phn syll peak>">.
```

However, in one class of verbs, the (relevant) vowel undergoes umlaut in past tense forms:

```
Verb_U:
    ...
    <phn syll peak past> == Umlaut:<"<phn syll peak>">.
```

All three languages exhibit this type of allomorphic variation in their numeral forms, with variation between, for example, *twee/twin-, two/twen-, zwei/zwan-*. In our account of the numerals expressions (Cahill & Gazdar, 1996) we capture this alternation by the use of morphosyntactic features to indicate the "teen" and "ty" forms of the numerals. Thus, the form of a numeral combined with either "teen" or "ty" is marked with an attribute `bound` that encodes a variant morphological property of morphemes (free/bound). Given this attribute, the variation in forms can be stated as follows:

```
Phon002:
    <> == Syllable
    <M phn body bound>  == n

    <D phn onset>       == t w
    <D phn peak>        == 'e:'
    <D phn peak bound>  == I

    <E phn onset>       == t
    <E phn onset bound> == t w
    <E phn peak>        == 'u:'
    <E phn peak bound>  == E

    <G phn onset>       == t s v
    <G phn peak>        == a i
    <G phn peak bound>  == a.
```

4.2 Conditional statements

The use of path extensions is the natural way to deal with allomorphic variation that is conditioned by variant properties of the unit involved. But it cannot be used for inherent properties of the unit since such properties will not be represented in the attributes that specify the inflected form. In such cases a different approach is required.

The approach adopted in PolyLex employs one of the most common idioms of modern programming languages, the **if ... then ... else ...** construct. In DATR, this construct takes the following form:

`IF:<`*condition* `THEN` *value1* `ELSE` *value2*`>`

where the condition is stated as some boolean combination of atomic truth-valued statements and *value1* and *value2* are phonological units (segments, for example). The atomic statements may involve predicates, such as `SCHWA`, `SIBILANT`, `VOICED`, `FEMININE`, and `ANIMATE` applied to arguments denoting phonological, morphological, or lexical units.

The *condition* can thus refer to any lexical information available, not just phonological. So, for instance, the realization of a phonological constituent may be determined by phonological aspects of the root or suffix, syntactic gender of the root or even semantic properties of the root (e.g., animacy in Russian noun inflection).

One of the major noun classes in Dutch suffixes an *-s* in the plural. The phonological realization of this is dependent on whether the root ends in a sibilant or not, taking the form /ə s/ if it ends in a sibilant and /s/ otherwise. This suffix node is defined in PolyLex as follows:

```
Suffix_S:
    <> == Affix
```



```
    <phn syl1 tail>  == s
    <phn syl1 peak>  == IF:<SIBILANT:<"Root:<phn syl1 tail>">
                        THEN ə
                        ELSE Null>.
```

German has an identical suffix, with identical variants. English also has an *-s* suffix, but because of the absence of final consonant devoicing in English, it also has a voicing contrast. This therefore requires two conditional statements, one for the peak which is identical to that for Dutch and one for the tail, stating that if the root final consonant is either voiced or a sibilant then the tail of the suffix is /z/ and otherwise it is /s/:

```
Suffix_Z:
    <> == Affix
    <phn syl1 peak> == IF:<SIBILANT:<"Root:<phn syl1 tail>">
                       THEN ə
                       ELSE Null>
    <phn syl1 tail> == IF:<OR:<VOICED:<"Root:<phn syl1 tail>">
                            SIBILANT:<"Root:<phn syl1 tail>">>
                       THEN z
                       ELSE s>.
```

German has two noun classes which suffix *-e*, one with umlaut and the other without. We include in these classes nouns which do not inflect in their plural (or which only umlaut the peak) where this is phonologically determined. The phonological requirement in these cases is that the final syllable must have a schwa peak. So the noun *Adler* has the singular and plural form /a:dlər/. We capture this in PolyLex by defining a suffix node `Suffix_e2`, distinct from `Suffix_e1`, which incorporates the alternation between /ə/ and null:

```
Suffix_e2:
    <> == Affix
    <phn syl1 rhyme> == IF:<SCHWA:<"Root:<phn syl1 peak>">
                        THEN Null
                        ELSE "Suffix_e1:<phn syl1 rhyme>">.
```

4.3. Combined path extensions and conditionals

Some allomorphic alternations involve both variant and inherent properties. In such cases it is appropriate to combine the two approaches to allomorphy outlined above. In the following hypothetical (but linguistically plausible) example, the plural form of the peak is /ə/ if the final segment of the root is a sibilant and /I/ otherwise, and the singular is always /a/:

```
Noun_a:
    <phn syl1 peak plur> == IF:<VOICED:<"Root:<phn syl1 tail>">
                           THEN e:
```

```
                                ELSE e>
    <phn syll peak>         == a.
```

There are not many examples of this kind in the PolyLex lexicons but one notable one is the German singular genitive suffix which only appears on masculine and neuter nouns. We can encode this fact as follows:

```
Suffix_s:
    <> == Affix
    <phn syll rhyme sing gen> ==
                        IF:<FEMININE:<"Root:<syn gender>">
                        THEN Null
                        ELSE "Suffix_S">.
```

The rhyme of this suffix is null if the syntactic gender of the noun is feminine and otherwise is inherited from the `Suffix_S` node, which as we have seen above, incorporates an additional phonological condition.

5. Conclusions

We have described the principal ways of representing allomorphic variation within the PolyLex lexicons. Alternations which are dependent solely on variant properties of the unit in question are captured with path extensions. Alternations which are dependent solely on inherent properties of the unit in question are captured with conditionals. These conditional statements may refer to any lexically available inherent information. In the case of the PolyLex languages, this includes morphosyntactic information (noun gender) but mostly involves phonological information about the root. When an alternation involves both variant and inherent properties of the units involved, then it is necessary to combine the use of path extensions with the use of conditionals. We have illustrated these methods with examples drawn from the PolyLex lexicon(s) for Dutch, English and German.

References

[1] Harald Baayen, Richard Piepenbrock & H. van Rijn (1995) The CELEX Lexical Database, Release 2 (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

[2] Doris Bleiching (1992) Prosodisches Wissen in Lexicon. In G. Görz, ed., *Proceedings of KONVENS-92,* Berlin: Springer-Verlag, 59-68.

[3] Doris Bleiching (1994) Integration von Morphophonologie und Prosodie in ein hierarchisches Lexicon. In Harald Trost, ed., *Proceedings of KONVENS-94,* Vienna: Österreichische Gesellschaft für Artificial Intelligence, 32-41.

[4] Doris Bleiching, Guido Drexel & Dafydd Gibbon (1996) Ein Synkretismusmodell für die deutsche Morphologie. In Dafydd Gibbon, ed., *Natural Language Processing and Speech Technology: Proceedings of KONVENS-96, Bielefeld.* Berlin: Mouton de Gruyter, 237-248.

[5] Dunstan Brown, Greville Corbett, Norman Fraser, Andrew Hippisley & Alan Timberlake (1996) Russian noun stress and network morphology. *Linguistics,* 34.1, 53-107.

[6] Dunstan Brown & Andrew Hippisley (1994) Conflict in Russian genitive plural assignment: A solution represented in DATR. *Journal of Slavic Linguistics,* 2.1, 48-76.

[7] Lynne Cahill (1990a) Syllable-based morphology. *Proceedings of the 13th International Conference on Computational Linguistics, COLING-90,* Vol. 3, 48-53.

[8] Lynne Cahill (1990b) Syllable-based morphology for natural language processing. DPhil Dissertation, University of Sussex, available as Cognitive Studies Research Paper 181.

[9] Lynne Cahill (1993) Morphonology in the lexicon. *Sixth Conference of the European Chapter of the Association for Computational Linguistics,* 87-96.

[10] Lynne Cahill & Gerald Gazdar (1996) A lexical analysis of numeral expressions in three related languages. *Proceedings of the AISB Workshop on multilinguality in the lexicon,* Brighton, UK, 69-75.

[11] Lynne Cahill & Gerald Gazdar (1997) The inflectional phonology of German adjectives, determiners and pronouns. *Linguistics* 35.2, 211-245.

[12] Lynne Cahill & Gerald Gazdar (forthcoming) German noun inflection. To appear in *Journal of Linguistics.*

[13] Andrew Carstairs[-McCarthy] (1987) *Allomorphy in Inflection.* London: Croom Helm.

[14] Roger Evans & Gerald Gazdar (1996) DATR: A language for lexical knowledge representation. *Computational Linguistics,* 22.2, 167-216.

[15] Norman Fraser & Greville Corbett (1995) Gender, animacy, and declensional class assignment: a unified account for Russian. In Geert Booij & Jaap van Marle, eds. *Yearbook of Morphology 1994.* Dordrecht: Kluwer, 123-150.

[16] Norman Fraser & Greville Corbett (in press) Defaults in Arapesh. To appear in *Lingua.*

[17] Dafydd Gibbon (1990) Prosodic association by template inheritance. In Walter Daelemans & Gerald Gazdar, eds. *Proceedings of the Workshop on Inheritance in Natural Language Processing.* Tilburg: Institute for Language Technology, 65-81.

[18] Dafydd Gibbon (1992) ILEX : a linguistic approach to computational lexica. In Ursula Klenk, ed. *Computatio Linguae: Aufsatze zur algorithmischen und quantitativen Analyse der Sprache (Zeitschrift fur Dialektologie und Linguistik,* Beiheft 73), Stuttgart: Franz Steiner Verlag, 32-53.

[19] Dafydd Gibbon & Doris Bleiching (1991) An ILEX model for German compound stress in DATR. *Proceedings of the FORWISS-ASL Workshop on Prosody in Man-Machine Communication*, 1-6.

[20] Jaap van Marle (1985) *On the Paradigmatic Dimension of Morphological Creativity*. Dordrecht: Foris.

[21] Peter Matthews (1972) *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.

[22] Sabine Reinhard (1990) Verarbeitungsprobleme nichtlinearer Morphologien: Umlautbeschreibung in einem hierarchischen Lexikon. In Burghard Rieger & Burkhard Schäder *Lexikon und Lexikographie*. Hildesheim: Olms Verlag, 45-61.

[23] Sabine Reinhard & Dafydd Gibbon (1991) Prosodic inheritance and morphological generalisations. *Fifth Conference of the European Chapter of the Association for Computational Linguistics*, 131-136.

[24] Stuart M. Shieber (1986) *An Introduction to Unification Approaches to Grammar*. Stanford: CSLI/Chicago University Press.

[25] Gregory T. Stump (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, 211-241.

[26] Gregory T. Stump (1993a) On rules of referral. *Language* 69, 449-479.

[27] Gregory T. Stump (1993b) Reconstituting morphology: The case of Bantu preprefixation. *Linguistic Analysis* 23, 169-204.

[28] Gregory T. Stump (1993c) Position classes and morphological theory. In Geert Booij & Jaap van Marle, eds. *Yearbook of Morphology 1992*. Dordrecht: Kluwer, 129-180.

[29] Gregory T. Stump (1995) The uniformity of head marking in inflectional morphology. In Geert Booij & Jaap van Marle, eds. *Yearbook of Morphology 1994*. Dordrecht: Kluwer, 000-000.

[30] John C. Wells (1987) Computer coded phonetic transcription, *Journal of the International Phonetic Association*, 17:2, 94-114.

[31] Wolfgang Wurzel (1990) The mechanism of inflection: lexicon representations, rules, and irregularities. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison, eds. *Contemporary Morphology*. Berlin: Mouton de Gruyter, 203-216.

[32] Arnold Zwicky (1985) How to describe inflection, *BLS* 11, 371-386.

COVERT GENERALIZATION

Paul Kiparsky
Stanford University

In a context where allomorphs /A/ and /B/ merge to [A], what is the underlying representation of non-alternating [A]?[1]

[1] Underlying: /A/ /B/
Output: [A]

Analogical change shows that it is analyzed as /A/. The evidence is that when a neutralization process /B/ → [A] is lost, non-alternating outputs of the form [A] remain unchanged (Kiparsky 1968, 1973). This preference for "face value" analyses is predicted by several theories, including Natural Phonology (Stampe 1972/1980), Natural Generative Grammar (Vennemann 1973, Hooper 1976), Lexical Phonology and Morphology (Kiparsky 1982), and Optimality Theory (Prince & Smolensky 1993).

But analogical change also provides evidence that this is not the whole story. The face-value analysis can be overridden by positive evidence for a deeper underlying form. I shall argue that the correct generalization is as follows:

[2] a. Non-alternating forms are assigned the optimal lexical representation.
b. Of several equally optimal lexical representations, the one closest to the output is preferred.

By [2a], non-alternating [A] is analyzed as underlying /B/ rather than as /A/ when /B/ conforms better to constraints on lexical representations, such as those relating to the phonological inventory, phonotactics, or the structure of particular classes of morphemes. Case [2b] is then only the tie-breaker.

This view is actually a consequence of Lexical Phonology and Morphology (LPM), and it is at least compatible with Natural Phonology. On the other hand, it seems irreconcilable with Vennemann's and Hooper's NGG. And if Prince and Smolensky are right that the form of lexical representations is derivative of constraints on the output (Richness of the Base, Lexicon Optimization), then lexical constraints could never choose /B/ as the preferred underlying form of an output that is always overtly realized as [A]. In that version of OT, case [2a] cannot be distinct from case [2b].

I will present evidence which indicates that case [2a] is not reducible to case [2b], and supports a version of OT phonology where lexical and postlexical phonology constitute

[1] I would particularly like to acknowledge Wayne Redenbarger, who long ago suggested, as a corollary of his analysis of Latin morphology in Redenbarger 1974, that Nom.Sg. *honor* could be a case of intervocalic rhotacism. Thanks are also due to Edward Flemming, Andrew Garrett, Bruce Hayes, and Donca Steriade for critical discussion.

separate, serially related constraint systems. Such a theory allows properties of lexical representations to be determined by phonological and morphological constraints at the lexical level, not just derivatively by constraints on output forms.

[2a] implies that lexical constraints can induce *covert reanalysis* of surface [A] as /B/, which may be overtly manifested in analogical change. Latin morphology provides a case in point. The constraint at stake is a preference for vocalic endings over consonantal endings, which drives a series of analogical suffix replacements, of which some are overt, others covert. In the covert cases, the replacement of the original consonantal ending by a vocalic ending at the *underlying* level is manifested indirectly through its contextual effects, which include the famous generalization of *-r* in from the oblique stem to the nominative singular in certain nouns and adjectives, e.g. *honōs* > *honor*.

Such covert generalizations are of theoretical interest in another respect as well. They are incomprehensible in terms of proportional analogy, output-output constraints, or other surface-oriented approaches to lexical relationship. Viewed in terms of the pre-reanalysis underlying form /A/, the overt consequences of the reanalysis to /B/ can appear as complications of the grammar (exceptions, morphological conditions), or as "paradigm uniformity" effects seemingly restricted to arbitrary contexts. At least in the present case, as well as in the analogous Gothic case presented in Kiparsky (in press), there is no complication of the grammar, nor even any change in the constraint system. What happens rather is that an exceptional morpheme is brought into line with the constraints that organize the language's morphology.

In Latin, intervocalic *s* becomes *r*, a process known as RHOTACISM (for the sound change, see e.g. Leumann 1963:140). Synchronically, rhotacism applies only in *derived environments*. Morpheme-internal invariant /s/ occurs both in native words such as [3a] and in presumed loanwords such as [3b].

[3] a. *miser* 'miserable'
b. *rosa* 'rose', *asinus* 'donkey'

Rhotacism does not apply to intervocalic *s* from *-ss-*, *-dt-*, *-tt-* by degemination after a long vowel, a word-level constraint:

[4] /haus+s+ī/ *hausī* 'I drained', /kad+t+us/ *cāsus* 'fallen', /vid+t+us/ *vīsus* 'seen'.

In LPM, these properties follow from the *cyclic* status of rhotacism.

Synchronically, rhotacism is *productive*. It functions as a rule of Latin morphophonology accounting for *s* ∼ *-r* alternations at morpheme edges in both derivational and inflectional morphology:

[5] a. *fūnus, fūner+is* 'body', *fūnes+tu+s* 'funereal'
scelus, sceler+is 'crime', *sceles+tu+s* 'criminal'
mās, mar+is 'male' (N.), *mas+cul+u+s* 'male' (A.)
corpus, corpor+is 'body' , *corpus+culum* 'little body', *in+corpor+ō* 'incorporate', *corpor+āl+is* 'bodily'
latus, later+is 'side', *later+āl+is* 'lateral', *latus+cul+u+m* 'little side'
lepus, lepor+is 'rabbit', *lepus+cul+u+m* 'bunny', *lepor+āri+a* 'rabbit meat'
glomus, glomer+is 'ball', *(con)+glomer+ō* 'to compact into a ball'
iūs 'justice', *iūs+tu+s* 'just', *in+iūr+ia* 'injustice', *iūr+ā+re* 'swear'
fās '(divine) justice', *ne+fār+iu+s*

b. *ger+ō* 'wear, carry', *ges+t+u+s* (pp.)
quaer+ō 'seek, interrogate', *quaes+tiō* 'search, interrogation'
haur+īre 'drain', *haus+t+u+s* '(pp.)'

c. *dis+put+ō, dis+curr+ō, dis+tine+ō* vs. *dir+im+ō, dir+ibe+ō*.

Conclusive evidence of the synchronic productivity of rhotacism is the fact that it applies in words newly introduced into Latin after the historical change had already taken place. These include loanwords ([6a]), and new *-s*-stems resulting from reanalysis ([6b]):

[6] a. *tūs, tūr+is* 'incense' (borrowed from Greek *thūos*).

b. *glōs, glōr+is* 'sister-in-law' (*/glō+s/ reanalyzed as /glōs/)
femus, femor+is 'thigh', Late Latin for classical *femur, femor+is* (/femur/ reanalyzed as /femus/),
bover+um 'oxen's' (Gen.Pl., Varro), with a stem /bouis/ resulting from reanalysis of Nom. *bov+is* (itself from *bōs* /bou+s/, see below) as /bouis+s/.

In fact, once its character as a derived-environment process is understood, it can be seen that *rhotacism is virtually exceptionless*. The only exception in noun inflection that I am aware of is *vās, vāsis* 'vessel'.[2]

The stem-final *-s* ∼ *-r* alternation resulting from rhotacism is leveled in some *s*-stem nouns by changing stem-final *-s* to *-r* in the nominative singular (Saussure 1916:230, Hoenigswald 1960:108-111, Kiparsky 1972, Hooper 1976:95, Kuryłowicz 1977:14, Hock 1979:179-180, Dresher and Lahiri 1983, Wetzels 1981, Ch. 4, Wetzels 1984, Kenstowicz 1996, Hale, Kissock, and Reiss 1998). The vowel in the innovating form is shortened in obedience to a constraint prohibiting polysyllabic words in *-V̄r, -V̄l*.

[7]

	Old Latin		Classical Latin	
Nom.	**honōs**	honōr+ēs	**honor**	honōr+ēs
Acc.	honōr+em	honōr+ēs	honōr+em	honōr+ēs
Gen.	honōr+is	honōr+um	honōr+is	honōr+um
Dat.	honōr+ī	honōr+ibus	honōr+ī	honōr+ibus
Abl.	honōr+e	honōr+ibus	honōr+e	honōr+ibus

In classical Latin, this analogical change affects primarily **masculine and feminine disyllabic and polysyllabic nouns**, including underived nouns, such as [8a], and derived nouns such as [8b], including about 60 nouns in *-or* from **-ōs*:

[8]

a.			b.		
arbōs > arbor	'tree'		**pallōs > pallor*	'pallor'
honōs > honor	'honor'		**albōs > albor*	'whiteness'
odōs > odor	'odor'		**vigōs > vigor*	'vigor'
vapōs > vapor	'vapor'		**rigōs > rigor*	'rigor'
colōs > color	'color'		**terrōs > terror*	'terror'
**augus > augur*	'omen'		*fulgus > fulgur*	'lightning'
vōmis > vōmer	'plowshare'			

[2]Even this may be only an apparent exception, if the word is really /uās/, /uāss+is/, with regular degemination after long vowels. At least some Latin speakers seem to have analyzed it that way, for the Nom.Pl. is spelled *vassa*, with *-ss*, in the Milan MS of Plautus. There is independent evidence that final geminates "count" in Latin: words such as /uad+(i)s/ → *vass* → *vās* 'guarantor' and /farr+(i)s/ *fār* 'spelt', seem to satisfy prosodic minimality only in virtue of their underlying /-CC/.

Most **neuter nouns** retain *-s*:[3]

[9] *corpus, corpor+is* 'body' ↛ **corpor, genus, gener+is* 'kind' ↛ **gener, scelus* 'crime', *crūs* 'leg', *onus* 'burden', *fūnus* 'funeral', *vulnus* 'wound', *stercus* 'dung', *tempus* 'time', *sīdus* 'star', *pecus* '(head of) cattle', *pectus* 'chest', *tergus* 'rear', *glomus* 'ball', *nemus* 'woods', *latus* 'side'

Adjectives fall into two classes. Adjectives derived from noun compounds (the *bahuvrīhi* type) level the alternation in all three genders (see [10a]), while other adjectives retain *-s* in the neuter (see [10b]):

[10] a. **bicorpus* > *bicorpor* ' two-bodied' (m., f., and n.)
**dēgenus* > *dēgener* 'degenerate' (m., f., and n.)

b. *melius* (n.), *melior* (m., f.) 'better', *meliōr+is* (gen.sg)
plūs (n.) 'more' (no m.f. sg.), *plūr+is* (gen.sg.)

Monosyllables retain *-s*:[4]

[11] *flōs* (m.) 'flower', *mōs* (m.) 'custom', *mūs* (m.) 'mouse, rat', *rōs* (m.) 'dew', *vīs* (f.) 'power', *mās* (m.) 'male'

Since rhotacism continues to govern productive alternations in Latin (see [5]), the putative reanalysis /honōs/ > /honōr/ would not be a simplification of the grammar. Moreover, some words that level the nominative singular keep the *-s* ∼ *-r* alternation in derivation:

[12] *arbor* > *arbus+tus* 'wooded', *arbus-cul-a* 'sapling'
honor > *hones+tu+s* 'honorable'
rōbur > *rōbus+tu+s* 'strong'
augur > *augus+tu+s* 'sacred, exalted'

These data somewhat undermine the credibility of an analysis according to which /-s/ stems are restructured as /-r/ stems, e.g. /honōs/ > /honōr/. Such a restructuring would entail complementing rhotacism with a backwards rhotacism ("sigmacism") process of the form $r \rightarrow s$ / __ C, or with the equivalent constraint or constraints barring *r* before consonants, in order to account for the alternations in [12]. Sigmacism adds a further measure of complexity to the grammar because it must be prevented from applying when the base ends in a "real" /r/, e.g. /fur+tu+m/ 'theft' (↛ **fustum*), from *fūr* 'thief'. Thus a putative reanalysis of /-s/ stems to /-r/ stems would not be an optimization. Some other motivation for the directionality of the change would be required, perhaps having to do with the special status of the nominative singular in the case paradigm (see Hooper 1976, Dresher & Lahiri 1983, Wetzels 1984 for discussion).

An adequate historical account of the Latin change should not only address this point, but also the following questions:

[13] a. Why does the analogy not apply in neuter nouns?

[3] *-r* is extended to neuters in a few words that have a masculine gender doublet: **rōbus, rōbor+is* (n.) 'oak heartwood, strength' → *rōbor*, cf. *rōbor* (m.); *fulgus, fulger+is* → *fulgur, fulguris* (n.) 'lightning', cf. *fulgor, fulgōr+is* (m.) 'id.'. But *decus, decor+is* (n.) versus *decor, decōr+is* (< **decōs*) (m.) 'dignity', *frīgus, frīgor+is* (n.) versus *frīgor, frīgor+is* (m.) 'cold' (Kieckers 1931:36).

[4] But *Lās* > *Lār* 'household god'; it mostly occurs in the plural *Larēs*, and has no derivatives with *-s*, so it may have been synchronically reanalyzed as /lar/.

b. Why does the analogy apply in some neuter adjectives but not in others?

c. Why does the analogy not apply in monosyllabic nouns (Kuryłowicz 1977:14)?

d. Why does the analogy not apply in verbs and prefixes? For example, why not **gertus* for *gestus*, or for that matter **gesō* for *gerō*? Why not **dir+tineō* for *dis+tineō* 'separate', on the analogy of *dir+imō* 'separate', or **dis+imō* by the converse analogy?

e. *honōs* > *honor* eliminates the *-s* ∼ *-r* alternation but in turn introduces an *-ō-* ∼ *-o-* alternation into the paradigm (Kiparsky 1972, Hale, Kissock, and Reiss 1998). Then in what sense can it be characterized as a leveling?

f. Why not instead *honōrem* : *honōs* = *sorōrem* = X (X = **sorōs*) (Kuryłowicz 1977:14), or indeed *soror* : *sorōrem* = *honōs* : X (X = **honōsem*)?

For an answer, let us take a closer look at the third declension, to which the *s*-stems belong.

[14]

	'circle'	'city'	'leader'	'mountain'	'enemy'	'sister'	'dog'
N.	orbis	urbs	dux	mōns	hostis	soror	canis
A.	orbem	urbem	ducem	montem	hostem	sorōrem	canem
G.	orbis	urbis	ducis	montis	hostis	sorōris	canis
D.	orbī	urbī	ducī	montī	hostī	sorōrī	canī
Ab.	orbe	urbe	duce	monte	hoste	sorōre	cane
N.	orbēs	urbēs	ducēs	montēs	hostēs	sorōrēs	canēs
A.	orbēs	urbēs	ducēs	montēs	hostēs	sorōrēs	canēs
G.	orbium	urbium	ducum	montium	hostium	sorōrum	canum
D.	orbibus	urbibus	ducibus	montibus	hostibus	sorōribus	canibus
Ab.	orbibus	urbibus	ducibus	montibus	hostibus	sorōribus	canibus

From a historical point of view, the third declension includes several classes of stems which were morphologically distinct in Indo-European, including *-i-*stems and *-C-*stems, which contrast in genitive plurals, cf. [15a] and [15b]:[5]

[15]

	Nom.Sg.		Gen.Pl.		
a.	/imbrek+s/	imbrex	/imbrik+um/	**imbricum**	'tile'
	/parent+s/	parēns	/parent+um/	**parentum**	'parent'
b.	/simpleki+s/	simplex	/simpleki+um/	**simplicium**	'simple'
	/ingenti+s/	ingēns	/ingenti+um/	**ingentium**	'huge'

The Nom.Sg. ending *-is* is reduced to *-s* in polysyllables, and, with some exceptions, after consonant clusters. This NOMINATIVE *i*-DELETION process applies to /-i+s/ and /-C+is/ but not to /-i+is/. The stems that undergo it will be referred to as LONG STEMS.

[16] a. /kom+potis/ *compos* 'in control' vs. /potis/ *potis* 'able'
/inter+kutis/ *intercus* 'subcutaneous' vs. /kutis/ *cutis* 'skin'

b. /nostrāti+s/ → *nostrās* 'native, domestic' (gen.pl. *nostrātium*, Old Latin *nostrātis*)

[5] Also in Acc.Pl. *-ēs* vs. *-īs*, in neuter Nom.Pl. *-a* vs. *-ia*, in Abl.Sg. *-e* vs. *-ī*, and sometimes in Acc.Sg. *-em* vs. *-im*. The analysis proposed here considerably regularizes the distribution of these endings.

c. /menti+s/ → *mēns* 'mind' (gen.pl. *mentium*)
/morti+s/ → *mors* 'death' (gen.pl. *mortium*)
/nokti+s/ → *nox* 'night' (gen.pl. *noctium*)
/imbri+s/ → *imber* 'shower' (gen.pl. *imbrium*)

Short -C stems, where deletion does not apply, show that there are two nominative singular allomorphs /-s/ and /-is/; contrast [17a] and [17b]:

[17]

	Nom.Sg.		Gen.Pl.		
a.	/op+s/	**ops**	/op+um/	opum	'help', pl. 'means'
b.	/kan+is/	**canis**	/kan+um/	canum	'dog'

Long *-i* stems also provide evidence for the distinction between the allomorphs /-s/ and /-is/; contrast [18a] and [18b].

[18]

	Nom.Sg.		Gen.Pl.		
a.	/nokti+s/	**nox**	/nokti+um/	noctium	'night' (f.)
	/urbi+s/	**urbs**	/urbi+um/	urbium	'city' (f.)
	/sekstanti+s/	**sextāns**	/sekstanti+um/	sextantium	'one sixth As' (m.)
	/imbri+s/	**imber**	/imbri+um/	imbrium	'shower' (m.)
b.	/uekti+is/	**vectis**	/uekti+um/	vectium	'lever' (m.)
	/orbi+is/	**orbis**	/orbi+um/	orbium	'circle' (m.)
	/sēmenti+is/	**sēmentis**	/sēmenti+um/	sēmentium	'planting' (f.)
	/febri+is/	**febris**	/febri+um/	febrium	'fever' (f.)

Similarly /sēcūri+is/ *sēcūris* 'axe' (f.) vs. /dispāri+s/ *dispār* 'unequal'. Generally, /-is/ is more frequent in feminines (as is the accusative ending /-im/, which I take as underlying /-i+im/, e.g. /febri+im/ *febrim* 'fever'). In feminine *-er* adjectives, /-is/ is regular:

[19] /ākeri+s/ → *ācer* (masc.), /ākeri+is/ → *ācris* (fem.) 'sharp'

In sum, there are three inflectional patterns in the long stems. The absence of the fourth possible pattern is explained by the proposed analysis, for two of the four potential inputs, [20c] and [20d], converge on the same output, and by the same token, no input yields a paradigm Nom.Sg. *-is*, Gen.Pl. *-um* for long stems:

[20]

	Nom.Sg.	Gen.Pl.	Nom.Sg.	Gen.Pl.
a.	/-i+is/	/-i+um/	*-is*	*-ium*
b.	/-i+s/	/-i+um/	*-s*	*-ium*
c.	/-is/	/-um/	*-s*	*-um*
d.	/-s/	/-um/	*-s*	*-um*

The Nom.Sg. /-is/ ~ /-s/ variation is part of a larger pattern of allomorphy. Other inflectional endings that begin with consonants have developed an allomorph in /-i/, which is favored after -C stems, where it eventually replaces the consonantal allomorph completely. In IE terms, vocalic inflection replaces consonantal inflection, and in Romance terms, parisyllables replace imparisyllables. Thus Dat./Abl.Pl. /-bus/ > /-ibus (see [21a]) and 2.3.Sg. /-s, -t/ > /-is, -it/ ([21b]) are parallel to Nom.Sg. /-s/ > /-is/ (see [21c]):

[21] a. **rēg+bus* > *rēgibus* 'king' (Dat./Abl.Pl.)

b. *ēst* > *ēdit* (/ēd+t/ > /ēd+it/) 'eats'
fert > *ferit* 'carries' (postclass.)

c. **aus* > *auris* 'ear'
bōs > *bovis* 'bovine animal' (Varro)
mōns > *montis* 'mountain' (Ennius)
mēns > *mentis* 'mind' (Ennius)
grūs > *gruis* 'crane' (Phaidr.)
sors > *sortis* 'lot' (Plautus)
frōns > *frontis* 'forehead'
frōns > *frondis* 'frond'
glāns > *glandis* 'acorn'
mūgil > *mūgilis* 'mullet' (Juvenal)

The selection of vocalic and consonantal allomorphs is done by the syllable structure constraints *CODA and ONSET: -C stems get the vocalic Dat./Abl.Pl. allomorph /-ibus/, while -V stems get the consonantal allomorph /-bus/. Similarly, -C stems get Gen.Pl. /-um/, while -V stems get /-rum/.[6]

[22] a.

Input		Candidates	*CODA	ONSET
/sorōr+bus, -ibus/		sorōr+bus	*	
	☞	sorōr+ibus		
/rē+bus, -ibus/	☞	rē+bus		
		rē+ibus		*

b.

Input		Candidates	*CODA	ONSET
/sorōr+rum, -um/		sorōr+rum	*	
	☞	sorōr+um		
/rē+rum, -um/	☞	rē+rum		
		rē+um		*

As shown in [20], NOMINATIVE *i*-DELETION neutralizes the potential distinction between /-s/ and /-is/ in long -C stems. A word like Nom.Sg. *parēns* could be underlying /parent+s/ or /parent+is/, and *sorōr* 'sister' could be underlying /sorōr+s/ or /sorōr+is/. This is a concrete instance of the neutralization schema in [1].

Theories which countenance constraints on lexical representations, such as LPM, make the following prediction here. The underlying representation of structurally ambiguous non-alternating outputs, *whatever their surface realization*, is the one that best conforms to lexical constraints (case [2a]). In particular, this means that the *CODA constraint in the lexical phonology will cause ambiguous non-alternating *-s* after long -C stems to be analyzed as underlying /-is/.

[23] Underlying: /-s/ /-is/
Output: -s

I will assume a lexical phonology with the constraints in [24] and [25]:[7]

[6] Stems in long *-ū* often pattern with -C stems, e.g. /fū+it/ *fuit* 'was' vs. /dā+t/ *dat* 'gives', /grū+is/ *gruis* 'crane'.

[7] Please note that these constraints represent a very preliminary analysis, and even so are grossly simplified

[24] Cyclic phonology:

a. *VsV (the constraint that drives rhotacism)

b. *CODA: A syllable must lack a coda.

c. ONSET: A syllable must have an onset.

[25] Word-level phonology:

a. STEM-FORM: A stem must contain at least a two-mora foot (not counting a stem-final -C, which is not moraic).

b. *VVR]$_\omega$: A word cannot end in a long vowel followed by *-l, -r*. (Dominated by STEM-FORM, hence no shortening in monosyllables).

c. '*-RC]$_\sigma$: A syllable cannot end in a sonorant+obstruent cluster.[8]

d. *NOM-*i*: *-i-* is deleted in the final syllable of nominative singular forms. (Probably not a single constraint but a complex of constraints. Synchronically, it has to be restricted to nominatives because of endings like gen.sg. *-is*, historically **-es*.)

As mentioned, the rhotacism constraint *VsV is virtually exceptionless. *CODA and ONSET, of course, are less often seen in action in Latin because they are dominated by Faithfulness constraints. Their role in allomorphy selection is thus a case of the emergence of the unmarked in the sense of McCarthy and Prince.

Tables [26] and [27] show the analysis of Nom.Sg. *-s* as /-is/ in long stems. Syllable-driven allomorphy selection in the cyclic lexical system yields *sorōr+is, rē+s* (see [26]), and the word-level constraints reduce *sorōr+is* to *soror* (see [27]).

[26]

Cyclic	Candidates	*CODA	ONSET
/sorōr+s, -is/	sorōr+s	*	
☞	sorōr+is		
/can+s, -is/	can+s	*	
☞	can+is		
/rē+s, -is/ ☞	rē+s		*
	rē+is		

to save space. In particular, for a phonological constraint not to figure in the cyclic or word level phonology really means that it is dominated at that level by an antagonistic Faithfulness constraint. In the actual constraint system, the cyclic and word level systems are not disjoint, but include essentially the same constraints, and differ rather in the ranking of Faithfulness constraints among them.

[8] Unless the obstruent is reduced from a two-consonant sequence, as in /monti+s/ → *mōns*, /parti+s/ → *pars*. Let us assume that such sequences are allowed to persist because of a Faithfulness constraint that dominates *-RC]$_\sigma$, which demands the retention of the segmental content of multiply linked phonemes (geminates and quasi-geminates). That constraint could also be responsible for the failure of rhotacism to apply to degeminated *-ss-*, e.g. /haus+s+l/ → *hausī* (**haurī*), and for the retention of /-i+is/ as *-is* in the face of NOMINATIVE *i*-DELETION.

[27]

W.L.	Candidates	STEM-FORM	*VVR]$_\omega$	*-RC]$_\sigma$	*NOM-*i*
/sorōr+is/	sorōr+is				*
	sorōr+s			*	
	sorōr		*		
☞	soror				
/kan+is/ ☞	can+is				*
	can+s			*	
	can	*			

In the word-level phonology, *i* is deleted in *sorōris* in satisfaction of *NOM-*i*, but retained in short stems like *canis* because deletion would violate the STEM-FORM constraint.

It can now be seen that the leveling of *-s* ∼ *-r* is a consequence of the (covert) spread of the Nom.Sg. allomorph /-is/. The regularized inflection of /honōs-/ runs entirely parallel parallel to the inflection of /sorōr-/, except that the cyclic *VsV constraint, which was vacuously satisfied in /sorōr-/, enforces rhotacism in Nom.Sg. /honōs+is/ → *honōris* (see [28]). The word-level constraint system takes *honōris* to *honor* (see [29]):

[28]

Cyclic	Candidates	*VsV	*CODA	ONSET
/honōs+bus, -ibus/	honōs+bus		*	
	honōs+ibus	*		
☞	honōr+ibus			
/honōs+s, -is/	honōs+s		*	
	honōs+is	*		
☞	honōr+is			

[29]

W.L.	Candidates	STEM-FORM	*VVR]$_\omega$	*RC]$_\sigma$	*NOM-*i*
/honōr+is/	honōr+is				*
	honōr+s			*	
	honōr		*		
☞	honor				

We have seen that the Nom.Sg. allomorph /-is/ spreads through the third declension over a long period. On the surface, this spread results in diverse changes, or no change, according to what the regular phonology of Latin dictates:

[30]

Old system		New system		
Underlying	Surface	Underlying	Surface	overt change
/sorōr+s/	*soror*	/sorōr+is/	*soror*	none
/kan+s/	**can(s)*	/kan+is/	*canis*	ending
/honō+s/	*honōs*	/honōs+is/	*honor*	stem
/monti+s/	*mōns*	/monti+is/	*montis*	stem and ending

The questions we posed in [13]a-f can now be answered.

Re [13a]: the reason the analogy does not apply in **neuter nouns** is that neuters do not

have a Nom./Acc. Sg. ending, as shown by the absence of *-s* even in those stems where an underlying /-s/ would have to surface:

[31] *mare* 'sea' *(*maris)*, *lac* 'milk' *(*lactis, *lax)*, *caput* 'head' *(*capis)*.

Because neuters have no Nom.Sg/-s/, there is no occasion to regularize /-s/ to /-is/, and hence no rhotacism.

Re [13b]: neuter *s*-stem **adjectives** get analogical Nom.Sg. *-r* in those morphological classes which are inflected with an overt Nom.Sg. ending, and retain stem-final *-s* (as neuter nouns do) in those morphological classes which are inflected without an overt Nom.Sg. ending. In particular, neuter adjectives of the form X+Noun (compounds of the *bahuvrīhi* type) get Nom./Acc.Sg. *-s* (which is synchronically /-is/ by our hypothesis, see [32a]). Otherwise, neuter adjectives have no Nom./Acc.Sg. ending, like neuter nouns (see [32b]).

[32] a. *bidēns* /bi+dent+is/ 'two-toothed', *triceps* /tri+kapit+is/, 'three-headed', *quadrupēs* /k*adru+ped+is/ 'four-footed', *duplex* 'double', *triplex* 'threefold' (Leumann 1963:265).

b. *brevis* (m.,f.), *breve* (n.) 'short' (cf. *mare* 'sea')

Therefore, those adjectives in /-s/ which are *bahuvrīhi* compounds level out rhotacism in all three genders, whereas those adjectives in /-s/ which do not belong to this class adopt the majority pattern. Adjectives formed with the comparative suffix *-ior* ~ *-ius* and *plūs* belong to this type.[9]

[33] a. /bi+korpus+is/ → *bicorpor* 'two-bodied' (m., f., and n.)
/dē+genus+is/ → *dēgener* 'degenerate' (m., f., and n.)

b. /citius+is/ *citior* (m.,f.), /citius/ *citius* (n.) 'faster'

Re [13c]: **monosyllables** retain Nom.Sg. *-s* because a *covert* generalization of the vocalic ending /-is/ is ruled out, for *i*-deletion doesn't apply in short stems. Consequently, *glīs* does not become **glīr* the way *honōs* becomes *honor*. The analogically reformed /glīs+is/ is a short stem, so its *-i-* is not subject to deletion, and the expected analogical output, instead, is *glīris*.

In fact, *glīs* > *glīris* is actually attested in late Latin *(Appendix Probi)*. Thus, the real counterpart to the analogical spread of rhotacism in the nominative singular long stems *(honōs* > *honor)* is the analogical spread of the *overt* ending *-is* in the nominative singular of short stems *(glīs* > *glīris*, analogous to cases like *mōns* > *montis* etc., see [21c] and [30]). But the overt generalization of /-is/ has a more drastic effect on the output form than its covert counterpart, for it not only replaces the stem-final consonant but also adds a syllable. Because the short stems provide the learner with clearer overt evidence than the long stems, they change more slowly, with the time lag characteristic for the salient cases of an innovation (Naro 1981).[10]

In postclassical Latin, as the declensional classes tended to merge, many stems that retained the *-s* ~ *-r* alternation in the classical language leveled it out:

[9] Adjectives that usually modify human beings, like *vetus* 'old', *pūbēs* 'having reached maturity', *dīves* 'rich', *caelebs* 'unmarried', naturally rare in the neuter, seem to have *-s* throughout the nominative of all three genders in the classical language, as far as it is possible to tell.

[10] In *canis* 'dog' /-is/ was introduced early, because the inherited **can(s)* violates foot minimality; also interestingly enough in *iuvenis* 'youth' (an exception to NOMINATIVE *i*-DELETION!).

[34] *cinis* > *ciner* 'cinder'
pulvis > *pulver* 'dust' (Gloss.)
cucumis > *cucumer* 'cucumber'
vetus > *veter* 'old' (Ennius)
pūbēs > *pūber* 'mature'

In the Romance languages, the allomorphy is quite obliterated: Italian *fiore, monte* (but *corpo, tempo, petto* etc.), French *fleur, mont*.

Re [13d]: unlike the innovation that actually happened, such innovations as **gesō*, **gertus*, **dis+imō*, or **dir+tineō* would complicate the system: *ges-* ~ *ger-* and *dis-* ~ *dir-* are phonologically regular reflexes of underlying forms /ges/ and /dis/, so changing the output forms would not reduce allomorphy, but simply introduce gratuitous exceptions to the otherwise well-behaved rhotacism process.

Re [13e]: the *-s* > *-r* replacement is a side effect of the generalization of the vocalic allomorph /-is/, which is itself driven by the optimization of syllable structure in *lexical representations*. The same generalization has another side effect of the opposite kind, a *differentiation* of the vowel length between the Nom.Sg. and oblique stem *(honor* vs. *honōr-)*. Both are the expected phonological consequences of a *morphological* change.

Re [13f]: the hypothetical analogies resulting in **sorōs* or **honōsem* would not simplify either the phonology or the morphology; the latter would in fact complicate it. The analogical change that actually happened removes a morphological exception, regularizing the distribution of the nominative singular endings.

We may conclude that there is evidence for optimization of lexical representations, which implies that there is a system of constraints that characterizes regularities at that level. Analogical change can implement this constraint system, causing lexical representations to conform to the canonical distribution of allomorphs and segments of the language. The case examined here is not an instance of surface analogy. It involves the elimination of unmotivated language-specific restrictions on the *CODA constraint in allomorphy, independently manifested in other morphological alternations of the language. Thus it is consistent with the view that *analogical change is grammar optimization*.

Department of Linguistics
Stanford University
Stanford, CA. 94305-2150
U.S.A.

Bibliography

ALLEN, W. SIDNEY. 1973. Accent and rhythm: prosodic features of Latin and Greek: a study in theory and reconstruction. Cambridge: CUP.

LAHIRI, ADITI AND BEZALEL ELAN DRESHER. 1983. Diachronic and synchronic implications of declension shifts. *The Linguistic Review* 3:141-163.

HALE, MARK, MADELYN KISSOCK, AND CHARLES REISS. 1998. Output-Output correspondence in Optimality Theory. *WCCFL* 16.

HOCK, HANS HEINRICH. 1979. Principles of historical linguistics. The Hague: Mouton.

HOENIGSWALD, HENRY. 1960. Language change and linguistic reconstruction. Chicago: U. of Chicago Press.

HOOPER, JOAN. 1976. *An introduction to Natural Generative Phonology*. New York: Academic Press.

KENSTOWICZ, MICHAEL. 1996. Base-identity and uniform exponence: alternatives to cyclicity. In J. Durand and B. Laks (edd.) *Current trends in phonology: models and methods*. Salford: ESRI.

KIECKERS, E. 1931. *Handbuch der vergleichenden lateinischen Grammatik*. München: Hueber.

KIPARSKY, PAUL. 1968. Linguistic universals and linguistic change. In E. Bach and R. Harms (ed.), *Universals in linguistic theory*. New York: Holt.

KIPARSKY, PAUL. 1972. Explanation in phonology. In S. Peters (ed.) *Goals of linguistic theory*. Englewood Cliffs: Prentice-Hall.

KIPARSKY, PAUL. 1973. Phonological representations. In Osamu Fujimura (ed.), *Three dimensions of linguistic theory*. Tokyo: TEC.

KIPARSKY, PAUL 1998. Sievers' Law as prosodic optimization. In Jay Jasanoff, H. Craig Melchert, and Lisi Oliver (edd.) *Mír Curad: Studies in honor of Calvert Watkins*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

KIPARSKY, PAUL In press. Analogy as optimization: "exceptions" to Sievers' Law in Gothic.

KURYŁOWICZ, JERZY. 1977. *Problèmes de linguistique générale*. Warsaw: Polska Akademia Nauk.

LEUMANN, MANU. 1963. *Lateinische Laut- und Formenlehre*. München: Beck.

NARO, ANTHONY. 1981. The social and structural dimensions of a syntactic change. Language 57.63-98.

PRINCE, ALAN AND PAUL SMOLENSKY. 1993. *Optimality theory: Constraint interaction in generative grammar*. Ms., Rutgers University and U. of Colorado.

REDENBARGER, WAYNE. 1974. Vowel-lowering and *i*-epenthesis in classical Latin. In Marta Luján and Fritz Hensey (ed.) *Current Studies in Romance Linguistics*, Washington D.C.: Georgetown University Press.

SAUSSURE, FERDINAND. 1916. *Cours de linguistique générale*. Publié par Charles Bally et Albert Sechehaye. Lausanne/Paris: Payot.

STAMPE, DAVID. 1972. *How I spent my summer vacation*. Ph.D. Dissertation, University of Chicago. New York, Garland, 1980.

WETZELS, LEO. 1981. *Analogie et lexique: le problème de l'opacité en phonologie générative*. Nijmegen: De witte studentenpers.

WETZELS, LEO 1984. Paradigm leveling in Latin and Old French: a critical view of rule inversion. In Philip Baldi (ed.), Papers from the XIIth Linguistic Symposium on Romance Languages. Amsterdam: Benjamins.

Franz Rainer (Wirtschaftsuniversität Wien)
franz.rainer@wu-wien.ac.at

PARADIGMATIC FACTORS IN THE IRRADIATION OF ALLOMORPHY: THE REANALYSIS OF THE LATIN TYPE *MANUALIS* IN ITALIAN

Abstract

Latin derivatives in *-alis*, *-are* (infinitive), *-arius*, and *-osus* show a *-u-* between root and suffix if the root is a fourth declension noun: we thus have *manualis* from the fourth declension noun *manus*, but *fiscalis* from the second declension noun *fiscus*. Italian inherited a lot of these Latin bases and derivatives, but not the various declension classes. Since the distribution of *-u-* thus became opaque, Tekavčič 1980 hypothesizes that this allomorphic variation also became unproductive, a conjecture incompatible with the existence of quite a number of neologisms with *-u-* in Italian. A closer examination of the data shows that most neologisms are due either to local analogies based on existing derivatives with the same suffix or to the influence of co-derivatives, i.e. existing derivatives with the same root but a different suffix. In these cases, however, it is only the allomorphy that is determined by the co-derivative, while from a morpho-semantic perspective the derivatives follow the series of words with the same suffix.

1. Introduction

It is quite frequent, at least in the more familiar European languages, that an affix seems to be joined to a base not directly but through some intervening „element". In English, e.g., though the adjectival suffix *–al* is generally attached directly to the base noun, as in *organizational*, in other derivatives one may observe an intervening element such as *-in-* (e.g. *criminal*) or *-u-* (e.g. *habitual*). Such elements have received quite a number of different treatments in the literature, the main points of discordance being their morphemic vs. phonemic status, their proper segmentation and bracketing, their function, and, tightly linked to these different conceptions, terminology.[1] There seems to be more agreement about how such elements arise: as Malkiel 1958 has shown, they may be relics of earlier or foreign morphological systems or the result of reanalysis or contamination.

Once such elements occur in at least one word of a language, one may often observe that they spread to other words of that language, and it is this process of

[1] The most important contributions to these topics are the following: Malkiel 1958, Tekavčič 1968, Aronoff 1976: 98-114, Szymanek 1985, and Dressler / Merlini 1994: 529-557.

irradiation that will be at the center of the present investigation.[2] In the few occasions when Malkiel touches the problem of the irradiation of such elements, he invariably resorts to analogy as the underlying mechanism, a view that is also explicitly endorsed by Lázaro Carreter 1972 and Becker 1993: 206 (with respect to the linking elements of German compounds). Aronoff 1976: 112, on the other hand, describes English words such as the ones cited above in terms of „rules of allomorphy", i.e. rules that introduce an empty morpheme in the context of two other morphemes. More recently, Booij 1997 has drawn our attention to the fact that the choice of such elements may also be determined paradigmatically: thus the *-aan-* of the Dutch ethnic adjective *Amerikaans* ‚American' is taken from the inhabitant name *Amerikaan* ‚inhabitant of America' following a general rule of Dutch according to which ethnic adjectives are formed by joining *–s* to the corresponding inhabitant name. In the following analysis of the fate of the Latin type *manualis* in Italian we will see that both analogical models, i.e. existing words or groups of words with the same suffix, and co-derivatives, i.e. words with the same stem but a different suffix, may play a role in the irradiation of our elements.

2. The Latin type *manualis*

The conditions underlying the appearance of *-u-* in Latin derivatives are well-known and quite straightforward: it was applied, roughly, after bases of the fourth declension[3] and before the suffixes *-alis*, *-are* (infinitive ending), *-arius*, and *-osus*. There are only about twenty base nouns that do not pertain to the fourth declension, as well as some sporadic examples of suffixes different from the ones just mentioned, but on the whole the pattern was quite regular and productive.

The nouns of the Latin fourth declension may be grouped into three groups: action nouns, „status nouns" of the type *consulatus* ‚consulate', and a heterogeneous residual group. As table 1, based on Gradenwitz 1904 and Georges 1914, shows, action nouns were by far the largest group and produced the greatest number of derivatives with a *-u-*. The 80 status nouns interestingly did not give rise to any derivative with *-u-*, while more than half of the residual group did so.

morph. category	action nouns	status nouns	other nouns
example	*latratus*	*consulatus*	*mons* etc.
total number	785	80	59
derivatives with *-u-*	77	0	35

Table 1

[2] A more extensive version of the present analysis will appear in Italian in the proceedings of the annual meeting of the Società di Linguistica Italiana held in Padua in September 1997.

[3] Latinists also speak of „*-u-*stems", since the nouns of this declension have a *-u-* after the root in all case forms: nom.sg. *manus* ‚hand', gen.sg. *manuus*, nom.pl. *manuus*, etc. (*-uu-* here stands for a long vowel.) The *-u-* thus appears before the case endings and the four suffixes *-alis*, *-are*, *-arius*, and *-osus*, but not normally before other derivational suffixes.

Though, as we have seen, the presence of *-u-* in Latin was conditioned by the declension class of the base noun and the presence of one of the four suffixes mentioned above, it is interesting to note that there was also a strong co-variation between the appearance of *-u-* and the presence of a root final /t/ or /s/, due to the fact that the root of all action nouns of the fourth declension ended in /t/ or /s/ and that the residual group was largely due to analogical extension from the action noun group based on semantic, but also formal similarity (as in the classic example of *montuosus* ‚mountainous', from *mons*, *-tis* ‚mountain', which is held to have been formed on the model of almost synonymous *saltuosus*, from *saltus*, *-uus*).

3. Stratigraphic analysis of 152 Italian derivatives with *-u-*

As shown in table 2[4], many Latin derivatives with *-u-* were borrowed into Italian over the centuries. This stock of latinisms was then extended during the modern times by a number of gallicisms and anglicisms, many of which were not borrowed from Latin but formed by speakers or writers of French or English. And last but not least, there is a considerable number of derivatives, from the very beginning of the Italian language up to the present moment, which must be classified as „neologisms", i.e. formations on an independent Italian basis.

century	latinisms	gallicisms	anglicisms	neologisms
XIII	5			1
XIV	32			5
XV	5			
XVI	16	2		5
XVII	3	1		2
XVIII	5		1	2
XIX	8	7		8
XX		5	4	35

Table 2

It is this last group that is of particular interest to our concerns. How can there be Italian neologisms with a *-u-*, one may wonder, when one of the essential conditioning factors in Latin, the fourth declension, has disappeared in that language? And in fact Tekavčić 1980: §§ 1015.3 and 1040.4, the only diachronic manual that addresses the problem, holds that the choice of *-u-* has become opaque and hence unproductive in

[4] This table was compiled by checking all the Italian descendants from Latin words of the fourth declension in *DELI* and *GDLI*, as well as on a reverse word list based on *Zingarelli*, all available dictionaries of neologisms, and the author's personal collection of neologisms. The only formations that might have gone unnoticed are thus neologisms quoted in *GDLI* whose base does not go back to a Latin word of the fourth declension and which is not included in *Zingarelli* or any of the collections of neologisms.

Italian due to the disappearance of the fourth declension. But this view is clearly at odds with the results displayed in table 2. We will thus have to investigate which factors determine(d) the appearance of *-u-* in Italian neologisms.

4. Irradiation from analogical models

We have already seen that in Latin there has been, to a large extent, co-variance between the fourth declension and the presence of a root final /t/ or, less frequently, /s/. A first reasonable guess might thus be that Italians simply substituted this phonological conditioning to the morphological one. A first rough statistic indeed seems to corroborate this suspicion: 68 % of the Italian bases have a /t/ and 25 % an /s/ before the final vowel, while only 7 % have a consonant different from /t/ or /s/ in the same position. One might thus be tempted to formulate an Aronovian „rule of allomorphy" of the type: „Insert /u/ before the suffixes *–ale*, *-are*, *-ario*, and *–oso* if the base ends in /to/ or /so/."[5]

	normal derivative		derivative with *-u-*		
suffix	1 loan words	2 neologisms	3 loan words	4 neologisms	5 co-derivatives
X(u)ale			*attuale* XIV		
			contrattuale XIX		
			fattuale XX		
	dialettale XIX	*interlettale* XX	*aspettuale* XX	*brevettuale* XX	
			concettuale XIX	*oggettuale* XX	
			distrettuale XIV	*progettuale* XX	
			effettuale XVI	*soggettuale* XX	
			intellettuale XIV		
			conflittuale XX	†*affittuale* XVI	*affittuario* XVI
	†*affittale* XIV		†*fittuale* XVI	*relittuale* XX	
X(u)are		*ciabattare* XIX	*attuare* XV		
		contrattare XVI			
	architettare XVI	*banchettare* XVI	*effettuare* XVI	*eccettuare* XIII	
	brevettare XIX	†*concettare* XVI		†*affettuarsi* XIV	*affettuoso* XIV
	†*eccettare* XIII	*difettare* XIII			
	precettare XVII	*filettare* XVIII			
	traghettare XVI	*merlettare* XIX			
		picchettare XIX			
		sorbettare XVIII			
	complottare XIX	*biscottare* XVI			
		cazzottare XVIII			
		pizzicottare XIX			
		fluttare XX	*fluttuare* XIV		
		fruttare XIII	†*frutuare* XIV		
		†*luttare* XIII			

Table 3

A closer inspection of the data however reveals that such a solution would be too simplistic. Let us look, e.g., at the behaviour of Italian bases ending in /tto/ before the suffixes *–ale* and *–are*, as displayed in table 3. Our rule of allomorphy would predict that they should all uniformly have a *-u-*, at least all the neologisms. But this is manifestly not the case. We may observe, on the contrary, that the presence of *-u-* is dependent on the suffix: while most neologisms in *-ale* have a *-u-*, those in *-are* overwhelmingly do not show this insert.[6] One would thus at least have to modify the rule of allomorphy in the following way: „Insert *-u-* before the suffix *–ale* if the base ends in -/to/ or -/so/".

But even this restricted version turns out to be too powerful, as the data in table 4 show.[7] In fact, all the roots of table 4 end in /ss/ and should thus show *-u-* according to our revised version of the rule of allomorphy. Though this is the case for the large majority of the neologisms, closer scrutiny reveals that the presence of *-u-* seems to be determined not just by the root-final consonant but by the phonic form of the root starting from the stressed vowel. Thus, while the bases in -/esso/ consistently take the *-u-* after the model of *sessuale*, those where /ss/ is preceded by a vowel different from /e/ seem to avoid the insert. We must thus conclude that the presence of *-u-* in Italian neologisms is not determined by some general rule of allomorphy, but obeys a host of very parochial generalizations taking in consideration the suffix and the phonic material of the root from the stressed vowel rightwards. The irradiation of *-u-*, in other words, is mainly to be accounted for by local analogy.

	Xsale		*Xsuale*	
	1 loan words	2 neologisms	3 loanwords	4 neologisms
/e/			*sessuale* XIX	*accessuale* XX
				ascessuale XX
				congressuale XX
	†*processale* XVIII			*processuale* XIX
/i/	*abissale* XIX			
	affissale XX	*interfissale* XX		
/o/	*colossale* XVIII			
	paradossale XVI			

Table 4

[5] 90 % of the Italian bases end in -/o/ (which goes back to the Latin acc. ending *-u(m)*).

[6] The counter-examples to this generalization have a special explanation, which however need not concern us here.

[7] The first column of table 4 shows the vowel immediately preceding root-final /ss/.

Analogy here is mainly based on formal resemblance, as we have seen. Between *sessuale* ‚sexual' and the group of neologisms *accessuale* ‚concerning seizures', *ascessuale* ‚concerning abscesses', *congressuale* ‚concerning congresses', *processuale* ‚concerning trials', and even among the four neologisms, there is no semantic resemblance whatsoever, but they share the sequence /ess/ and the suffix *-ale*. Only in very few cases can one find a semantic base for the analogy besides the formal one. A case in point would be the recent nonce-formation *contornuale* ‚contextual', derived from *contorno* ‚context' after the model of the synonymous couple *contestuale* / *contesto*. But even here the fact that both bases begin with *con-* may have been helpful.

Why, one might ask, did and do speakers and writers of Italian resort to such parochial analogical strategies and not to a more general phonological conditioning as conjectured above? The reason for this is simple: Italian inherited from Latin not only many derivatives with *-u-* based on Latin fourth declension nouns, but also many derivatives without *-u-* based on formally similar bases from other Latin declensions, such as *fatale* (XIV; Latin *fatalis*, from *fatum* ‚fate'), *digitale* (XVI; Latin *digitalis*, from *digitus, -i* ‚finger'), *capitale* (XIV; Latin *capitalis*, from *caput, -itis* ‚head'), etc. Any sweeping generalization of the kind „Insert *-u-* after roots ending in /t/" was thus impossible: speakers and writers had and have to take into consideration more phonic material in order to identify groups of bases homogeneous with respect to the presence or absence of *-u-*.

Analogy, however, is not enough to explain *all* neologisms. In the remaining part of this paper, we will adress the question which other mechanisms were and are operative in the irradiation of *-u-*.

5. Irradiation from co-derivatives

Some neologisms seem to owe their *-u-* to the fact that some co-derivative, i.e. a derivative with the same stem but a different suffix, also has one. The arguments for the need of this additional mechanism are subtle and not always totally conclusive, but on the whole it would seem to me that the need for such an additional device is indisputable.

A first kind of argument is based on lexicographic practice. Italian lexicographers sometimes felt the need to justify the presence of *-u-* in certain derivatives, and in these comments they also resort to irradiation from co-derivatives. By explaining the nonce-formation *intellettuoso* ‚lit. intellectuous' as a derogatory variant of *intellettuale* ‚intellectual', Migliorini 1963, for example, seems to want to establish a direct relationship between these co-derivatives. Other lexicographers are more explicit. *GDLI*, e.g., notes with respect to C.E. Gadda's neologism *puntuare* ‚to gain in precision', derived from *punto* ‚point': „per influsso di *puntuale*" [i.e. influenced by *puntuale* ‚precise']. And in the *Novissimo Dardano* (Rome: Curcio) *sessuato* ‚sexuate' is said to be derived from *sesso* ‚sex' „con sovrapposizione di *sessuale*" [i.e. with superposition of *sessuale* ‚sexual']. The latter explanation is certainly incorrect from an historical perspective, since *sessuato* is a loan-translation of either French *sexué* or English *sexuate*, where the *-u-* is already present, but it is nevertheless interesting as an intuition of a native speaker.

The Italian lexicographers quoted in the last paragraph seem to have had in mind a kind of double derivation for words like *puntuare* or *sessuato*: while from a morpho-semantic point of view they are said to derive from their bases *punto* and *sesso* according to the general model of verbs in *-are* and adjectives in *-ato*, their allomorphic shape is attributed to the influence of the co-derivatives *puntuale* and *sessuale*. In other words, we would have here cases of paradigmatically determined allomorphy in the sense of Booij 1997.

This is not, however, the only possible interpretation. Scholars uneasy with this idea of double derivation might want to argue that these Italian words are directly derived from their co-derivatives, both semantically and formally. And, indeed, one can also find on some occasions comments pointing in that direction in Italian dictionaries. Under *puntuario*, e.g., an obsolete synonym of *puntuale*, *GDLI* writes: „Deriva da *puntuale*, con cambio di suff." [Derives from *puntuale*, with suffix change]. Sceptics could also point to the fact that affix-substitution is a relatively common process both in inflection and in word-formation.[8] It has been argued, e.g., that analogical levelling of the type Old French *je claime* ‚I cry out'/ *nous clamons* ‚we cry out', *j'aime* ‚I love'/ *nous amons* ‚we love' to Modern French *je clame / nous clamons*, *j'aime / nous aimons* should be analysed as based on affix-substitutions of the type *Xons* > *Xe* and *Xe* > *Xons*. And Faitelson-Weiser 1981: 214-5 has pointed to an interesting case in the derivational system of Mexican Spanish, where the irradiation of allomorphic *-c-*, which is inserted between bases of a certain phonic shape and the suffix, from diminutives in *-ito* to augmentatives in *-ote* can be shown to be due to the fact that *-ito* and *-ote* are in a productive antonymic relationship in that variety of Spanish, which allows speakers to form a corresponding derivative in *-ote* to almost any derivative in *-ito*. In the same fashion, one might be tempted to derive *puntuare* from *puntuale* by substituting *-are* to *-ale*, and similarly in the other cases.

Though I agree that affix-substitution is much more widespread than is generally believed, I do not think that it is the right solution for our Italian cases, at least not for the great majority of them. If *XA* is derived from *XB* by substituting *A* for *B*, then normally there is independent evidence that derivatives of the form *XA* should stand in a direct relationship to derivatives of the type *XB*. In inflectional paradigms, e.g., the development of mutual relationships among the different slots is only to be expected. And in the Spanish case the existence of a direct antonymic relationship between diminutives in *-ito* and augmentatives in *-ote* is proved, among other things, by their frequent co-occurence in texts. No such independent evidence, however, is available for our Italian derivatives.

Besides the ones quoted above, the following examples should also probably be considered to contain a *-u-* that originated from a co-derivative. As we have already seen in table 3, verbs in *-are* derived from bases in *-/etto/* normally do not take *-u-*. The only exception to this rule is Old Italian *affettuarsi* ‚to fall in love': morpho-semantically, it is derived from *affetto* ‚love', but the *-u-* certainly comes from the co-derivative *affettuoso* ‚loving'. The verbal nonce-formation *rituare* ‚to respond with stereotyped phrases' (*GDLI*), derived from *rito* ‚rite', probably owes its *-u-* to the co-

[8] Cf., among others, Marle 1985, Becker 1990 and 1993, Ford / Singh / Martohardjono 1997.

derivative *rituale* ‚ritual', even though one cannot totally exclude the possibility that its shape might be the result of a formal analogy to the couple *sito* ‚place' / *situare* ‚to locate'. The case where the irradiation of the *-u-* from a co-derivative is most obvious, however, is the long series of neologistic derivatives from *sesso* ‚sex': contrary to what we observe in English *sexology* and similar terms in other European languages, Italian *sessuologia* shows a *-u-*, which must have irradiated from the co-derivative *sessuale* ‚sexual' since there is no other Italian derivative in *-ology* with a *-u-* that might have served as an analogical model. The same behaviour, by the way, may be observed in most other neologisms on the basis of *sesso*: *sessuofobia, sessuomania, sessuomorfo, sessuofilico*, etc. What distinguishes the *sesso*-case from the others is that the irradiation of *-u-* goes even beyond the limits of the four suffixes *-ale*, *-are*, *-ario*, and *-oso*. It seems that Italian speakers have extracted a stem allomorph *sessu-* from *sessuale* which is now applicable before all suffixes beginning with a back vowel.[9]

The different behaviour of Italian and other European languages with respect to the irradiation of the *-u-* inside the sex-family suggests that it is not possible to formulate sufficient conditions under which such paradigmatic irradiation of allomorphy takes place. The same conclusion is also prompted by opposing to *sesso* / *sessuale* / *sessuologia* other Italian couples like *congresso* ‚congress' / *congressuale* / *congressomania*, *istinto* ‚instinct' / *istintuale* / *istintolatra*, *reddito* ‚income' / *reddituario* / *redditometro*, etc., where no irradiation of *-u-* may be observed under comparable circumstances. The best one can do, it seems, is to formulate necessary conditions for this kind of paradigmatic irradiation. Just as in analogical levelling, some kind of higher-order paradigmatic organization of morphological categories seems to be a precondition for paradigmatic irradition to take place. In the case of our *-u-*, there is an incipient paradigmatic network in the sense that the 152 Italian derivatives with *-u-* are not distributed randomly, but over about ninety Italian nouns and four suffixes.

6. Conclusion

This study of the irradiation of *-u-* in Italian has shown that it cannot be totally accounted for by analogy alone. As we have seen, allomorphy may also spread from co-derivatives.[10] Since we have argued against an affix-substitution account of the Italian facts, we have to conclude that the creation of a neologism may, if the word is integrated into a certain type of paradigmatic network, follow two models at the same time, one for the morpho-semantic side and one for the allomorphy.

[9] The limitation to back vowels is necessary in the light of *sessismo* ‚sexism' and similar derivatives.

[10] Things are even more complicated, since other Italian neologisms with *-u-* can only be explained as „virtual latinisms": Italian writers and speakers in several cases seem to have inserted a *-u-* only because they knew that the base noun originated from a Latin noun of the fourth declension and that derivatives from such nouns were supposed to have a *-u-*. Still another, marginal, source for *-u-* not treated here is contamination.

Bibliography

Aronoff, Mark (1976) *Word Formation in Generative Grammar*. Cambridge, Mass.: MIT Press.

Battaglia, Salvatore (1961ff.) *Grande dizionario della lingua italiana*. Turin: UTET.

Becker, Thomas (1990) *Analogie und morphologische Theorie*. Munich: Beck.

Becker, Thomas (1993) „Morphologische Ersetzungsbildungen im Deutschen", *Zeitschrift für Sprachwissenschaft* 12, 185-217.

Booij, Geert (1997) „Allomorphy and the Autonomy of Morphology", *Folia Linguistica* 31, 25-56.

Cortelazzo, Manlio / Paolo Zolli (1979-88) *Dizionario etimologico della lingua italiana*. Bologna: Zanichelli.

DELI = Cortelazzo / Zolli (1979-88).

Dressler, Wolfgang U. / Lavinia Merlini (1994) *Morphopragmatics*. Berlin-New York: Mouton de Gruyter.

Faitelson-Weiser, Silvia (1981) *Les suffixes quantificateurs de l'espagnol moderne*. Paris: Editions Hispaniques.

Ford, Alan / Rajendra Singh / Gita Martohardjono (1997) Pace *Panini. Towards a Word-Based Theory of Morphology*. New York: Peter Lang.

GDLI = Battaglia (1961ff.).

Georges, Karl Ernst (1914) *Ausführliches lateinisch-deutsches Handwörterbuch*. 8th ed. Reprint 1995: Hannover: Hahn.

Gradenwitz, Otto (1904) *Laterculi vocum latinarum*. Leipzig: Hirzel.

Lázaro Carreter, Fernando (1972) „¿Consonantes antihiáticas en español?", *Homenaje a Antonio Tovar ofrecido por sus discípulos, colegas y amigos*. Madrid: Gredos, 253-64.

Malkiel, Yakov (1958) „Los interfijos hispánicos", *in* Diego Catalán (ed.) *Miscelánea Homenaje a André Martinet*. II. *Estructuralismo e historia*. Universidad de La Laguna, 107-199.

Marle, Jaap van (1985) *On the Paradigmatic Dimension of Morphological Creativity*. Dordrecht: Foris.

Migliorini, Bruno (1963) *Parole nuove*. Milan: Hoepli.

Szymanek, Bogdan (1985) „On intermorphemic extensions in English and Polish", *in* Edmund Gussmann (ed.) *Phono-Morphology*. Lublin, 176-191.

Tekavčić, Pavao (1968) „Sull'analisi morfematica di un tipo di derivati italiani. Il problema dei cosiddetti interfissi", *Studia romanica et anglica zagrebiensia* 25-26, 69-85.

Tekavčić, Pavao (1980) *Grammatica storica dell'italiano*. III. *Lessico*. Bologna: Il Mulino.

Zingarelli. Bologna: Zanichelli 1995.

Anna M. Thornton (Università dell'Aquila & IP/CNR, Roma)
annathor@kant.irmkant.rm.cnr.it

STEM ALLOMORPHS OR SUFFIX ALLOMORPHS? ON ITALIAN DERIVATIVES WITH ANTESUFFIXAL GLIDES.

This paper examines Italian derivatives with the suffixes *-ale*, *-oso*, *-ario* and *-are* which display /j/ or /w/ between the stem of the base and the suffix. These derivatives are not generated by categorical rules. Rather, they conform to schemata of the sort of those discovered by Bybee & Slobin (1982). The bases of derivatives with /w/ conform to a source-oriented, phonologically and partially morphologically defined schema, which establishes as prototypical for the insertion of /w/ bases ending in C$t. Bases which insert /j/ can be captured with similar schemata, but a subpart of the derivatives which display /j/ between stem and suffix, those in *-iale*, can be described also as the result of a morphophonological or readjustment rule inserting /j/ between certain stem-final suffixes and *-ale*.

0. Introduction

This paper addresses a question of allomorphy in Italian. There is a set of Italian denominal adjectival suffixes, *-ale*, *-oso*, *-ario*, and one denominal verbal suffix, *-are*[1], which sometimes appear preceded by orthographic <i> or <u>, which can be realized phonologically as high vowels /i,u/ or glides /j,w/ (cf. (1a)). Following my own usage, I will consider these segments glides. Examples of words showing the alternance under discussion are given in (1b). At first sight it seems that the glides appear unpredictably in contexts (1bii-iii) which are virtually identical from the segmental point of view to the ones in which no glide appears (1bi). Our task is then to discover conditions that govern the appearance vs. absence of the glides in derivatives of this kind.

(1) Pretheoretical overview of the data

a.

Suffixes	
-ale/-iale/-uale	[[X]N + suffix]A
-oso/-ioso/-uoso	"
-ario/-iario/-uario	"
-are/-iare/-uare	[[X]N + suffix]V

b. Examples of derivatives[2]

-ale	i.	strad+a strad+ale	ii.	mond+o mond+i+ale	iii.	grad+o grad+u+ale
-oso	i.	ferr+o ferr+oso	ii.	mister+o mister+i+oso	iii.	mostr+o mostr+u+oso
-ario	i.	second+o second+ario	ii.	fond+o fond+i+ario	iii.	cens+o cens+u+ario
-are	i.	sched+a sched+are	ii.	distanz+a distanz+i+are	iii.	accent+o accent+u+are

I have collected a corpus of derivatives of the kind presented in (1bii-iii), by selecting all the derivatives of this kind present in a reverse dictionary of Italian[3], and in five dictionaries of Italian neologisms of the Eighties[4]. The number of words with each final shape in the corpus is shown in (2):

(2) Number of words with each final shape in the corpus

iale	uale	ioso	uoso	iario	uario	iare	uare
95	49	19	27	15	17	29	12

The data in (2) must be compared with the overall number of derivatives with each of the four suffixes. In fact, there are about 850 derivatives in *-ale*, about 540 in *-oso*, about 360 in *-ario*, and many denominal verbs in *-are*[5] in which these suffixes are not preceded by any glide. Clearly, the derivatives which display a glide between the root of the base and the suffix are the exception rather than the rule. We will now try to determine what conditions the appearance of the glides.

1. Derivatives with /w/

Historically, most of the bases of the derivatives with /w/ go back to Latin fourth declension nouns, whose stem ended in *-u*; there are, however, a number of derivatives from bases of other kinds, as the data in (3) show:

(3) Bases of derivatives with /w/ (total = 82)
Derivatives already existing in Latin are underlined.
Neologisms are in **boldface**.

a. Descendants of Latin fourth declension nouns (*-u* stems) (total = 48)
Examples:

accento 'stress'	-->	accentuale; accentuare 'to accentuate'
arco 'bow, arch'	-->	arcuare; arcuato 'to bend; bent, arched'
caso 'case'	-->	casuale 'casual'
lusso 'luxury'	-->	lussuoso 'luxurious'
porto 'harbour'	-->	portuale, portuoso, portuario

[1] A more precise citation form for this suffix would be *-0-a*, that is, a zero suffix which forms verbs from adjectives and nouns, and is associated to the first verbal conjugation, distinguished by the thematic vowel *-a-*; for convenience, this suffix will hereafter be referred to as *-are*, a citation form which is homophonous with the infinitive ending of the verbs which it forms; by no means it should be implied, though, that I take the infinitive (inflectional) ending as having category-changing power.

[2] Representation is in the standard orthography of Italian. Glosses are not provided for lack of space. The boundary symbol "+" is used pretheoretically in (1), just to show the result of a blind segmentation procedure.

[3] Based on the merging of Zingarelli (1983) and Garzanti (1987), two dictionaries of usage including around 100.000 words each. Courtesy of Tullio De Mauro.

[4] Quarantotto (1987), Cortelazzo & Cardinale (1989), Vassalli (1989), Forconi (1990), Lurati (1990).

[5] The data about the number of derivatives in *-ale*, *-oso* and *-ario* are drawn from an unpublished database constructed by Burani & Thornton, based on Ratti et al. (1989), a reverse dictionary based on an abridged dictionary of Italian, containing about 45.000 lemmata. It is hard to determine the number of denominal verbs in *-are* because in a reverse dictionary they are not set apart from underived verbs in *-are*. In BDVDB (Thornton, Iacobini, Burani 1997) they are 187 over a total of 1007 verbs in *-are*, that is 18,6%. Projecting from this percentage, I have estimated the number of denominal zero-suffixed *-are* verbs in a 45.000 lemmata dictionary at about 900.

b. Descendants of Latin nouns alternating between fourth and second declension (total = 8)
Examples:
tumulto 'riot' --> tumultuoso 'tumultuous'
punto 'point' --> puntuale 'punctual'

c. Descendants of Latin nouns of the second declension (*-o* stems) (total = 18)
Examples:
santo 'saint' --> santuario 'sanctuary'
mostro 'monster' --> mostruoso 'horrible'
talento 'talent' --> **talentuoso** 'talented'
delitto 'crime' --> delittuoso 'criminal'

d. Descendants of Latin nouns of the third declension (-C stems) (total = 4)
monte 'mount' --> montuoso 'mountainous'
ponte 'bridge' --> pontuale (rare)
volutta 'pleasure' --> voluttuoso, voluttuario
bound root *mens-* 'month' --> mensuale'monthly' (rare, antiquated)

e. New bases (not attested in Latin according to Lewis & Short)
brevetto 'patent' --> **brevettuale** (Rainer 1997)
contorno 'contour' --> **contornuale**
percento 'per cent' --> percentuale
rapporto 'relationship' --> **rapportuale**

These bases do not share any morphological property, except being almost all of masculine gender. We will comment later on the relevance of this feature.

From the segmental point of view, though, it is striking that the overwhelming majority of the bases (95,1%) has a root ending in a coronal anterior segment. However, a root ending in such a segment does not automatically yield the insertion of a /w/ glide before one of our suffixes, as there are over 300 derivatives in *-ale*, and about 100 derivatives in *-oso* and in *-ario*, from bases whose root ends in a coronal anterior segment which do not display a /w/ glide between root and suffix (examples: *fato* 'fate' > *fatale* 'fatal', *dialetto* 'dialect' > *dialettale* 'dialectal', *naso* 'nose' > *nasale* 'nasal', *cultura* 'culture' > *culturale* 'cultural', *spirito* 'humor' > *spiritoso* 'humorous', *sasso* 'stone' > *sassoso* 'full of stones'). The property of the root of ending in a coronal segment, then, seems to be at most a necessary, but by no means a sufficient condition, to yield a derivative with a /w/ between root and suffix. Therefore, it cannot be incorporated in a phonological rule or a readjustment rule.

I would like to propose that this condition functions rather as a schema, in the sense of Bybee and Slobin (1982) and Bybee and Moder (1983).

These authors introduce the concept of schema in relation to the past tenses of English irregular verbs. These past tense forms "are rote-learned and stored in the lexicon, but this does not prevent speakers from formulating generalizations about these forms. These generalizations are [...] [called] SCHEMAS. A SCHEMA is a statement that describes the phonological properties of a morphological class [...]. It is not a constraint which rigidly specifies what can and cannot occur, but it is rather a much looser type of correlation [...]" (Bybee & Slobin 1982:267).

Bybee & Slobin (1982:279) exemplify their notion of schema with a class of English verbs whose past tense has the shape in (4) :

(4) ...æŋ(k)] verb
past

They comment that "the schema defines a prototype of the category (in the sense of Rosch & Mervis 1975), in that *sing* or *drink* are the best exemplars -- but *swim* and *begin* may also belong to the category because they end in nasals, although not velar nasals" (Bybee & Slobin 1982:279).

The theory of Bybee and colleagues, then, is that a schema defines the prototype of a category that functions as a natural class, i. e. a schema defines/describes a class by referring to its prototype, which is defined on the basis of its phonological shape. The phonological shapes of the members of the class form a series of family resemblances rather than sharing a discrete set of features. The "most common and best exemplars" of the class conform to the schema, i.e. to the prototype.

Using this notion of schema, and the idea that morphological classes, like natural categories, can be defined not by a necessary and sufficient set of features but by their clustering around a prototype, we could say that there is a prototypic shape of the root of bases displaying the /w/ between root and suffix; bases whose root differs from the prototype by one feature only often still behave like the prototype, i.e. belong to the class of bases which insert a /w/, while progressively more distant bases display progressively less often the /w/ between root and suffix.

The prototype of bases displaying /w/ is shown in (5):

(5) Root ends in C$t
i.e., root ends in a Coda Consonant followed by a
[- Continuant, - Sonorant, - Voice, Coronal, +Anterior] C

Bases differing from the prototype by one or more features are schematized in (6):

(6) a. Root ends in C$s [+ Continuant]
b. Root ends in V$t no Coda Consonant
c. Root ends in V$s no Coda Consonant, [+ Continuant]
d. Root ends in C$n [+ Sonorant]
e. Root ends in V$n no Coda Consonant, [+ Sonorant]
f. Root ends in C$k not Coral
...

The number and percentage of bases of each type in the corpus are shown in (7).

(7)

Final shape	Number of bases	Percentage	Cumulative percentage
C$t	51	61,4%	61,4%
C$s	10	12,0%	73,4%
V$t	9	10,8%	**84,2%**
V$s	3	3,6%	
Other Coronal	6	7,2%	
Other	4	4,8%	

It is clear that the majority of bases that have a root allomorph ending in /w/ either correspond to the prototype or differ from it by only one feature.

A chi-square test was run to determine whether the distribution of the bases in the first three rows differed significantly from what could be expected given the number of bases with the relevant shape in the language. It turned out that the difference between the prototype and the C$s shape was not significant (χ^2 (1) = 0.69, p = n.s.), while the difference between the prototype and the V$t shape was highly significant (χ^2 (1) = 22,81, p <.001) .

This leads to the hypothesis (already put forward by Bybee & Moder) that there is a hierarchy among the features characterizing the prototype: having a closed syllable before the last segment is more important than having exactly /t/ as last segment. This means that prosodic conditions are more important than segmental conditions in defining the shape of the prototype.

Even more striking are the data which result from taking into account only the bases of the "new" derivatives with /w/, i. e. the bases of those derivatives that did not exist in Latin (according to Lewis & Short). These data are shown in (8).

(8) Number and percentage of new derivatives from bases with different final shapes

Final shape of the base	Number of derivatives	Percentage	Cumulative percentage
C$t	47	68,1%	68,1%
C$s	9	13,0%	81,1%
V$t	6	8,7%	**89,8%**
C$n	2	2,9%	
d	3	4,3%	
k	1	1,4%	

It seems that in the history of Italian the likelihood for a new derivative displaying a root allomorph with a final /w/ to come up was directly proportional to the closeness of the base to the prototype defined by the schema in (5), and decreased abruptly for bases whose root differs from the prototype by more than one feature.

The characteristics recognized by Bybee & Slobin in the schemas for English irregular past tenses, and considered by these authors as general characteristics of all morphological schemata, are the ones in (9):

(9) Characteristics of the schemas for English irregular past tenses
(a) Their defining properties are phonological and can range over more than one segment [...].
(b) Classes of items covered by schemas are defined in sets of family resemblances, not by sets of strictly shared properties [...] .
(c) Though schemas do not in themselves change features, they are used in lexical selection; and they may serve as the basis of new formations occasionally, either in speech errors [...] or in so called analogical formations [...].
(Bybee & Slobin 1982:285)

The schemata we have established to describe the class of bases that may insert a /w/ before one of the suffixes in (1a) have the characteristics (9b) and (9c); as far as characteristic (9a) is concerned, although the definition of our schema is primarily phonological, some morphological conditions seem to play a role: the bases should be masculine, and should not contain the suffix *-mento*, although it has the relevant phonological shape. The generalization in (10) holds both in the attested lexicon and in neologisms:

(10) -mento --> mentale, *-mentuale

e.g.	ornamento -->	ornamentale	*ornamentuale
	fondamento -->	fondamentale	*fondamentuale

There is one characteristic, however, which distinguishes the schema we have established to describe the set of bases that may undergo a /w/ insertion from Bybee & Slobin's schemata. They claim that schemata are product-oriented generalizations: one of their schemata "does not relate a base form to a derived one, as a rule does, but describes only one class of forms (the product class, in terms used by Zager 1980)" (Bybee & Slobin 1982:267). Our schema in (5), instead, is not product-oriented but base oriented (or "source-oriented", in Zager's terms): the relevant conditions are defined over the base and not over the derivative (or at least, there is no gain in defining them over the derivative).

So it seems that the study of the Italian bases which insert a /w/ before certain suffixes has led us to widen the concept of morphological schema, to include also source-oriented generalizations that define classes of bases rather than classes of outputs.

2. Derivatives with /j/

I have been able to collect 118 bases which have at least one derivative in which one of the suffixes in (1a) is preceded by /j/. There is no strong historical relation among these bases, comparable to belonging to the Latin IV declension for the bases of /w/ derivatives.

At first sight, it is striking that many of the bases of the derivatives inserting /j/ end in certain suffixes, as shown in (11):

(11) Number of bases in *-tore, -anza, -enza* which have one or more /j/ derivatives

-tore	(deverbal suffix forming agent nouns)	15
-anza	(deadjectival/deverbal suffix forming quality nouns)	5
-enza	"	48

So we could hypothesize that at least in these cases it is the suffix in the base which is responsible for the glide insertion, and we could try to write a readjustment rule inserting /j/ after these suffixes. But there are two problems with this account. In the first place, about half of the derivatives with /j/ would remain unexplained, as they do not come from bases with these suffixes; some examples are given in (12):

(12) *ministero* 'ministry' > *ministeriale* 'ministerial', *mondo* 'world' > *mondiale* 'world-wide', *razza* 'race' > *razziale* 'racial', *grande* 'big' > *grandioso* 'grand', *umile* 'humble' > *umiliare* 'to humiliate', *terzo* 'third' > *terziario* 'tertiary'

More crucially, there are counterexamples to the generalization that bases with these suffixes yield a derivative with /j/. These counterexamples are shown in (13):

(13) Derivatives from bases in *-anza, -enza, -tore* without /j/

-ale: semenzale, influenzale, pastorale, elettorale, dottorale[6]
-oso: burbanzoso, baldanzoso, speranzoso, **vacanzoso**
-ario: scadenzario
-are: burbanzare, fidanzare, speranzare, quietanzare, incombenzare, piacenzare, cadenzare, scadenzare, mordenzare, agenzare, urgenzare, influenzare, addottorare

Therefore, it is not possible to predict the occurrence of /j/ before one of our suffixes on morphological grounds, at least if we want to predict the distribution of /j/ in the attested lexicon.

Nevertheless, before abandoning the hypothesis of a morphological condition on the appearance of /j/, let us revise the quantitative distribution of this glide with bases in *-anza*, *-enza* and *-tore* in the attested lexicon (excluding neologisms). The number of derivatives with and without the glide for each base and suffix combination is shown in (14).

(14) Number of derivatives with and without /j/ for each base and suffix combination

-anza	*-ale*	*-ario*	*-oso*	*-are*
with /j/	4	1	1	5
without /j/	-	-	4	4
-enza	*-ale*	*-ario*	*-oso*	*-are*
with /j/	42	8	5	7
without /j/	2	1	-	8
-tore	*-ale*	*-ario*	*-oso*	*-are*
with /j/	14	-	-	-
without /j/	2	-	-	1

We will come back to these data in section 4.

As far as the data observable in the attested lexicon are concerned, the best way of accounting for derivatives with /j/ between stem and suffix seems to be the hypothesis of the existence of phonologically defined schemata. There are two schemata around which the majority of the derivatives with /j/ cluster. They are shown in (15).

(15) Schemata of the bases which yield /j/ derivatives

a. **"ts" schema**
Prototype: Root ends in C$ts,
i.e., Root ends in C$C [- Sonorant, - Voice, Coronal, +Anterior]
/\
[- Continuant] [+ Continuant]

[6] To these, we can add the couple of synonyms *rettorale/rettoriale* (observed by Sgroi 1997).

Shapes differing from the prototype by one feature:

Root ends in C$t lacks the [+ Continuant] C
Root ends in C$s lacks the [- Continuant] C
Root ends in V$ts lacks the Coda Consonant

b. **"r" schema**
Prototype: Root ends in V$r, i.e., Root ends in an open syllable followed by a [+ Sonorant, + Continuant, + Voice, Coronal, +Anterior, - Lateral] C

Shapes differing from the prototype by one feature:

Root ends in V$l [+ Lateral]
Root ends in C$r Coda Consonant

The table in (16) shows the number and percentage of bases with each final shape.

(16)

Final shape	Number_of bases	Percentage	Cumulative percentage
a. C$ts	55	46,6%	46,6%
C$t	8	6,8%	53,4%
C$s	2	1,7%	**55,1%**
b. V$r	29	24,6%	24,6%
V$l	6	5,1%	29,7%
C$r	2	1,7%	**31,4%**
			86,5%

A chi-square test showed that the difference between the number of bases in the prototype and in the second and third row in (16a) respectively is highly significant (C$ts vs. C$t: χ^2 (1) = 129,73, p < .001; C$ts vs. C$s: χ^2 (1) = 20,47, p <.001).

The table in (17) shows the number and percentage of new derivatives with /j/ from bases of different final shapes:

(17)

Final shape	Number_of derivatives	Percentage	Cumulative percentage
a. C$ts	77	62,6%	62,6%
C$t	1	0,8%	63,4%
C$s	1	0,8%	64,2%
V$ts	1	0,8%	**65,0%**
b. V$r	24	19,5%	19,5%
V$l	6	4,9%	24,4%
C$r	2	1,6%	**26,0%**
			91,0%

Here again, bases conforming to the prototype are the vast majority, and prosodic conditions prove more important than segmental ones in defining the prototype.

Now that we have seen the schemata at work with derivatives which display a /j/, we can compare them with the schema at work with /w/ derivatives.

The interesting thing is that there are certain phonological configurations that could be marginal members of classes defined by two different schemas, yielding different glides. For example, roots ending in C$s differ from roots ending in C$t (prototype of the class defined by the schema in (5)) only in the feature [+continuant], and from roots ending in C$ts (prototype of the class defined by the schema in (15a)) by the lack of the [-continuant] unit. The prediction, in such a case, would be that we should find, at least occasionally, derivatives from bases of these shapes with both glides. This prediction is borne out by data such as the ones in (18):

(18)	C$s root	derivative with /j/	derivative with /w/
	asse 'axis'	*assiale* 'axial'	–
	sesso 'sex'	–	*sessuale* 'sexual'

3. Productivity

Let us consider now the neologisms formed with our suffixes, to see whether the glides appear with new derivatives, besides appearing in part of the attested lexicon.

The formation of denominal verbs with a -O-*are* suffix seems to be slightly, if at all, productive in contemporary Italian (cfr. Iacobini & Thornton 1992). Therefore, we will concentrate on the three adjectival suffixes.

All three are productive in a Schultinkian sense, i. e. new words are formed with them in contemporary Italian. For *-oso* and *-ario*, though, the number of neologisms from bases that match the prototype of one of the schemata we have defined are so few as to render almost inconclusive the evidence. With *-ario* there are no neologisms with any of the glides, and a couple of glideless neologisms from bases that correspond to a schema: *eccedentario, carcerario*. With *-oso*, there is one neologism with /w/, *talentuoso*, and no glideless neologism from bases that match the C$t schema; the schemata for /j/, on the contrary, seem inactive, as there are glideless neologisms from bases that match them both: *vacanzoso, incazzoso, caciaroso, ceroso, paperoso*, the last three, crucially as we will see, from feminine bases. It is with *-ale* that derivatives with a glide from bases that match one of the schemata are encountered more frequently: there are 6 *-uale* neologisms from C$t bases (*rapportuale, gestuale, fattuale, oggettuale, progettuale, conflittuale*), 9 *-iale* neologisms from C$ts bases (*adolescenziale, dirigenziale, tangenziale, emergenziale, coscienziale, consulenziale, demenziale, differenziale, vertenziale*), and 8 *-iale* neologisms from V$r bases (*manageriale, teenageriale, datoriale, amatoriale, genitoriale, monitoriale, settoriale, autoriale*). There is also an occasional extension of the /j/ to a non prototypical base (*Bardot --> bardottiale*). Counterexamples (i.e., neologisms without a glide from bases which match one of the schemata) exist, but are few: some derivatives from feminine bases (*giuntale, figurale, congiunturale*) and *cantautorale*, from the blend *cantautore* 'singer-author'. It is possible that masculine gender is a morphological condition to be added also to the definition of the prototype of the V$r schema, as derivatives from feminine bases that match the phonological definition of this schema fail to display the glide

also with *-oso*, as we have seen. Neologisms in *-iale* from bases of the V$r shape are all from nouns ending in the Agentive/Instrumental suffix *-tore* or from English bases with the comparable suffixes *-er*, *-or*, (*manageriale, teenageriale, monitoriale*); neologisms in *-iale* from bases of the C$ts shape are all from bases in *-enza* .

4. Concluding remarks

To conclude, let me summarize my findings and the analysis which I propose for the two sets of data we have seen.

In the case of /w/ insertion, I think the best analysis is that certain bases, those which correspond to a prototype described by the phonologically and morphologically definable, base-oriented schema in (5), may have or develop a root allomorph ending in /w/, which is employed in the derivation of adjectives in *-ale* and occasionally in *-oso*, and which is observable in the attested lexicon also in derivatives in *-ario* and *-are*. Productivity is scanty (only 7 neologisms), as expected with morphological processes regulated by a schema rather than by a rule.

Bybee & Moder, following Rosch, call our attention to the factor of 'cue validity' as predictor of the productivity that a morphological class defined by means of a schema can attain.

According to Rosch, "cue validity is a probabilistic concept: the validity of a given cue X as a predictor of a given category Y [...] increases as the frequency with which cue X is associated with category Y increases, and decreases as the frequency with which cue X is associated with categories other than Y increases" (Rosch 1978:30).

In the case of the category of bases which display a root allomorph ending in /w/, the cue validity of the schema we have established is very low, as most of the bases that match the prototype in the language do not in fact display a glide-final root allomorph. There are almost 2000 masculine bases whose root ends in C$t, and only 51 have a root allomorph with final /w/.

So the schema for root allomorphs with final /w/ is not very productive, but, as some of the schemata discovered by Bybee and colleagues, it can "serve as the basis of new formations occasionally" (Bybee & Slobin 1982:285, cfr. (9) above). There are in fact a few neologisms in *-uale* and *-uoso* from bases that conform to the prototype, and in a pilot test I have been able to elicit oral production of *-uale* derivatives from bases which do not have an established adjectival derivative in the language and whose phonological shape conforms to the schema in (5).

As far as cases in which a /j/ appears, different analyses are possible. If we want to take into account all the attested lexicon, the analysis will be parallel to the one offered for /w/ derivatives: certain bases, conforming to one of the two phonologically defined schemas in (15), may have a root allomorph ending in /j/, which is employed in derivatives with one of the suffixes in (1a).

But another analysis is possible for the derivatives in *-iale*: if we go back to the data in (14), we can see that the presence of one of the three suffixes *-anza*, *-enza* and *-tore* has high cue validity in predicting that a derivative in *-ale* will display a /j/ before this suffix, as there are very few counterexamples to this generalization. So if we hypothesize the existence of morphologically defined schemata such as the ones in (19), which define a base containing one of these suffixes as prototypic for derivatives in -/jale/, such schemata would have a high cue validity, contrary to the low cue validity of the purely phonologically defined schemata in (15).

(19) Morphologically defined schemata for derivatives in *-iale*

a. Base ends in *-anza*
b. Base ends in *-enza*
c. Base ends in *-tore*

The morphologically defined schemata in (19) are not mutually exclusive with the phonologically defined ones in (15). Of course there is overlapping between the sets of bases captured by the morphologically defined schemata in (19a-b) and the phonologically defined schema in (15a), and by the schema (19c) and the schema (15b). But the interesting point is that the two sorts of schemata have different cue validity: this is quite low for the phonologically defined schemata in (15), but quite high for the morphologicaly defined ones in (19).

According to Bybee's (1988) approach, in which the difference between rules and schemata is not qualitative, but purely quantitative, in that "rules are highly reinforced representational patterns or schemata" (Bybee 1988:135), we would predict that a "highly reinforced" schema, i.e. a schema with high cue validity, such as the ones in (19), is almost non-distinct from a rule. And in fact, this is the case: remember that all but one of the neologisms with *-ale* from bases conforming to the morphologically defined schemata in (19) display the */j/* and conversely, only one neologism displaying the */j/* (*bardottiale*) is not derived from a base defined by one of the schemata in (19). This categorical behaviour is typical of a productive word formation rule.

So, if we do not aim at generating all the attested lexicon, but limit our aim to the characterization of productive processes only, the establishment of a morphological condition is possible. As we have seen, */j/* appears in neologisms only in derivatives with *-ale* from bases ending in the suffixes *-enza*, *-tore* or English *-er*, *-or*. In this case, we might analyze the data both as cases of morphologically governed base allomorphy (as in (20a)) or of morphologically governed suffix allomorphy (as in (20b)):

(20) Two possible analyses for the appearance of */j/* in neologisms with *-ale*

a. +tor]$_N$ --> +tori]$_N$ /__+ale]$_A$
   +enz]$_N$ --> +enzi]$_N$ /__+ale]$_A$
   ...

b. +ale]$_A$--> +jale]$_A$ / { +tor]$_N$, +enz]$_N$, +er]$_N$, +or]$_N$ } +_____

The decision between analyses (20a) and (20b) is up to a point arbitrary, as both correctly describe the facts.

Analysis (20a) would be preferred on historical grounds, as the source for the observed allomorphy is in the fact that the Latin suffix *-entia* yielded derivatives in *-entialis* and the Latin suffix *-torius* (*+tor+ius*) yielded derivatives in *+torialis*, which then formed the model for analogical creations in *-enziale*, *-toriale* from Italian bases in *-enza*, *-tore*.

Analysis (20b) would be preferred on economic grounds, as it reduces the number of allomorphic entities in the language (only the suffix *-ale* would have an allomorph, vs. the four suffixes *+tore*, *+enza*, *+er* and *+or*). But in the age of morphology by itself it is not clear whether the reduction of allomorphy must be pursued as the most desirable outcome.

I believe the decision will have to be taken on more general grounds, considering also the tendency in allomorphic phenomena in other parts of the Italian language and in other languages.

References

Bybee, J.L., 1988, "Morphology as Lexical Organization", in M. Hammond & M. Noonan (eds.), *Theoretical Morphology*, San Diego et alibi, Academic Press, 1988, 119-141.

Bybee, J.L. & Slobin, D.I., 1982, "Rules and schemas in the development and use of the English past tense", *Language*,58, 265-89.

Bybee, J.L. & Moder, C.L., 1983, "Morphological classes as natural categories", *Language*, 59, 251-270.

Cortelazzo, M., e Cardinale, U., 1989, *Dizionario di parole nuove 1964-1987*, Torino, Loescher.

Forconi, A., 1990, *Dizionario delle nuove parole italiane*, Milano, SugarCo.

Garzanti, 1987 = *Il grande dizionario Garzanti della lingua italiana*, Milano, Garzanti.

Iacobini, C. & Thornton, A.M., 1992, "Tendenze nella formazione delle parole nell'italiano del ventesimo secolo", in *Linee di tendenza dell'italiano contemporaneo*, a cura di B. Moretti, D. Petrini & S. Bianconi, Roma, Bulzoni, 25-55.

Lewis & Short = *A Latin Dictionary*, Oxford, OUP, 1984.

Lurati, O., 1990, *3000 parole nuove. La neologia negli anni 1980-1990*, Bologna, Zanichelli.

Quarantotto, C., 1987, *Dizionario del nuovo italiano*, Roma, Newton Compton.

Rainer, F., 1997, "Le condizioni d'impiego dell'allomorfo in *-u-* (tipo: *man-u-alis*) in latino e in italiano", *Bollettino della Società di Linguistica Italiana (SLI)*, XV,2, 77-79.

Ratti, D., Marconi, L., Morgavi, G., Rolando, C., (a cura di), 1989, *Flessioni, rime, anagrammi: l'italiano in scatola di montaggio*, Bologna, Zanichelli.

Rosch, E., 1978, "Principles of categorization", in E. Rosch & B.B. Lloyd (eds.), *Cognition and Categorization*, Hillsdale, NJ, Erlbaum, 27-48.

Rosch, E. & Mervis, C.B., 1975, "Family resemblances: Studies in the internal structure of categories", *Cognitive Psychology*, 7, 573-605.

Sgroi,S.C., 1997, "I derivati in *-iale* e in *-uale*: dai prestiti lessicali all'induzione di morfema", *Bollettino della Società di Linguistica Italiana (SLI)*, XV,2, 95-97.

Thornton, A.M., Iacobini, C., Burani, C., 1997, *BDVDB. Una base di dati per il Vocabolario di base della lingua italiana*, Roma, Bulzoni.

Vassalli, S., 1989, *Il neoitaliano. Le parole degli anni Ottanta*, Bologna, Zanichelli.

Zager, D., 1980, *A real-time process model of morphological change*, Buffalo, SUNY Dissertation.

Zingarelli, N., 1983, *Vocabolario della lingua italiana*, XI ed., Bologna, Zanichelli.

Section III
Compounding

Loren A. Billings (Leipzig)
billings@uni-leipzig.de

MORPHOLOGY *AND* SYNTAX: DELIMITING STUMP COMPOUNDS IN RUSSIAN

Determining the degree to which syntax is involved in morphological compounds—the central issue of the Mytilene conference's session on compounding—is elucidated by so-called stump compounds in Russian. This paper uses two particularly interesting case studies to illustrate stump compounds' hybrid properties.[1]

This paper consists of four parts: First I discuss the properties distinguishing stump compounds. Then I present two case studies which show that some stump compounds involve internal case-assignment: In the first case study the stump assigns quirky case to the following part of the compound. The second study shows that a stump compound and its non-stump counterpart correspond to differing syntactic case-assignment. In the final section I suggest, using colloquial data, that syntactic relations within stump compounds are unstable and this situation is rectified in one of two ways: Either the stump starts to act as a full-fledged word or the syntactic relation disappears.

1. Background, definitions, diagnostics

I begin by distinguishing between two types of truncative morphology in Russian.[2] Following Ward (1965:156-63), I refer to the examples in (1) as STUMP COMPOUNDS, in which the first of two stems is shortened (usually) to its first syllable.[3]

---

1 Most of this paper originally appeared in Billings (1995) and was presented in lectures at Leipzig and Princeton, and as a talk at the first Mediterranean Conference of Morphology. I express my appreciation to the audiences at all of these venues, especially to G. Adassovsky, L. Babby, T. Beyer, D. Carius, B. Cetnarowska, R. Cleminson, G. Corbett, A.-M. Di Sciullo, L. Downing, S. Franks, S. Harves, A. Israeli, G. Jarema, U. Junghanns, E. Komar, R. Leed, A. Ralli, K. Robblee, G. Rowicka, M. Schoorlemmer, A. Spencer, A. Thornton, M. Vigario, M. Yadroff, G. Zybatow and the Slavic and East European Languages e-mail list (seelangs@cunyvm.cuny.edu). Any shortcomings in this paper are solely my own responsibility.

2 As Andrew Spencer has pointed out to me, stump compounds (and perhaps even clips) might better be referred to not as WORD-FORMATION, but as WORD-CREATION. This is because such forms are generally coined not spontaneously, but by bards, advertising executives, authors and the like. I admit that while their coinage might be restricted in this way, stump compounds' properties are quite rigid prosodically. That is to say, there is a real grammar constraining their production.

3 The following special notations are used in this paper. Abbreviations: ACC(usative), ADJ(ective), ADPAUC(al), DAT(ive), F(eminine), GEN(itive), INFIN(itive), INST(rumental), LOC(ative), M(asculine), N(oun), NOM(inative), NT: neuter, NUM(eral), SG: singular, P(reposition), PL(ural). Stress notation: Main word stress is shown in ALL-CAPS; secondary stress, in SMALL CAPS; other syllables, in plain text. Truncated (either stumped or clipped) words are underlined. Unless otherwise marked, the form is in the NOM and SG. Finally, a bullet (•) indicates the boundary within complex word-formations.

(1) Stump compounds

a.	gorodSKOJ	soVET	GOR•soVET	[GOR.sʌ.V^{j}ET]
	city$_{(ADJ)M}$	council$_{(N.M)}$	'city council'	
b.	informaciONnoe	bjuRO	inFORM•bjuRO	[in.FORM.b^{j}u.RO]
	information$_{(ADJ)NT}$	bureau$_{(N.NT)}$	'information agency'	
c.	kommunisTIčeskaja	PARtija	KOM•PARtija	[KOM.PAR.t^{j}i .jə]
	communist$_{(ADJ)F}$	party$_{(N.F)}$	'communist party'	
d.	komanDIR	polKA	KOM•polKA	[KOM.pʌl.KA]
	commander$_{(N.M)}$	regiment$_{(N.M)GEN}$	'regimental commander'	
e.	koMANDnyj	soSTAV	KOM•soSTAV	[KOM.sʌ.STAF]
	command$_{(ADJ)M}$	composition$_{(N.M)}$	'command personnel'	
f.	poloVIna	GOda	POL•GOda	[POL.GO.də]
	half$_{(N.F)}$	year$_{(N.M)GEN}$	'half (a) year'	
g.	profsoJUZnoe	soBRAnie	PROF•soBRAnie	[PROF.sʌ.BRA.n^{j}i.jə]
	trade-union$_{(ADJ)NT}$	meeting$_{(N.NT)}$	'trade-union meeting'	
h.	xoZJAJstvennyj	rasČËT	XOZ•rasČËT	[XOS.rʌš^{j} ,Š^{j}OT]
	household$_{(ADJ)M}$	calculation$_{(N.M)}$	'self-supporting basis'	
i.	upraVLJAjuščij	deLAmi	uPRAV•deLAmi	[u.PRAF.di^{e}.LA.m^{j}i]
	manager$_{M}$	affairs$_{(N.NT)INST.PL}$	'on-site manager'	
j.	zaVEdujuščij	laboraTOrijej	ZAV•laboraTOriej	[ZAF.lə.bə.rʌ.TO.r^{j}i.jij]
	director$_{M}$	laboratory$_{(N.F)INST}$	'laboratory director'	

Stress notation follows Borunova *et al.* (1989). Cf. Hamilton (1980) and the references therein regarding the the phonetics of Russian.

The forms in (2), which I will simply call CLIPS, are another type of abbreviation.

(2) Clips

a.	parTIJnyj	organiZAtor	part•ORG	[pʌr.TORK]
	party$_{(ADJ)M}$	organizer$_{(N.M)}$	'(political-)party organizer'	
b.	polnoMOČnyj	predstaVItel'	pol•PRED	[pʌl.PR^{j}ET]
	plenipotentiary$_{(ADJ)M}$	representative$_{(N.M)}$	'ambassador plenipotentiary'	
c.	profsoJUZnyj	organiZAtor	prof•ORG	[prʌ.FORK]
	trade-union$_{(ADJ)M}$	organizer$_{(N.M)}$	'trade-union organizer'	
d.	proizvoDItel'	raBOT	pro•RAB	[prʌ.RAP]
	producer$_{(N.M)}$	operations$_{(N.F)GEN.PL}$	'construction superintendent'	
e.	SEL'skij	korresponDENT	SEL'•KOR	[S^{j}EL^{j}.KOR]
	rural$_{(ADJ)M}$	correspondent$_{(N.M)}$	'rural correspondent'	
f.	ŠKOL'nyj	raBOTnik	ŠK•RAB	[ŠKRAP]
	school$_{(ADJ)M}$	staffer$_{(N.M)}$	'schoolteacher'	
g.	sociAL'naja	straXOVka	SOC•STRAX	[SOTS.STRAX]
	social$_{(ADJ)F}$	insurance$_{(N.F)}$	'social insurance'	
h.	upraVLJAjuščij	DOmom	uprav•DOM	[u.prʌv.DOM]
	manager$_{M}$	house$_{(N.NT)INST.PL}$	'manager of a block of flats'	
i.	xoZJAJstvennyj	magaZIN	XOZ•MAG	[XOS.MAK]
	household$_{(ADJ)M}$	store$_{(N.M)}$	'household(-goods) store'	

The descriptive properties in (3a-d) distinguish stump compounds (1) from clips (2)—as well as from non-truncating compounds:

(3) Properties of stump compounds
- a. The final stem is **not** truncated; any non-final stem is truncated to a "stump".
- b. Stumps are invariably consonant-final; if this is an obstruent, then it is devoiced.
- c. Stumps are monosyllabic if consonant-initial and disyllabic if vowel-initial.
- d. Stumps bear secondary stress (shown in SMALL CAPS in this paper).

As the following discussion shows, these properties all point to a minimal-PrWd template for stumps, whereas no such template applies consistently to clips.

First, note that the wording in (3a) allows for stump compounds with more than one non-final stump, so long as the last stem appears in full. The example in (4) consists of three stems: two stumps (*sov•* and *inform•*) followed by the full word (*bjuro*). Neither *sov•* nor *inform•* can be free-standing words in Russian (with these meanings).

(4)

soVETskoe	informaciONnoe	bjuRO	sov•inFORM•bjuRO
[soviet	information]$_{(ADJ)NT}$	bureau$_{(N.NT)}$	[SOF.in.FORM.b^{j}u.RO]

I return shortly to the disyllabicity of *inform•* in (1b) and (4).

The observation in the first clause of (3b), that stumps are consonant-final, distinguishes stump compounds from other kinds of compounds, which take a linking vowel between the stems. An example of such a non-stump compound is *sever•o•vostok* 'northeast'. The criterion in (3b) also distinguishes between stump compounds and clips; cf. (2d) *pro•rab*, in which the initial stem is truncated to vowel-final *pro•*. As the second part of (3b) states, any stump-final obstruent must be [–voice]. Russian has **PrWd**-final devoicing; other languages, such as German, have **syllable**-final devoicing. This is additional evidence that stumps are PrWds.

The syllabic criterion in (3c), which is invariably true of stumps, does not always hold for clips. One example, *šk•rab* (2f), has the first stem truncating to non-syllabic *šk•*. Furthermore, stumps not only must be syllabic, they conform to a specific number of syllables—monosyllabic if consonant-initial and disyllabic if vowel-initial.[4] As (1b) *inform•bjuro* and (1i) *uprav•delami* show, there can occasionally be disyllabic stumps. This is because of an apparent correlation between disyllabicity and initial onsetlessness. Similar templatic phenomena, where certain morphological operations

[4] I am grateful to Anna Thornton for pointing out this correlation to me. In an earlier version of this paper I incorrectly characterized the following truncated form as a stump compound: *in•otdel* (shortened from *inostrannyj otdel*) 'foreign(-affairs) section'. I have not been able to find this particular truncated form in any dictionary. A similar word is listed in Borunova *et al.* (1989) **without** secondary stress on the initial, truncated stem: *In•turist*, truncated—according to Kramer (1965:145)—from *Vsesojuznoe akcionernoe obščestvo po inostrannomu turizmu v SSSR* 'All-USSR Joint-Stock Company for **Foreign Tourism**'. I am unable to conclude whether these forms bear secondary stress on *in•*. (As the discussion below in this paper shows, distinguishing between unstressed and secondary-stressed syllables headed by high vowels is very difficult.) I preliminarily conclude, based on the judgments in Borunova *et al.* (1989) regarding *In•turist*, that neither *in•otdel* nor *In•turist* is a stump compound. What then are these two forms? All clips that I'm aware of truncate the final stem. In any event, this issue is orthogonal to this paper's main issue: whether there can be syntactic reltions within morphological structures.

such as PrWd-reduplication require an onset, are discussed in Downing (1998). While there is disagreement in the prosodic-morphology literature about why an onsetless initial syllable is deficient, there is a consensus (based on numerous language families) that certain special morphological phenomena require an onset in the resulting form.

Finally, according to (3d), each stump has a secondary-stressed syllable; the final untruncated stem bears the compound's main stress. This is a property shared with other compounds; cf. *sever•o•vostok* 'northeast' [sʲɛ.vʲi.rə.vʌ.STOK]. With regard to the clips listed above, only (2e, g, i) have secondary stress on the non-final member. Generally speaking, secondary stress is limited to compounds in Russian, but there are exceptions. One such example is the non-compound borrowing in (5a):

(5) Non-compound words and secondary stress

a.	kONgeniAL'nyj	[kɔn.gʲi.nʲi.ALʲ.nyj]	'congenial'
b.	konfronTAcija	[kən.frʌn.TA.tsy.jə]	'confrontation'

Perhaps the most reliable test for whether a syllable bears stress—either secondary or primary—is the roundness of underlying /o/. As (5a) and several of the examples in (1) show, secondary stress on /o/ maintains lip-rounding. The example in (5b)—also a borrowing beginning in the same /kon-/ prefix—does not have secondary stress.[5]

The exact reduction of unstressed /o/ takes two forms after a non-palatalized consonant: In the syllable immediately preceding the main stress /o/ reduces to [ʌ], and elsewhere /o/ reduces to [ə]. See the first two syllables of (5b). What is important for these purposes is that unstressed /o/ loses lip-rounding. Thus, while secondary stress is *required* of stumps, with clips and other non-compounds secondary stress is *possible*.

I should add that some older stump compounds have become re-analyzed as simplex stems. Examples of this are shown in (6a-b):

(6) Stump compounds which have lost the internal morphological boundary

a.	professioNAL'nyj	soJUZ	profsoJUZ	[prəf.sʌ.JUS]
	professional$_{(ADJ)M}$	union$_{(N.M)}$	'trade union'	
b.	podVODnaja	LODka	podLODka	[pʌd.LOT.kə]
	under-water$_{(ADJ)F}$	boat$_{(N.F)}$	'submarine'	

Note the reduction of the former stumps' /o/ vowels in (6a-b) to [ə] and [ʌ], and the non-devoicing of the stump-final obstruent [d] in (6b). The lexical representations of (6a-b) should therefore not include a morpheme boundary. Note as well that the adjective formed from this erstwhile stump compound in (6a), /profsojuzn-/, can be further truncated into a stump, as in (1g) *prof•sobranie*, which doesn't mean 'professional meeting', but rather 'trade-union meeting'. The stump *prof•* can mean both 'trade-union', as in (1g), or 'professional', as in (11) below. Data like (6a-b) are

5 Bożena Cetnarowska has suggested to me that the contrast between (5a-b) might have to do with the number of syllables between /kon-/ and the main stress. In Polish primary stress is on the penult; secondary stress appears on any odd-numbered syllable from the beginning of the word as long as one unstressed syllable intervenes (e.g., *RE.vo.LU.cjo.NI.sta* 'revolutionary-NOM' and *RE.vo.LU.cjo.ni.STA.MI* 'revolutionaries-INST'; not **RE.vo.LU.cjo.NI.STA.mi*); see Rubach and Booij (1985) for details. In Russian the distribution of secondary stress within non-compounds appears to be restricted to prefixes separated from the main stress by at least two syllables.

not a problem for the proposed analysis, because compounds are often re-analyzed over time as simplex stems. One contributing factor in the re-analysis of (6b) might be that *pod*, the erstwhile stump, is itself the proclitic prefix (which means 'under'); I mention below that proclitic prepositions don't take secondary stress.

The properties listed in (3b-d)—final devoicing, onsetfulness, and secondary stress—show conclusively that stumps are PrWds.[6] Clips do not consistently adhere to all of these. I will not consider non-stump compounds or clips further in this paper. In order to clinch the argument that stump compounds are real compounds, however, it remains to be shown that stumps are neither proclitics/prefixes nor full syntactic words.

That stumps are not prefixes or proclitics is very clear. Monosyllabic prefixes and proclitics often consist of a single, vowel-final syllable, which isn't possible for stumps. Furthermore, most (but not all) prefixes fail to exhibit secondary stress.

It is not as easy, however, to show that stumps are not fully accented syntactic words. Reformatskij (1967) considers all of the forms I've called compounds here—stumps and otherwise—to be mere syntactic phrases. This view is supported by the fact that secondary stress sounds very similar to the less prominent word in a syntactic phrase. English distinguishes compounds from syntactic phrases by means of the so-called Compound Stress Rule, as the data in (7) show; only the loudest stress is shown:

(7) Compounds versus syntactic phrases in English

a. *BLACK•bird* (= a species of bird) b. *black BIRD* (= a bird which is black)

This contrast is not available in Russian, since compounds, as in (8a), and syntactic phrases, such as the adnominal-genitive structure in (8b), each assign greater stress to the latter element. There is a slight difference in minimal pairs such as (8a-b):

(8) Compounds versus syntactic phrases in Russian

a.	POL•	KOMnaty	'(the/a) half room'
	half$_{(NUM)}$	room$_{(N.F)GEN}$	
b.	POL	KOMnaty	'(the/a) floor of (the/a) room'
	floor$_{(N.M)}$	room$_{(N.F)GEN}$	

In both of (8a-b) the strongest stress is on *komnaty*, with a less prominent stress on *pol•* 'half' or *pol* 'floor'. The **difference** in prominence between the stresses in the two examples, however, is not the same. The difference in amplitudes and durations of the stressed syllables is closer in (8b) than in (8a). This suggests that (8b) has two main stresses (with different phrasal prominence), while (8a) there is secondary and main stress. Alas, I have not conducted instrumental studies to confirm this impression.

Reformatskij (1967) gives particularly enticing evidence that stump compounds are syntactic words; sentential Wackernagel's Law clitics (shown in italics) can separate a stump from the following stem, as in (9):

6 Observe also that the final consonant of a stump does not become syllabified as the onset of a following vowel-initial stem: *zav•otdelom* [zʌF.ʌdʲ.DʲE.ləm] 'department director' (*otdelom* 'department$_{INST}$'). See also the onsetless second stump in (4): [sʊF.in.FORM.bʲu.RO]. Cf. the forms in (2a, c) where vowel-initial *org* does trigger coda-capture: [pʌr.TORK] and [pʌr.FORK].

(9) a v SEL'• *to* soVET i ne uSPEL zajTI
but to village CL council$_{(N,M)ACC}$ even not managed$_{(V)M.PAST}$ stop.by$_{(V)INFIN}$
'but at the village council {he/you$_{M.SG}$/I$_M$} didn't even manage to stop by'

Note that the clitic not only breaks up the stump compound, it also interrupts a prepositional phrase. I show in Billings (1996:76) that a sentential clitic can even appear between a preposition and pronoun in sandhi environments, as shown in (10):

(10) { JA ne ZNAju, { otnoSItel'no *li* **neGO** oNI govoRJAT. } }
regarding Y/N him$_{GEN}$ they speak$_{3.PL}$
'I don't know if it's regarding **him** (that) they're talking.' (clausal stress on *nego*)

In (10) the orthotonic preposition *otnositel'no* triggers a special, nasal-initial form of the personal pronoun *nego*. Nonetheless, a sentential clitic can intervene. My answer to sentential clitics is prosodic inversion; the clitic appears after the first prosodic-word stress.[6] Stumps are prosodic words. If a PrWd ends within a morphologically (and prosodically) compound word, prosodic inversion places a clitic inside that compound.

Alekseev (1968:119) adds that non-clitic parentheticals can likewise interrupt stump compounds, as shown (11):

(11) meNJA [...] VYgnali za PROF•, *TAK nazyVAemuju*, nepriGODnost'
me$_{ACC}$ expelled$_{3.PL}$ for professional so-called$_{F.ACC}$ uselessness$_{(F)ACC}$
'I was expelled for so-called professional uselessness'

Parentheticals, in my view, constitute a valid argument for the syntactic-phrase-hood of stump compounds. I return to this issue below.

One piece of evidence in support of my morphological analysis is that stumps are bound to the following full stem. That is, the following stem can't be elided or moved syntactically. I show examples in support of this argument below.

2. First case study: De-participial stumps assigning quirky case

Having shown that stump compounds are indeed morphological structures, I turn to the first of two case studies which show that there is a syntactic relation within a morphological compound. Note the case-marking in the forms repeated in (12a-b):

(12) Stump assigning quirky (instrumental) case to the following stem
a. zaVEdujuščij laboraTOrijej ZAV•laboraTOriej [ZAF.lə.bə.rʌ.TO.rʲi.jij]
director$_M$ laboratory$_{(N.F)INST}$ 'laboratory director'
b. upraVLJAjuščij deLAmi uPRAV•deLAmi [u.PRAF.dʲiˑ.LA.mʲi]
manager$_M$ affairs$_{(N.NT)INST}$ 'on-site manager'

The stumps in (12a-b) are truncated from stems that were present-active participles which were in turn derived from verbs, /zavedova-/ 'direct' and /upravljaj-/ 'manage', which both assign quirky instrumental case to their objects. Throughout these morphological derivations, the assignment of instrumental case has been maintained, even by the stumps. All available prosodic tests confirm that these stumps have secondary stress: Because the stump's vowel is /a/ in both, the /o/-roundness test is not available. Because the stump precedes the main stress by at least a syllable, however, a less perspicuous but nonetheless just as conclusive a test is available:[7] Unstressed /a/ (after a non-palatalized consonant) reduces to [ʌ] in the syllable immediately preceding the stress and to [ə] elsewhere. In (12a-b) the stumps' secondary-stressed vowel surfaces as [a], not as [ʌ] or [ə]. These stumps likewise undergo devoicing of /v/ to [f].[8]

This section has shown that stump compounds can involve very specific kinds of syntactic relation (quirky case assigned by the stump to the following stem). Unfortunately, there is little evidence that *zav•* and *uprav •* aren't simple syntactic words. The one proof of their stump-hood is that they are bound to the following stem.

3. Second case study: *pol•* 'half'

The next case study shows this bound-ness more clearly: The element *pol•* 'half', introduced in (1f) and exemplified again in (8a), is unique among numerals in that it must immediately precede the element which it quantifies. That is, the noun quantified by *pol•* must be phonetically overt and must not be moved. In this section I begin by showing that *pol•* is syntactically a numeral. Many of these tests come from Babby (1987) and Mel'čuk (1983; 1985). Then I use my own tests to show that *pol•* is a bound morpheme, which supports my contention that *pol•* is a stump.

Russian has, in the last several hundred years, developed a syntactically distinct category of numerals, of which *pol•* is a member. The most conclusive test for numeral-hood is the case assigned by so-called paucal numerals (those less than 5 in cardinality). This includes the fraction numerals *pol•* 'half' and *četvert'* 'quarter', shown in (13b-c):

(13) Assignment of distinctive "adpaucal" GEN case by paucal numerals
a. TRI čaSA [TRI čʲi.SA]
three$_{(NUM)}$ hour$_{(N.M)ADPAUC}$ 'three hours'
b. POL• čaSA [POL.čʲi.SA]
half$_{(NUM)}$ hour$_{(N.M)ADPAUC}$ '(a) half hour'
c. ČETvert' čaSA [ČʲET.vʲirt' čʲi.SA]
quarter$_{(NUM)}$ hour$_{(N.M)ADPAUC}$ '(a) quarter hour'

Like the other paucal numerals in (13), *pol•* assigns the special adpaucal GEN form to the quantified noun. Only numerals assign this special form. Usually this form is referred to as the GEN.SG, because most nouns don't show a distinction between the adpaucal and GEN.SG forms. The noun in (13) and a handful more MASC nouns do, however, show a distinction: The GEN.SG form has initial stress, as *ČAsa* in (14a-c) shows, compared to the end-stressed *časA* in (13a-c).

---

7 Some characterizations of the reduction of unstressed /a/ and /o/ differ from my description here: In first-pretonic position (after a non-palatalized consonant) /a/ and /o/ neutralize to [a] instead of [ʌ] (and reduce to [ə] elsewhere); see Hamilton (1980) and the sources cited therein. With such a simplified system my vowel-reduction arguments no longer hold in section 2. Either characterization would function, however, for my arguments about the reduction of /o/ above in section 1, where the only relevant criterion is whether /o/ maintains lip-rounding.

8 These two tests—secondary stress and final devoicing—are illustrated opportunely by comparing the stump compound in (1i) *uprav•delami* [u.PRAF.dʲiˑ.LA.mʲi] 'on-site manager' with the clip in (2h) *uprav•dom* [u.prav.DOM] 'manager of a block of flats'. While it remains a mystery why these two forms have resulted in divergent morphological structures, the prosodic effects follow quite straightforwardly from the distinction between stump compound and clip.

(14) Assignment of normal GEN.SG case in adnominal-GEN environments

a.	naČAlo	ČAsa	[na.Čʲ A.lə	Čʲ A.sə]
	beginning$_{(N.NT)}$	hour$_{(N.M)GEN}$		'beginning of {the/an} hour'
b.	poloVIna	ČAsa	[pə.la.Vʲ I.nə	Čʲ A.sə]
	half$_{(N.F)}$	hour$_{(N.M)GEN}$		'half of {the/an} hour'
c.	ČETvert'	ČAsa	[Čʲ ET.ˌvʲ irtʲ	Čʲ A.sə]
	quarter$_{(N.F)}$	hour$_{(N.M)GEN}$		'quarter of {the/an} hour'

Replacing the end-stressed form *čaSA* with stem-stressed *ČAsa* in (13a-b), or replacing *ČAsa* with *čaSA* in (14a-b) results in ungrammaticality. Other assigners of GEN case, such a verbs and prepositions, invariably trigger the stem-stressed GEN.SG form.

Note that (13c) and (14c) have identical first words: *četvert'*. I argue in Billings (1995) that *četvert'* 'quarter' is both a numeral and a noun (homophonous words). This is supported by the test in (15) and (16), where the ADPAUC-triggering numeral takes PL agreement, in (15), while the noun that triggers GEN.SG morphology on the following noun triggers SG agreement on the determiner, in (16):

(15) Plural agreement (on demonstrative) if there is a numeral

a.	TE	TRI	čaSA	'those three hours'
	those$_{NOM.PL}$	three$_{(NUM)}$	hour$_{(N.M)ADPAUC}$	
b.	TE	POL•	čaSA	'that half hour'
	those$_{NOM.PL}$	half$_{(NUM)}$	hour$_{(N.M)ADPAUC}$	
c.	TE	ČETvert'	čaSA	'that quarter hour'
	those$_{NOM.PL}$	quarter$_{(NUM)}$	hour$_{(N.M)ADPAUC}$	

(16) Singular agreement (on demonstrative) if there is a noun

a.	TO	naČAlo	ČAsa	'that beginning of {the/an} hour'
	that$_{NT}$	beginning$_{(N.NT)}$	hour$_{(N.M)GEN}$	
b.	TA	poloVIna	ČAsa	'that half of {the/an} hour'
	that$_{F}$	half$_{(N.F)}$	hour$_{(N.M)GEN}$	
c.	TA	ČETvert'	ČAsa	'that quarter of {the/an} hour'
	that$_{F}$	quarter$_{(N.F)}$	hour$_{(N.M)GEN}$	

The agreement combinations in (15) and (16) are the only ones possible. If the nominal expressions in (15) and (16) are the sentential subject, then they trigger PL or SG agreement on the verb, respectively. Thus, while homophonous noun and numeral versions of *četvert'* 'quarter' co-exist in the lexicon (each with its own syntactic properties), the forms meaning 'half', *pol•* and *polovina*, are not homophonous. Nor do they share the same morphological properties: *pol•* is a stump; *polovina*, a word.

One final syntactic test for numeral-hood is shown in (17) and (18).[9] In time expressions (meaning 'at ... o'clock') the preposition *v* 'in' triggers the ACC case in the

[9] Above I argue for the co-existence of noun and numeral forms of *četvert'* 'quarter'. I have no explanation for why *četvert'* cannot take the LOC case as in (18); only the ACC-assigning construction is available: *v ČETvert' PJAtogo* (in$_{(P)}$ quarter$_{(NUM)ACC}$ fifth$_{(ADJ)M.GEN}$) 'at quarter past four'. Apparently only its numeral form is allowed in time expressions. Still, when an unquantified nominal expression is the object of this preposition, a clear LOC case is attested.

numeral (which then triggers the GEN case in the following nominal expression), as in (17); but *v* triggers the LOC case in a non-numerical nominal expression, as (18) shows.

(17) The preposition *v* with numerically quantified time expressions

a.	v	TRI	čaSA	'at three o'clock'
	in$_{(P)}$	three$_{(NUM)ACC}$	hour$_{(N.M)ADPAUC}$	
b.	v	POL•	PJAtogo	'at half past four'
	in$_{(P)}$	half$_{(NUM)ACC}$	fifth$_{(ADJ)M.GEN}$	

(18) The preposition *v* with non-quantified time expressions

a.	v	naČAle	PJAtogo	'at just after four (o'clock)'
	in$_{(P)}$	beginning$_{(N.NT)LOC}$	fifth$_{(ADJ)M.GEN}$	
b.	v	poloVIne	PJAtogo	'at half past four (o'clock)'
	in$_{(P)}$	half$_{(N.F)LOC}$	fifth$_{(ADJ)M.GEN}$	

Once more, (17b) and (18b) show that *pol•* is a numeral and *polovina* is a noun.

Before continuing my analysis of the two words that mean 'half', I should show two tests which prove that *pol•* is indeed a bound morpheme. The contrast in (19) shows that all non-bound numerals can take an elided noun after them:

(19) Ellipsis of the quantified noun

a.	podoŽDĚM ešČĚ TRI (čaSA)	'Let's wait another three hours.'
b.	podoŽDĚM ešČĚ POL•*(čaSA)	'Let's wait another half hour.'
c.	podoŽDĚM ešČĚ ČETvert' (čaSA)	'Let's wait another quarter hour.'

Although it is not always felicitous to elide the noun after the numeral, it is **possible** in (19a) or (19c). With *pol•* in (19b) the following noun is **obligatory**.

The other indicator that *pol•* is morphologically bound is the approximative-inversion construction, discussed in Billings (1995) and exemplified in (20). Whereas most numerals and the following quantified noun can be juxtaposed to express approximate cardinality, this is not possible with *pol•*.

(20) Approximative inversion with most (non-bound) numerals

a.	TRI čaSA	'three hours'	čaSA TRI	'*approximately* three hours'
b.	POL•čaSA	'half hour'	*čaSA POL•	'*approximately* a half hour'
c.	ČETvert' čaSA	'quarter hour'	čaSA ČETvert'	'*approximately* a quarter hour'

The contrasts between (19b-c) and (20b-c) are especially convincing because the numerals are both fractions. Syntactically *pol•* and *četvert'* are numerals, yet these two possess quite divergent morphological properties.

Having proven that *pol•* is syntactically a quantifier, it remains to be proven that *pol•* is not some prefix or proclitic. The vowel-reduction test discussed above in section 1 clarifies this issue: *pol•* always maintains lip-rounding in its vowel /o/: [POL.čʲ i.SA]; this eliminates any analysis of it as prefix or proclitic.

Returning to *pol•* and *polovina*, I propose that these two forms are, morphologically, a stump and full form, despite the fact that *pol•* is a numeral and *polovina* is a noun. This analysis is not, however, accurate etymologically; *polovina* is the form historically derived from the Common Slavic root **pōl* (cf. Czech *půl*, < /pōl/ 'half'), of which

Russian *pol•* is also a reflex. Stump compounds other than *pol•* (and clips) are discussed extensively in the sociolinguistic literature, because their emergence largely coincided with the 1917 revolution (and many of the coinages have to do with that political system). According to Maksimov (1973), the diachronic introduction of *pol•* in constructions like (17b), as opposed to (18b), began in the 1800s and really became widespread during this century. The re-analysis of *pol•* as the stump corresponding to *polovina* is something akin to back-formation. Still, the proper **synchronic** analysis for these two forms' morphology is that of a stump and its corresponding full-form.

How then is it possible that a stump and full form differ as to their syntactic category? (I am grateful to Dirk Carius for posing this seeming inconsistency to me so clearly.) This is not really a problem for two reasons: If morphology and syntax are truly autonomous, then two lexical items that are somehow linked morphologically need not possess identical syntactic subcategorization. Additionally, after a stump compound is derived, it can evolve its own lexical properties. Panov (1968:117) shows that in rare cases stump compounds have developed meanings unrecoverable from their untruncated parts. 'He gives (1h), where the two full words *xozjajstvennyj rasčët* mean something like 'household accounting', while the stump compound *xoz•rasčët* means 'self-supporting basis'. This, too, seems to be what's going on with *pol•* and *polovina*.

4. On the stability of syntactic relations within stump compounds

This final section re-assesses whether syntax is possible within a word. The assumption made by many linguists is that this isn't possible. Worth (1959) concludes that *pol•* can't be a bound morpheme because of its ability to assign case. Reformatskij (1967) mentions that stump compounds like (12) are "syntactically 'uncomfortable' ". Note also that the majority of stumps are truncated adjectives followed by full forms of nouns; cf. (1a-c, e, g-h), (4), (9) and (11). There is no morphological evidence of syntax within such forms. The two case studies represent a small minority of all stumps.

More colloquial versions of the data in the two case studies suggests, however, that syntax within words is somewhat unstable: I mention above that stumps cannot be separated from the following full forms. Comrie, Stone and Polinsky (1996:141) report that in colloquial speech just a few stumps, including *zav•* 'director' in (12a), can be pronounced by themselves. I've further confirmed that both stumps in (12) can be case-marked when standing alone (e.g., *zavu* 'director$_{DAT}$'). This colloquial development suggests, then, that the stumps in (12) are merely becoming full-fledged words, with primary stress. I have detected no difference in meaning, aside from register differences, between *zav•*, free-standing *zav*, and their corresponding untruncated form.

Moving to *pol•*, there is a feint indicator that a different repair strategy is being employed to eliminate syntax within this type of compound. The examples in (21) show colloquial and slang forms of one example:

(21) Loss of case-assignment in one form with *pol•*

a.	polbanki	POL• half$_{(NUM)}$	BANki jar$_{(N.F)ADPAUC}$	[POL.BAN.k^{j}i]	'half-liter bottle of vodka'
b.	polbanka	pol	BANka	[pal.BAN.kə]	'half-liter bottle of vodka'

[Zalucky (1991:563); stress and phonetic notation elicited]

In (21a) the word after *pol•* is in the ADPAUC (homophonous with the GEN.SG in this stem). In (32b), however, the following word appears to end in a caseless stem. This is similar to adjectives in a German nominal expression following a definite article, where the adjectives just have a [ə] inflection. The extremely substandard form in (21b) is not even known to all my informants; the one informant who does accept (21b) pronounces *pol* without lip-rounding. This suggests that this form involves neither morphological nor syntactic structure. If this is any indication of things to come, the case-assignment within stump compounds with *pol•* may be on the wane.[10]

Therefore, the colloquial indicators in section 4 indicate, by no means conclusively, that syntactic relations within compounds are not stable.

To conclude this paper, I have shown how stump compounds in Russian pose a challenge to the autonomy hypothesis. Two types of stumps assign very specific cases to the word form to which they are bound morphologically. Still, the rarity of such forms, and possible indicators that this interference of syntax into the internal affairs of words is being avoided, suggest that the autonomy of syntax from morphology is alive and, if not well, at the very least showing signs of recovery.

Alekseev, D. I. (1968) [§§61-64 (pp. 91-93) of] *Slovoobrazovanie sovremennogo russkogo literaturnogo jazyka.* [= *Russkij jazyk i sovetskoe obščestvo. Sociologo-lingvističeskoe issledovanie*, **2**.] M. V. Panov (ed.). Moskva: Nauka.

Alekseev, D. I. / I. G. Gozman / G. B. Saxarov (1983) *Slovar' sokraščenij russkogo jazyka.* D. I. Alekseev (ed.). 3rd, expanded ed. Moskva: Russkij jazyk.

Babby, Leonard H. (1987) "Case, prequantifiers, and discontinuous agreement in Russian." *Natural language and linguistic theory* **5**:1, 91-138.

Billings, Loren A. (1995) *Approximation in Russian and the single-word constraint.* Princeton University Ph.D. dissertation.

Billings, Loren A. (1996) "Sandhi phenomena and language change." In *Interfaces in phonology*. U. Kleinhenz (ed.) [= Studia Grammatica, **41**.] Berlin: Akademie Verlag, 60-82

Borunova, Svetlana L. / Vera L. Voroncova / Natal'ja A. Es'kova (1989) *Orfoëpičeskij slovar' russkogo jazyka. Proiznošenie, udarenie, grammatičeskie formy. Okolo 65 000 slov.* R. I. Avanesov (ed.). 5th (corrected and expanded) ed. Moskva: Russkij jazyk.

Comrie, Bernard / Gerald Stone / Maria Polinsky (1996) *The Russian language in the twentieth century.* 2nd, revised and expanded edition [of Comrie, B. / G. Stone (1978) *The Russian language since the Revolution.*] Oxford: Clarendon Press.

10 Forms like (21b) share a striking similarity with colloquial stump compounds. Mel'čuk (1983) reports that constructions involving *pol•* 'half', when externally assigned oblique cases (GEN, DAT, INST and LOC), display such case endings instead of the ADPAUC ending triggered by *pol•*. Indeed, I have determined that a GEN-assigning preposition will cause the noun quantified by *pol•* to take the non-ADPAUC GEN form: *bez pol•ČAsa* 'without half an hour$_{GEN}$' (not **bez pol•čaSA*). Leonard Babby informs me that similar case marking is observed with de-adnominal stump compounds like *zam•direktora* 'deputy director$_{GEN}$' in colloqual speech when the entire compound is assigned a non-GEN oblique case, as in *k zam•direktoru* 'to (the) deputy director$_{DAT}$'.

Downing, Laura J. (1998) "On the prosodic misalignment of onsetless syllables." To appear in *Natural language and linguistic theory* **16**:1.
Graudina, L. K. (1968) [§§61-62 (pp. 91-94) of] "Morfologija." [Chapter 1 (pp. 9-232) of] *Morfologija i sintaksis sovremennogo russkogo literaturnogo jazyka.* [= *Russkij jazyk i sovetskoe obščestvo. Sociologo-lingvističeskoe issledovanie*, **3**.] M. V. Panov (ed.). Moskva: Nauka.
Hamilton, William S. (1980) *Introduction to Russian phonology and word structure.* Columbus, Ohio: Slavica.
Ickovič, V. A. (1971) "Sovremennye abbreviatury." *Russkaja reč'* (no.) 2. 74-79.
Jakobson, Roman O. (1921) "Vliv revoluce na ruský jazyk. Poznámky ke [Mazon (1920)]." *Nové Atheneum* (roč.) **2** [Praha: Tiskem Edvarda Leschingra v Praze], 3-32. [MIT Archives (Jakobson Papers): MC 72 Box 7 Folder 21 (offprint)]
Karcevskĭj, S. Ž. (1923) *Jazyk", vojna i rvoljucĭja.* Berlin": Russkoe universal'noe izdatel'stvo.
Kramer, Alex A. (1965) *Abbreviations in Soviet publications.* Trenton, New Jersey: Scientific Russian Publishing Service.
Maksimov, V. I (1973) "*Polovina pervogo, polpervogo?*" *Russkaja reč'* (no.) 6, 51-53.
Mazon, Andrè (1920) *Lexique de la guerre et de la rèvolution en Russie (1914-1918).* (= Bibliothèque de l'Institut français de Petrograd (tome) **6**.) Paris: Librarie ancienne honoré Champion.
Mel'čuk, Igor [sic.] A. (1983) "The numeral *pol*[1] 'half' in modern Russian: An autonomous word or a bound morpheme." *Journal of linguistic research* **2**:1, 49-62.
Mel'čuk, Igor' A. (1985) *Poverxnostnyj sintaksis russkix čislovyx vyraženij* [The surface syntax of Russian numeral expressions]. (= *Wiener slawistischer Almanach* Sonderband **11**.) Wien: Institut für Slawistik der Universität Wien.
Panov, M. V. (1968) [§§82-83 (pp. 116-120) of] "Morfologija." [Chapter 1 (pp. 9-232) of] *Morfologija i sintaksis sovremennogo russkogo literaturnogo jazyka.* [= *Russkij jazyk i sovetskoe obščestvo. Sociologo-lingvističeskoe issledovanie*, **3**.] M. V. Panov (ed.). Moskva: Nauka.
Reformatskij, Aleksandr A. (1967) ["Složenija." §50 (pp. 289-293) of his] *Vvedenie v jazykoznanie.* 2nd ed. Moskva: Prosveščenie.
Rubach, Jerzy and Geert Booij (1985) "A grid theory of stress in Polish." *Lingua* **66**, 281-319.
Seliščev, A. M. (1928) *Jazyk revoljucionnoj èpoxi. Iz nabljudenij nad russkim jazykom poslednix let (1917-1926).* 2nd ed. Moskva: Rabotnik prosveščenija.
Ward, Dennis (1965) *The Russian language today. System and anomaly.* London: Hutchinson University Library.
Worth, Dean S. (1959) "Grammatical and lexical quantification in the syntax of the Russian numeral." *International journal of Slavic linguistics and poetics* **1-2**. 117-132. [Reprinted in his *On the structure and history of Russian.* (1977) München: O. Sagner. 43-58.]
Zalucky, Henry K. (1991) *Compressed Russian. Russian-English dictionary of acronyms, semiacronyms and other abbreviations used in contemporary standard Russian.* Amsterdam etc.: Elsevier.

Janet DeCesaris - e.mail: janet.decesaris@trad.upf.es
Maria Rosa Bayà - e.mail: rosa.baya@trad.upf.es
Elisenda Bernal - e.mail: elisenda.bernal@trad.upf.es
Institut Universitari de Lingüística Aplicada - Universitat Pompeu Fabra (Spain)

NOUN-NOUN SEQUENCES IN CATALAN AND SPANISH

Abstract: This paper examines noun-noun sequences in Catalan and Spanish in which the second, non-head noun appears to be modifying the first, head noun. It has been previously claimed that most of these sequences correspond to a specific semantic class, but the examples show that such a restriction cannot account for the attested data. The prototypical interpretation for these sequences involves assigning a figurative sense to the non-head noun. It is suggested these sequences are generated in the syntax, thus expanding the nature of noun phrase structure in these languages.

0. Introduction

Compounding is at the heart of much research in morphology because it lies at the intersection of two components, one of which is responsible for forming words and the other of which is responsible for forming phrases and sentences. In languages like Catalan and Spanish, it has been claimed (Rainer and Varela 1992) that there are not always observable differences between the morphosyntactic behavior of lexical compounds and that of syntactic strings, and therefore crucial arguments in favor of classifying a sequence as a compound as opposed to as a syntactic phrase are often based on semantics, and many linguists feel uncomfortable with classifications based on claims on what a sequence means as opposed to its formal behavior.[1] Exactly what constitutes a compound in any given language is thus not always straightforward. This is the case for English, if we consider that many of the strings claimed to be compounds by Lieber (1992) have been argued to not be compounds but rather be mere quotations of syntactic strings (cf. Weise 1996).[2] The characteristics of compounds in Romance languages, particularly noun-noun compounds, are even less well defined, to judge

[1] We are not endorsing this lack of confidence in semantically based analyses, but rather stating what to us seems to be a fact in morphological analysis: linguists prefer to base claims for lexical compounds as opposed to syntactic phrases on easily quantifiable behavior such as pluralization, type of complement allowed, relationship to other word-formation processes, etc. rather than on assertions of meaning, which are more open to differing interpretations.

[2] We cite Weise (1996) because he specifically discusses forms claimed to be compounds by Lieber, but take this opportunity to note that the idea of quoting syntactic strings and inserting that information into a context did not originate with. Spencer (1991) cites Bauer (1988) on this, and it is unfortunate that Weise makes no mention of these two well-known, easily accessible sources.

In fact, one of the common threads underlying his discourse is that most of these forms are unnatural in the language in some sense and should be replaced by more Spanishlike formations. He is probably correct in assuming that many are loan translations from English, but the fact remains that modern Spanish has been able to incorporate these strings to such extent that the structure is productive.

Noun-noun sequences take the gender marking of the left noun. Plural formation of these sequences varies, not only with the string but also with the speaker. As Rainer and Varela note for Spanish (1992: 126), plural marking is usually only found on the left noun, but can occasionally also appear on the right noun. This is also the case for Catalan; e.g. conservative speakers will say *problemes clau* 'key/main problems,' but all three of us have heard, both on the radio and in conversations we ourselves have had. *problemes claus.*[5]

2. Rainer and Varela's discussion of the data

Rainer and Varela point out that "left-headed n-n compounds and restrictive appositions" are the hardest sequences in Spanish to classify as either compounds or phrases. They provide examples of nine types of noun-noun sequences, and suggest that "the hard core of n-n compounds is thus constituted by the coordinative type *cantante autor* ['singer-songwriter'] and the subordinative type *ciudad dormitorio* ['bedroom community'], while the borders both towards apposition and adjective phrases seem to be somewhat fuzzy" (1992: 120). They go on to claim that contrary to what occurs in Germanic, Spanish compounds of this sort "in productive use only follow a restricted set of semantic types" (1992: 126), although they do not identify what exactly defines these semantic types.

Let us consider their idea in more detail, particularly with reference to other word-formation processes in Spanish and Catalan. Obviously, semantics plays an important role in determining whether a given formative can actually combine with a given stem in derivation. There are surely semantic restrictions on compounding as well, so that a combination like Sp. *ciudad cantante* / Cat. *ciutat cantant* is not likely to exist, because in the real world that we refer to we know of no 'singer cities' nor are we likely to require this reference But this does not seem to be the type of semantic restriction Rainer and Varela are alluding to. Rather, they state that the words *modelo* ('model'), *estrella* ('star'), *límite* ('limit'), *fantasma* ('ghost'), and *clave* ('key') form a semantic class of nouns which lends itself to an adjectival interpretation, and thus these forms are beginning to act like adjectives in terms of pluralization and predicative use. It is not clear to us exactly what this semantic class is or how it can be described. We note that even if this semantic class existed, it would be unique in terms of word-formation in Catalan and Spanish. Restrictions on word-formation can be based on argument structure or semantic features such as [animacy], but we know of no semantic feature grouping together the nouns in question. Moreover, the paraphrase offered by Rainer and Varela, "an n_1 which is like an n_2 " (1992: 126) is too general to effectively

[5] We note that Rainer and Varela have the same intuitions about *clave* in Spanish.

identify a semantic class. Rainer and Varela's comments seem to be based on the idea that these are the only (or practically only) nouns used productively in this sort of construction, but that is not the case, as witnessed by the examples given above.

3. The semantics of subordinative compounds

Although we are not convinced that there is a single semantic feature or set of features shared by nouns appearing in non-head position, it does seem to be the case that these nouns are typically not interpreted literally but rather figuratively. Let us consider the following examples:

(3) a. el plan/programa/proyecto piloto 'pilot plan/program/project'
b. el paper moneda 'bill' (lit. paper coin)

The only possible way to interpret *piloto* is figuratively, as neither plans nor programs nor projects can be pilots because pilots must be animate. Likewise, paper is not like coins in a literal sense but rather serves the same purpose as coins do.

The behaviour of the noun *bomba* 'bomb' is particularly interesting. In both languages, *bomba* combines with other nouns to refer to the bomb's packaging:

(4) cotxe bomba (Catalan) / coche bomba (Spanish) 'car bomb'
paquet bomba / paquete bomba 'letter bomb'

Although Rainer and Varela analyze *coche bomba* as a subordinative compound, we suggest that the literal interpretation given to both nouns would put it on a par with *sofá cama*: i.e., in our view it is a coordinative compound. *Bomba*, however, can also be understood figuratively:

(5) Els preus bomba (Catalan) 'shockingly low prices'
Una noticia bomba (Spanish) 'shocking news item'

If semantics is the basis for comparison of coordinative vs. subordinative compounds, then *preu bomba* must be different in structure from *cotxe bomba* because *bomba* is not understood in the same fashion.

Another interesting set of examples are those listed in (6), taken from a document on pharmaceutical procedures written in Spanish.

(6) a. la solución patrón 'standard reference solution'
b. la solución problema 'test solution'
c. la solución reactivo 'reagent solution'

The sequences in (6a) and (6b) can be analyzed as subordinative compounds in that the solution serves as the standard reference or as the test case being run through the equipment. Neither (6a) nor (6b) appears to have arisen as a result of preposition

deletion, as the corresponding phrases **solución de(l) patrón/problema* do not occur.[5]

The phrase in (6c), however, may be related to *la solución del reactivo*, with deletion of the preposition *de* 'of' and of the definite article.

Preposition deletion has not received much attention in Spanish or Catalan. Rainer and Varela mention that forms which seem to have lost a preposition, such as *tren mercancías* or *moto todo terreno* (their examples) probably do not qualify as lexical structures but rather as syntactic strings, and we would agree. Examples of noun-noun sequences that appear to be missing a preposition are commonplace in advertising, where they often coexist with the full forms:

(7) Examples taken from hardware store catalogue in Catalan, summer 1996

a. Both forms:	ventilador sostre	'ceiling fan'
	ventilador de sostre	'ceiling fan'
b. No preposition:	bústies jardí	'garden mailboxes'
	mobles jardí	'garden furniture'
	carret jardí	'garden cart'
	dutxa jardí	'outdoor shower'
	taules resina	'resin tables'
	taula ordinador	'computer desk'

In an advertising context, in which a speaker's knowledge of real world references is all-important, the semantic information contributed by the preposition is easily obtained without the preposition itself. It is what we know about the world, and not some characteristic of noun-noun sequences, that allows us to interpret *taula resina* as a table made out of a certain substance yet *taula ordinador* as a table designed to hold a computer while in use. That explains why the specific preposition that can be deleted is not always *de* 'of', which is surely most often used to relate nouns, but can also be Catalan *per a* 'for', *amb* 'with', Spanish *por, para* 'for', *con* 'with', etc. The semantics of these noun-noun sequences is thus crucially dependent on external information, whereas the semantics of true subordinative noun-noun sequences is more dependent upon a non-literal interpretation of the non-head noun.

4. Noun-noun sequences with proper names

A subset of subordinative noun-noun sequences are those strings in which the righthand noun is a proper name:

(8)

una cafetera Bosch	'Bosch coffeemaker'
ventanas Velux	'Velux-brand windows'
la garantía Nissan	'the Nissan guarantee'

Zwanenburg, in his discussion of the parallel forms in French, suggests that this structure is used for name-giving and then becomes more or less fixed in meaning

[5] *La solución del problema* would be interpreted as the answer to the problem.

(1992: 225). It is certainly true that this kind of structure can become lexicalized, as in the Spanish phrase *cuerpo Danone* 'body by Dannon Yogurt', which is not used to describe the bodies of yogurt-eating individuals but rather to describe a young, physically fit person, but the great majority of these strings do not exhibit this behavior. The prototypical interpretation for these strings is 'made, manufactured, produced by..,' but again, the role of what a speaker knows about the real world is important in being able to interpret the phrase. We may compare two phrases:

(9) a. madre Nutribén 'a Nutriben mother, a mother who uses Nutriben products'

b. potito Nutribén 'a Nutriben jar of baby food'

The possible relationship between the nouns in (9a) cannot be the same as that in (9b) because as people we know that mothers are not made by the manufacturers of commercially prepared baby food. Unlike the subordinative sequences discussed above, however, the non-head noun *Nutribén* is not understood figuratively. Rather, this seems to be a syntactic structure in which a commercial company is related to a head noun. This structure probably originated from one with the preposition *de* 'of', but can no longer only be explained by preposition deletion precisely because of examples like (9a), as **una madre de Nutribén* is ungrammatical. Nor does the structure seem to display the semantic unity that is typically claimed for lexical compounds. We therefore suggest that these strings are the result of a syntactic structure combining two nouns.

5. Lexical vs. syntactic origin of compounds

As we stated at the outset, one of the major debates in morphology addresses whether compounds are lexical or syntactic in nature. Both views have eloquent defenders: Lieber (1992) claims compounds are syntactic because all morphological processes are syntactic; some years earlier Di Sciullo and Williams (1987) argued that what had previously been claimed to be compounds in French were lexicalized syntactic phrases listed in the lexicon and not the output of word-formation in the lexicon, which is basically the view exposed in Zwanenburg (1992). Alternatively, a recurring them in the work of linguists like Carstairs-McCarthy and Spencer is that word-formation is just that –word-formation– and not sentence formation. Choosing one approach or the other depends on one's overall view of what the organization of a grammar should be, what generalizations one wants to make in that grammar, and what tools one wants to use or has available to make those generalizations. Even for a single structure, those decisions can only be made when a substantial variety of word-formation structures have been closely examined. Since this paper has only dealt with a small number of examples of a few structures, it would be rather presumptuous to make any far-reaching claims based on such little data. We can suggest, however, whether the types of noun-noun sequences discussed should be generated by the word-formation component or by the syntax. To date, many of the arguments for a syntactic origin entail lexicalization of a set phrase, or at least a phrase naming a specific reference. We believe these two issues need to be separated from one another. The productive use of

nouns in the non-head position would seem to simply that lexicalization of a set phrase is not necessary for a noun to be used as a noun modifier. Strings like *piso muestra* (literally 'model apartment, apartment sample') or *preu bomba* are not lexicalized in our view, yet we think they should be generated in the syntax as opposed to in the word-formation component. This, in turn, means that nouns can act as modifiers of other nouns in Catalan and Spanish, which in effect expands the possible type of noun phrase structure usually assumed for these languages. The acceptability of these structures in the modern languages is probably aided by external factors such as increased influence from English, especially in journalistic prose, and an advertising style in which recoverable information is deleted to save space, as well as by the existence of other noun-noun sequences resulting from apposition. These syntactic phrases may lexicalize, as in *fecha límite / data límit* 'deadline', but it is equally as possible for the non-head noun to become increasingly more adjectivelike, as is the case with *clave / clau* 'key'.

6. Concluding remarks

The analysis outlined in this paper expands the roles nouns play in Catalan and Spanish. If this approach is correct, then we should expect to find nouns in other non-prototypical contexts, and we do:

(9) Spanish nouns in adverbial contexts:
Me lo pasé *muy bien / bomba*. 'I had a blast.'
No hagas nada *bestia* (lit. 'beast'). 'Don't do anything stupid.'
Les gustó *una barbaridad* (lit. 'horror'). 'They like it a lot.'

Catalan nouns in adverbial contexts:
No m'agrada *gota* (lit. 'drop'). 'I don't like it at all.'
Li van impressionar *quantitat* (lit. 'amount'). 'They impressed him/her a lot.'

The alternative, considering subordinative noun-noun sequences compounds, in our view would open the flood gates as to what constitutes a compound in these languages. That in itself is not argument against it, but lexical origin seems ill-equipped to account for proper nouns as non-head nouns, and if those noun-noun sequences are to be treated like other subordinative sequences –and we think they should– then a syntactic approach is preferable.

Noun-noun sequences in Catalan and Spanish are interesting proving ground for morphological analysis. We hope to continue gathering data on potential compound structures, and specifically on the semantics involved in subordinative noun-noun sequences, not only to achieve a better understanding of how these two languages work, but also to shed light on compounding and word formation in Romance. For now, we hope to have shown that the range of semantic relationships between the two nouns is wider than previously thought.

References

Bauer, L. 1988. *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.

Corbin, D. 1994. "Locutions, composés, unités polylexématiques: lexicalisation et mode de construction". *Actes du colloque «La locution, entre lexique, syntaxe et pragmatique*. Paris - Saint Cloud.

Di Sciullo, A. M.; Williams, E. 1987. *On the definition of word*. Cambridge, MA: MIT Press.

Lieber, R. 1992. *Deconstructing Morphology*. Chicago: University of Chicago Press.

Martínez de Sousa, J. 1996. *Diccionario de usos y dudas del español actual*. Barcelona: VOX-Biblograf.

Rainer, F.; Varela, S. 1992. "Compounding in Spanish". *Rivista di Linguistica*, 4/1, 117-142.

Scalise, S. (ed.) 1992. *The Morphology of Compounding*. Special issue of *Rivista di Linguistica*, 3/1. Turin: Rosenberg & Sellier.

Spencer, A. 1991. *Introduction to Morphological Theory*. Oxford: Blackwell.

Weise, R. 1996. "Phrasal Compounds and the Theory of Word Syntax". *Linguistic Inquiry*, 27/1, 183-193.

Zwanenburg, W. 1992. "Compounding in French". *Rivista di Linguistica*, 4/1, 221-240.

Anna-Maria Di Sciullo
Université du Québec à Montréal

Abstract

The formal and semantic differences between compounds and phrases are proposed to follow from general economy considerations, forcing X^0 and XP expressions to be distinct as soon as possible in the derivation. Some consequences of this hypothesis for the differences in conceptual interpretation between the two sorts of expressions are discussed. It is shown that the formal differences involved, basically that the spec-head and the head-complement relations and the features they bare are not visible for interpretation at the interface between word-structure and the Conceptual-Intentional system, correctly predict that the semantic features associated i) with the subject or external argument, ii) with delimiting and measuring modification, and iii) with referential specific categories, are not available for interpretation in compound structure, whereas they are in phrasal structure.

ECONOMY OF X^0/XP DERIVATIONS[1]

1. Purpose

The purpose of this paper is to account for the differences and the similarities between words (X^0s) and phrases (XPs) in terms of economy considerations. We focus on compounds, as they exhibit both X^0 and XP properties.

On the one hand, compounds are X^0s by phonological criteria, they have a unique stress on the non-head (Chomsky and Halle, 1965; Cinque, 1992), by conceptual criteria, they exhibit conceptual and referential opacity (Di Sciullo and Gruber, 1992), and by syntactic criteria, they are non separable (Di Sciullo and Williams, 1987). On the other hand, compounds are XPs by categorial criteria, viz., they include more than one categorial item, they may include a phrasal constituent, and they may instanciate pseudo-modification and pseudo-complementation (Di Sciullo, 1994, 1997). The question we address in this paper is the following: why are compounds both different and similar to phrases?

1.1. Proposal

We propose that the differences and the similarities between compounds and phrases are an effect of the economy of the grammar. Compounds and phrases may share properties in their derivations, which proceed in the same computational space. However, they crucially must differ at the interfaces with the performance systems by conceptual necessity, as the conceptual system must distinguish X^0s from XPs on configurational grounds in order to associate each expression to a different sort of interpretation.

[1] We thank Noam Chomsky, Ken Hale, Tom Roeper and the members of the Modularity project at the Université du Québec à Montréal for discussions on this and related topics.

This work is supported in part by the Social Sciences and Humanities Research Council of Canada (grant number 411-92-0012) as well as by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (grant number 95-ER-0401).

1.2. Assumptions

We assume that the derivation of the linguistic expressions, X^0s and XPs, takes place in the same computational space as an effect of the operation of the different modules of the grammar, including the morphology and the syntax (Chomsky, 1995). The modules are autonomous to the extent that they instanciate specific treatments of the grammatical primes, features and elemental configurations, as in Di Sciullo (1996a).

We assume further, as in Di Sciullo (1996b), that the computational space includes interacting types of derivations leading to optimal target configurations. The derivations lead to canonical target configurations (CTC), the features of which can be visible, i.e. interpreted, at the interfaces with the cognitive system. The canonical target configurations are limited to specifier-head, head-complement and adjunct-head configurations.[2] They are the elemental structures upon which the configurational asymmetries and the features they bare are legible at the interfaces. We assume that a feature is visible/interpreted only in a target configuration. X^0s CTC is an adjunct-head configuration, XPs CTC is a spec-head-compl configuration at the conceptual interface.

1.3. Hypothesis

We propose that the similarities between compounds and phrases may not survive in the derivation and crucially not at the conceptual interface, given the Optimality of X^0/XP derivations, that we define as follows:

(1) Optimality of X0/XP derivations
X^0/XP derivations must be distinct as early as possible in the derivation.

Thus, compounds and phrases may share configurations in the derivation and at the acoustic interface (PF), but they must differ at the conceptual interface.

This condition is part of the economy conditions internal to the grammar. It narrowly restricts the possible derivations of words and the possible derivations of phrases. It ranges over properties not covered by the economy conditions of the Minimalist Program (Chomsky, 1995).[3]

1.4. Consequences

If the derivation of the X^0s and XPs lead to distinct target configurations at the conceptual interface, we predict that X^0s and XPs have different conceptual properties at that interface. We show that they differ with respect to the visibility of the features of:

(2) a. the subject or external argument
b. delimiting and measuring modification
c. referential specific nominals

[2] We take the specifier-head configuration to be the target configuration for feature agreement under identity; the head-complement configuration to be the target configuration for feature matching under partial identity; and the adjunct-head configuration to be the target configuration for feature inclusion under distinctiveness, as in Di Sciullo (1997).

[3] Procrastinate. Greed and Minimal Link apply at every step in the derivation and allow to choose amongst convergent derivations, i.e. interpretable at the interface, optimal ones. Full Interpretation is a condition on representations requiring that no superfluous element be part of the interfaces. The Optimality distinguishes X^0s and XPs, notwithstanding their similarities.

These features, visible in spec-head and head-complement configurations in XP structure, are not visible in compounds, which are, in our theory, adjunt-head structures at the conceptual interface.

This paper is organized as follows. In section 2, we discuss the features supported by the predication relation and show that the features associated with the subject are not visible/interpretated in compounds. In section 3, we consider the differences between words and phrases with respect to aspectual modification, bringing evidence to the effect that at least one sort of aspectual modification cannot be obtained within compounds. In section 4, we show that, as it is the case for measuring modification and subjecthood, specific reference cannot be obtainted with compounds because the configuration supporting the features is not available at the morpho-conceptual interface. These facts are predicted by our theory and follow in a unified way. In the last section, we consider some theoretical advantages of our proposal over alternative treatments.

2. Predicates and Predication

In this section we focus on the differences between compounds and phrases with respect to predication. Compounds are X^0s, they are predicates, viz., open functions, and they include predicates. On the other hand, phrases are XPs, they are closed functions, and include closed functions. Functional closure may be achieved via predication (Williams, 1980, 1994), Theta-binding (Higginbotham, 1985) and feature checking (Chomsky, 1995).

2.1. Primary predication

In the Minimialist framework, (Chomsky, 1995; Ura, 1997) predication is a manifestation of the spec-head relation, where features of the external argument or the subject are in the spec position.

(3) $[\text{Spec}\ \ X]_x$
|
subject

While predicates are legitimate in XPs and X^0s, the predication relation may only be legitimate in XPs. In deverbal compounds, the bare nominal category does not qualify as an external argument or a subject. This is the case for compounds, where the nominal non-head qualifies as an internal argument, but not as an external one. The nominal non-head is interpreted as the theme, but not as an agent or an experiencer, compare (4) and (5).

(4) a. LI-readers
b. LI-reading

(5) c. *student-reader of LI
d. *student-reading of LI

The exclusion of subject within compounds follows, in our theory, from the fact that spec-head configurations are not canonical target configurations for words, assuming that sujets may only be visible under spec-head relation at the conceptual interface.

2.1.1. Relatives

Furthermore, assuming that the predication relation also holds between the head of a relative clause and that clause, as suggested in Williams (1994), our theory correctly

predicts that relative clauses may not be visible/interpreted as a compound at the morpho-conceptual interface.

In effect, in the case of the relativization of the subject, the head of the relative agrees in phi-features with the verb via the wh-element in spec-CP. Subject-verb agreement does not hold within compounds.

(6) a. the [one who writes] the book is out of town.
b. the [ones who write] the book are out of town.

(7) a. *the [[writes]-er] of SPE
b. the [[write]-er] of SPE

This again follows from our hypothesis that at the conceptual interface, the spec-head configuration and the features it bares are not visible/interpretable in X^0s, whereas they are in XPs.

Thus, our theory correctly predicts that compounds and phrases are different with respect to primary predication, and derives the fact that the features supported by a subject are not visible/interpreted within compounds.

2.1.2. Secondary predication

We also predict that secondary predication, depictive or resultative, may not be obtained within compounds, as it also is a manifestation of the spec-head configuration, which is not part of X^0s.

(8) a. [He [puts [the book on the shelf]]]. secondary predication on argument
b. *book-shelf-putting is boring

(9) a. They consider [smoking crazy] selected predication
b. *Smoking-crazy-considering is quite common.

(10) a. They eat [fish raw] depictive adjunct
b. *Fish-raw-eating is particular in Japan.

(11) a. They iron [shirts flat] resultative adjunct
b. *Shirt-flat-ironing is what you want to do.

This again follows from our hypothesis that the spec-head configuration is not visible in X^0s at the conceptual interface, while it is in XPs.

3. Aspectual modification

There is a difference in the range of aspectual modification that can be licensed in XPs and in X^0s. For concreteness, we take event structure to be a representation that covers the internal spatiotemporal constituency of a situation or an event denoted by a verbal expression (Vendler, 1967; Comrie, 1976; Dowty, 1979; Bach, 1986; Kipka, 1990; Parsons, 1990; Smith, C. 1991; Verkuyl, 1993). Various features and properties of this spatiotemporal structure can be referred to or modified by affixal or phrasal elements across languages. We examine the following three types of aspectual modification: sequencing, bounding and measuring.

The following paragraphs provide evidence to the effect that while sequencing modification may be licensed in compounds, this is not the case for delimiting and

measuring modification, which can occur only in XPs. This difference, not noticed before in the litterature, brings further support to the differences between words and phrases at the conceptual interface and provides a rationale to the X^0/XP differences with respect to aspectual modification.

3.1. Sequencing modification

Sequencing modification, as a manifestation of the adjunction configuration, may be licensed in X^0s and in XPs. It takes the form of iterative and inverse prefixes in X^0s and of adverbial phrases in XPs.

The scopal difference between XP and X^0 sequencing modification gives rise to the differences in aspectual interpretation (cf. Weshler, 1990; Di Sciullo, 1997; Roeper and Keyser, 1992, 1995).

(12) a. Mary wired a house <u>again</u> (a different house)
b. Mary <u>re</u>wired a house (the same house)

Sequencing modification can be licensed in verbal compounds, via an iterative prefix that modifies a directional preposition, as in (13d), but not the verbal complex formed by a verb and a particle, as in (13c).

(13) a. to turn the chair over again
b. to [turn-over] the chair again
c. *to [re-turn-over] the chair
d. to [re-over-turn] the chair
e. to [re-re-over-turn] the chair

These facts are consistent with the generalization that we established elsewhere, on the basis of Romance data (cf. Di Sciullo, 1996), that external prefixes, such as the iterative or the inverse prefix, c-command the internal prefixes, mainly the directional prefixes. Similar facts, even though more limited, are observed in English.

(14) a. porter/apporter/reporter
b. réapporter/*areporter
c. to lighten/to enlighten
d. to reenlighten/*to enrelighten

e. V
/ \
P_{ex} V
/ \
P_{in} V

Thus, sequencing modification can be lincensed within compounds via prefixation, as an adjunct to a prepositional prefix. This is predicted by our theory since this is achieved via adjunction.

3.2. Delimiting modification

Moreover, in XP structure, a DP or a PP complement may add an endpoint to the event denoted by the verbal projection, as discussed in Tenny, 1988, 1994; Di Sciullo and Klipple, 1994; Pustejovsky, 1995; Di Sciullo, 1997).

It has also been established that the presence of a specific DP object or locative PP may have a delimiting effect on the event denoted by a verbal projection. Where 'delimitedness' refers to the property of an event's having a distinct, definite, and inherent endpoint in time (Tenny, 1994).

(15) a. x ran for one hour/*in one hour. (activity)
b. x ran the mile *for one hour/in one hour. (accomplishment)
c. x drove the car for one hour/*in one hour. (activity)
d. x drove the car <u>to New York</u> *for one hour/in one hour. (accomplishment)

Non referential, non specific cognate objects do not have a delimiting effect.

(16) a. Mary laughed for an hour/*in an hour. (activity)
b. Mary laughed a mirthless laugh (in one minute/for one minute) (Tenny, 1994)
c. John ran a great run (in an hour/for an hour)

This is also the case for deverbal compounds, where the nominal non-head is not delimiting.

(17) a. Marathon-running (in an hour/for an hour) is fun.
b. Pasta-eating (in an hour/for an hour) is hard.
c. Newspaper-reading (in an hour/for an hour) is easy to do.

In compounds, the nominal expression included in a devernal compound cannot have a delimiting effet on the event denoted by the deverbal head because the non-referential, non specific expression is in adjunct position.

Thus, delimiting modification is blocked for deverbal compounds, where the nominal expression within has no effect on the event structure of the deverbal head. This follows in our theory from the fact that the nominal expression is not in a head-complement configuration at the conceptual interface. It is interpreted as non-referential and non-specific in the head-adjunction structure it is a part of at that interface.

3.3. Measuring modification

Measuring modification, as a manifestation of the head-complement configuration, is instanciated by a class of degree modifiers found in XP structure, which modifies the path or the change of state in the event structure.

(18) a. x closes the door <u>partway</u>
b. x walks <u>halfway</u> to New York

The adverb "partway" modifies the final resulting state of the event, to supply a new final state only part of the distance to the original final state. This kind of interpretation has been discussed by Tenny and Heny (1993) and has been represented formally by Parsons (1990).

Measuring modification, unlike sequencing or delimiting modification, may only occur in XP structure. Prefixes which are candidates for a measuring interpretation, such as <u>under-</u>, do not in fact provide that interpretation. Evidence comes from the possible interpretations of the sentences, as well as from the verb classes they may apply to, which is not the same set.

(19) a. x(*under)closed the door partway
b. x(*under)walked halfway to y
c. *x appreciated y halfway/x underappreciated y

Measuring modification is not found within the verbal morphology because it crucially involves the composition of a verb with its complement, which can only occur in syntax. Measuring cannot modify within a word because it requires a complement, a referential specific DP, not found within the word at the conceptual interface.

The following morphological consequences follow from the absence of measuring modification in the basic adjunct and head composition of words. First, deverbal compounds will not include a category that can be interpreted as a path.

(20) a. x mixed the paint completely
b. paint-mixing (*completely) is fun

Second, given D&W's (1987) definition of the head of the word and Kayne's (1994) LCA, the fact that measuring modification requires a complement and that such modification cannot be obtained within the word brings further support to the view that complements cannot be licensed within the word at the conceptual interface. In fact, aspectual modification in derivational morphology is only possible to the left, in adjunct position.

(21) a. to reoverprotect
b. *to reprotectover
c. to overoverprotect
d. *to reprotectover

This follows from our hypothesis that X^0s are canonical adjunct-head configurations at the conceptual interface, while XPs are canonical Spec-head-compl configurations.

4. Referentiality

In XPs, both spec-head and head-complement configurations may support the referential and the specific features for nominal expressions. We predict that these features are not visible at the conceptual interface for X^0 expressions such as compounds, the CTC of which is an Adjunct-head configuration.

Let us first distinguish two sorts of interpretation for a nominal expression, referential and non-referential, as depicted in (22), where the referential interpretation further ramifies into specifics and non-specifics. Let us assume with Diesing's (1994) instanciation of Heim/Kemp's theory, that specific move out of the nuclear scope (the VP), whereas referential non-specifics are subject to existential closure. Let us further assume that non-referential non-specific categories are subject to adjunction.

(22) a. referential: specifics, definites, quantifiers (move out of the nuclear scope)
non-specifics (subject to existential closure)
b. non-referential: non-specifics (subject to adjunction)

The referential and non-referential interpretations are exemplified in (23a,b,c) respectively. The difference in interpretation is reflected in a difference in structure, since while referential specific nominals may undergo passive for instance, this is not the case for non-referential non-specific ones.

(23) a. He ate the pasta. (referential specific)
b. He ate some pasta. (referential non-specific)
c. He ate pasta. (non-referential non-specific)

(24) a. The pasta was eaten.
b. Some pasta was eaten.
c. *Pasta was eaten.

Relevant to our purpose is the fact that a nominal expression included in a compound is non-referential and non-specific.

The lack of referentiality of the nominal expression in the compound is evidenced by the fact that it may not undergo passive, (25b), it may be the antecedent of a pronoun, (25c), and it can be licensed in gapping structures, (25d).

(25) a. John grocery-shopped today.
b. *Grocery was shopped by John today.
c. *His pen$_i$-holder is in marble and it$_i$ is in gold.
d. *Mary likes book-shelving and Paul newspaper.

Matsumoto (1996) observed that the referential as well as the non-referential interpretations were available for cognate objects:

(26) Mary danced a beautiful dance.

(27) a. A certain type of dance, say tango, is famous for its beauty (referential, specific)
b. The result of the activity of dancing is beautiful (referential, non-specific)
c. The activity of dancing is beautiful (non-referential, non-specific)

As expected, passive is not possible with the non-referential, non-specific reading (27c):

(28) a. A beautiful dance was danced by Mary and it was tango.
b. A beautiful dance was danced by Mary and everybody was delighted.
c. *A beautiful dance was danced by Mary and it was never-ending.

Moreover, a cognate object with a non-referential reading cannot be the antecedent of a pronoun nor can it be licensed in a gapping structure, as the following b. examples show.

(29) a. Mary danced a delightful dance, and it, was attractive.
b. *Mary danced a never-ending dance, and it, was attractive.

(30) a. Mary danced a mysterious dance and Jane an attractive dance.
b. *Mary danced a never-ending dance and Jane a sudden dance.

Borer (1994) following Stowell (1989) assume that the difference between referential and non-referential nominals is set in categorial terms. Referential nominals are DPs, while non-referential nominals are NPs. This position is also taken by Matsumoto (1996) to distinguish the interpretations of a cognate object.

(31) a. DP
/ \
D NP
a, some
many, three

b. DP
/ \
D NUMP
/ / \
[e] NUM NP
three, a

c. DP
/ \
D NP
/ / \
[e] NUM N'
three, a

(Borer, 1994)

(32) a. FP
/ \
Spec F'
/ \
F VP
/ \
[+acc] V DP
smiled a beautiful smile
[+acc]

b. FP
/ \
F VP
/ \
V NP
smiled a sudden smile

(Matsumoto, 1996)

However, analyses that capture the differences between referential and non referential nominals in terms of a categorial difference are difficult to hold for compounds, where the definite article, but not a numeral may be licensed. French compounds are ungrammatical with a numeral, compare (33) and (34).

(33) a. un coupe-la-soif, un trompe-l'oeil, un hors-la-loi, un sans-le-sou
a' 'a thurst-quencher','un trompe-l'oeil', 'an out-law', 'a poor-one'
b. *un coupe-une-soif,*un trompe-un-oeil,*un hors-une-loi, un sans-un-sou

The definite article projects a DP structure in these compounds. However, the interpretation of the DP is non-referential and non-specific.

(34) a. Ce vin est un vrai coupe-la-soif.
`This wine is a real thirst-quencher.'
b. *La soif a été coupée par ce vin.
`Thirst was cut by this wine.'
c. *Ce [coupe-la-soif$_i$] est efficace lorsqu'elle$_i$ est persistante.
`This thirst-quencher is good when it is strong.'

What prevents a referential specific interpretation for that DP? We claimed in Di Sciullo (1996b) that the nominal expression included in the compound is an adjunct to D before Spell-Out and at the conceptual interface.

(35)
N
/ \
V N
/ / \
coupe D [e]
/ \
D N
| |
la soif

Given the strong R feature of D in Romance (Longobardi, 1994), the overt adjunction is forced by the strong features of D in the language under consideration and is sustained by the fact that the adjunct structure is not subject to DP movement such as passive, as well as other syntactic properties of DPs.

5. Summary

Thus, notwithstanding the presence of XP structure in X^0s, we show that our theory captures, in configurational terms, basic differences between the interpretation of compounds and the interpretation of phrases with respect to predication, delimiting modification and specific reference at the conceptual interface. As manifestations of the spec-head-compl configuration, these relations and the features they bare are not interpretable in an X^0 adjunct-head configuration at the conceptual interface.

6. Advantages

In this last section, we identify three theoretical advantages of our proposal. The first is that such construction-specific conditions are no longer required in the grammar of compounding.

Thus, to rule out the compounding of subjects, such as in the examples in (5) above, construction-specific constraints have been proposed in the literature, including the First Sister Principle and the Subject Condition.

(36) **First Sister Principle**
All verbal compounds are formed by the incorporation of a word in first sister position of the verb.
Roeper and Siegel (1978: 208)

(37) **Subject Condition**
The SUBJ argument of a lexical item may not be satisfied in compound structure.
(Selkirk 1982:34)

A second theoretical advantage of our proposal is that no independent aspectual ASP node is required in the derivation of words, as the restrictions relative to aspectual modification follow from the configuration, head-adjunction preventing measuring modification to be obtained word internally.

In fact, ASP nodes (Travis, 1992; Borer, 1995) are partial descriptions for aspectual modification. They fail to predict the basic word/phrase asymmetry with respect to aspectual modification, as our configurational theory predicts.

(38)
N'
/ \
N ASPP [+event]
/ \ / \
V -ion Spec ASP'
| / \
destruc ASP VP
/ \
V DP
\
<- the city

In our theory, the bare output conditions impose a strong requirement on the form of words and phrases. The two sorts of grammatical objects must be configurationally distinct at the interface in order to be properly interpreted by the performance systems. One part of the distinctiveness lies on the visibility of the head-complement configuration in phrases and its non-visibility in words. This asymmetry makes the correct predictions with respect of aspectual modification.

A third theoretical advantage of our proposal is that no head-movement is required to account for the lack of specific reference for the nominal expressions included in compounds (Baker, 1988; Bock-Bannema, 1994).

This is a welcome result, as head-movement violates the Uniformity Condition on movement of the Bare Phrase Structure Theory (Chomsky, 1994).

(39) Uniformity
A chain is uniform with respect to its phrasal status.

The chain created by head-movement violates Uniformity since the tail of the chain is a minimal category and the head of the chain is both minimal and maximal.

(40)


Thus, head-movement is not a possible sort of movement and thus is not available for word-formation including compound formation, given Uniformity on the one hand and a configurational definition of categogy types on the other.

References

Baker, N. 1988. *Incorporation*. Chicago : University of Chicago Press.
Borer, H. 1994. "The Projection of Arguments". UMOP 17, 19-47.
Anderson, S. 1993. *A-Morphous Morphology*. Cambridge : Cambridge University press.
Aronoff, 1994. *Morphology by Itself*. Cambridge : The MIT Press.
Chomsky, N. 1995. *The Minimalist Program*. Cambridge: The MIT Press.
Chomsky, N. 1994. *Bare Phrase Structure*. Cambridge: MIT.
Diesing, M. 1994. *Indefinites*. Cambridge. MIT Press.
Di Sciullo, A.M. 1996a. "Atomicity and Relatedness in Configurational Morphology", in A.M. Di Sciullo ed. *Configurations*. Sumerville: Cascadilla Press. pp. 17-41.
Di Sciullo, A.M. 1996b. "Modularity and X^0/XP Asymmetries", *Linguistic Analysis* 26: 3-28.
Di Sciullo, A.M. forthcoming. "Verbal Structures and Variation", *LSRL* 26.
Higginbotham, J. 1985. "On Semantics", *Linguistic Inquiry* 16, 547-593.
Kayne, R. 1995. *The Antisymmetry of Syntax*. Cambridge: The MIT Press.
Lieber, R. 1992. *Deconstructing Morphology*. Chicago : University of Chicago Press.
Longobardi, G. 1994. "Reference and Proper Names", *Linguistic Inquiry* 25: 609-667.
Matsumoto, M. 1996. "Cognate Objects and Referentiality". GISSL Workshop, Girona.
Rothstein, S. 1983. *The Syntactic Forms of Predication*. Ph.D. dissertation, MIT.

Stowell, T. 1989. "Subject, Specifiers and X-bar Theory", Alternative Conceptions of Phrase Structure. M.R. Baltin and A.S. Kroch (eds.), 232-262, The University of Chicago Press, Chicago.
Tenny, C. 1994. *Aspectual Roles and the Syntax-Semantics Interface*, Kluwer, Dordrecht.
Williams, E. 1980. "Predication". *Linguistic Inquiry* 8, 101-139.
Williams, E. 1994. *Thematic Structure in Syntax*. Cambridge, Mass: MIT Press.

Claudio Iacobini - University of Ferrara (Italy) - e.mail: ibc@dns.unife.it

DISTINGUISHING DERIVATIONAL PREFIXES FROM INITIAL COMBINING FORMS

Abstract: The definition of prefix is based upon formal characteristics which do not allow for a clear distinction between prefixes and initial combining forms (icf's). Even though the differentiation between prefixes and icf's can be represented on a continuous scale, arguments are given in favour of a clear-cut distinction between typical prefixes (derivational elements) and typical cf's (compounding elements). A detailed definition of prefix is provided, and derivational affixation is compared with compounding. An explanation is given of the reasons why some icf's can lose their stem-like characteristics and lean towards prefixes to such an extent that they are fully identified with them. This paper ends with a list of Italian prefixes, and of the icf's which are most similar to them.

0. Introduction

This paper focuses on the definition of derivational prefix and on the criteria distinguishing prefixation from compounding and those which distinguish prefixes from initial combining forms (also called prefixoids, formants, neo-classical elements, confixes, etc.).

The commonly accepted definition of prefix — "a bound form that can be productively attached to the left of a word (in the sense of a free lexical element)" — is unsatisfactory because of its vagueness. Such a definition is not sufficient to identify the prefixes of a language in a consistent way. For example, the collection of elements defined as "prefix" in four authoritative grammars and in five dictionaries of contemporary Italian amounts to 106 items, but there is unanimous agreement only on 10 of them in spite of the fact that each source mentions approximately 80 to 90 items.

The main disagreement concerns the classification of elements which some regard as prefixes and others as initial combining forms (henceforth icf's). The principal reason for this disagreement is that the definition of prefix is essentially a formal one. Since prefixed words and words formed by an icf plus a word are similar, both from a formal (bound form / free form) and a functional point of view (non-head / head), we need more precise criteria to characterize prefixes and distinguish them from icf's, in addition to the positional criterion and the syntactic autonomy of the elements.

For this purpose, I will suggest a set of properties defining (proto)typical derivational prefixes. I will then compare the characteristics of affixal derivation to those of compounding. My definitions of prefix and of compound refer to the Italian language, but they may also be extended to the other Romance languages. The distinction between prefixes and icf's applies more generally and concerns other languages as well.

Even though the differentiation between prefixes and icf's can be represented on a continuous scale, I will argue in favour of a clear-cut distinction between typical prefixes (that is, affixes that operate according to derivational rules) and typical cf's (that is, stem-like elements that operate according to compounding rules).

Unlike words and affixes, cf's do not form a natural class; rather, they are a heterogeneous set of foreign elements interacting with the elements and the formative processes proper to a specific language. The wide usage of technical and scientific terms in everyday language, and processes of grammaticalization can make some icf's lean towards prefixation. The icf's which are at a more advanced stage of the grammaticalization process lose their original categorial identity, they show a tendency to play a subordinate role with respect to the other element in a complex word, and tend to be more general in meaning, expressing cognitively relevant, recurring components of meaning which are organized in paradigmatic relations with the meaning expressed by prefixes.

The suggested set of properties defining (proto)typical derivational prefixes makes it possible to draw a list of prefixes in the Italian language. By working out the criteria distinguishing prefixation from compounding, as well as the characteristics typical of cf's, we will be able to reduce to a minimum the number of elements for which there is still a margin of arbitrariness in distinguishing between prefixes and icf's. We will also be able to justify the decision to assign them to either the one kind of element or the other.

1. Theoretical disagreements

The disagreements in defining prefix and prefixation concern both the kind of formative process to which prefixation belongs (derivation vs. compounding), and the kind of lexical element to which prefixes can be ascribed (affix vs. lexeme). With regard to this, there are two main positions within generative morphology. The one we could call "lexicalist" (represented by Aronoff (1976), Booij (1977), Scalise (1984)) clearly distinguishes lexemes from affixes and states that prefixation belongs to derivation (together with suffixation) and not to compounding. The other, which we could call "syntacticist" (represented by Siegel (1974), Allen (1978), Lieber (1981, 1992)), tends not to distinguish derivation from compounding, and to equate lexemes and affixes. Items like *electromagnet* are considered compounds by the former and derivatives by the latter, like words such as *unhappy* or *happiness*. For instance, Lieber (1981) states that *hydro* and *macro*, are "Level II affixes", like the privative prefix *un-*. Therefore, the distinction between prefixes and cf's is of the least importance in the syntacticist view, whereas it is established in principle — but not in detail — in the lexicalist view.

2. Research model

For my research I have adopted the "word based morphology" model, where word means "lexeme" as proposed by Aronoff (1994). However, unlike Aronoff (1976), I believe the lexicon includes units of different kinds, which word formation rules combine to form new words (cf. Selkirk 1982, Corbin 1987). The lexicon, besides lexemes, includes stems and affixes with a proper lexical entry (cf. Lieber 1992). Consequently, affixes do not identify with word formation rules, but more than one affix can be associated to a single word formation rule. The separation between affixes and

word formation rules allows us to maintain the distinction between derivational and compounding processes and at the same time to handle intermediate phenomena which otherwise would be difficult to classify. The classification of prefixes I am suggesting and the distinction between prefixes and icf's is not based on a number of examples selected from those which are clearest, but is the result of an analysis covering all the bound-elements preposed to words in the word formation system of contemporary Italian (cf. Iacobini, 1992).

3. The distinction between prefixation and compounding

The distinction between derivation and compounding, between affixes and lexical elements centred on the syntactic autonomy criterion shows its limits when we consider words formed by two cf's (e.g. *anthropology, geography*) which are rightfully regarded as compounds even if they violate the definition of compounding based on the syntactic autonomy of components.

Scalise (1984: 75) gives several arguments — which I have summarized in the following five points — for these elements to be considered as learned stems, combined in compounds, by postulating properties of affixes that these elements do not share:

a. An affix cannot be a prefix in some words and a suffix in others.
b. Affixes cannot be "factored out".
c. The position of affixes is fixed, that of cf's is not.
d. A free element (= a word) cannot consist only of affixes.
e. Cf's have a suppletive semantic relation with words.

Let us now look carefully at Scalise's arguments one by one.

a) The first point is valid, but it has a limitation. It only applies to the limited number of cf's that can occupy both the initial and the final position with no difference in meaning (e.g. It. *bio* in *bioritmo* "biorhythm" and *anfibio* "amphibious"; *geo* in *geologia* "geology" and *ipogeo* "hypogean"). However, it does not help us to distinguish prefixes from icf's combined with words.

b) Prefixes too can be factored out. For example, Engl. *pro- and enclitics*, Germ. *be- oder entladen* ("charge or discharge"), Dutch *ge- en verboden* ("orders and bans").

c) Scalise gives examples of copulative compounds, a kind of compounding in which the order of elements is not significative. He makes the example *Italo-Anglo-Soviet* compared to *Anglo-Italo-Soviet*. However this property is the consequence of this specific type of compounding and not of the use of cf's. In compounds having the structure specifier/head, the order is meaningful: see It. *logopatia* "logopathy", *logopedia* "logopedics" as compared to *patologia* "pathology" *pedologia* "pedology"; or the order is determined by the fact that the cf's can only occupy one position, either initial or final: see the English *pseudomorph* vs. the impossible word **morphopseudo*.

d) If it is true that outside the foreign-learned vocabulary it is not possible that a word is formed by two affixes only, it is also true that in Italian it is impossible to find words formed by two or more stems (Italian compounds are generally formed by the combination of two free words). Moreover, being foreign to common lexicon is not a feature pertaining only to cf's: in many languages there is a subset of the vocabulary that behaves unusually in certain morphological processes, and consists of loan-words having a "learned" connotation. Besides cf's, it contains clear affixes (e.g. the English suffix *-ity*, which is usually joined to stems of Latin origin, as in *curiosity, profanity*), and words (like the English *curriculum*: pl. *curricula* besides *curriculums*).

e) Scalise's examples (*bio-* meaning "life" and *anthropo-* meaning "man") point to semantic relationships between cf's and words; but there are also semantic relationships between affixes and words, e.g. *pre-* and *before*, *post-* and *after*, *re-* and *again*; and between cf's and words belonging to closed classes (for example, locative prepositions and adverbs): *eso-, ecto-* "out", *endo-* "in", *cata-* "down", *opistho-* "behind".

To sum up, Scalise's criteria, apart from the first (which is useful, but of limited scope), are not very helpful in clearly distinguishing between affixes and cf's. However these criteria give a very useful, yet often overlooked insight: cf's are a heterogeneous set with different characteristics and behaviours.

4. Further complications

Three more phenomena make the distinction between compounding and prefixal derivation more intricate.

First and most important: some icf's can be employed as determiners even before words (and not only in combination with other cf's), that is, in conformity with the typical structure and function of prefixation (bound form / free form) (specifier / head) (e.g. It. *narcotrafficante, euroburocrate, pornoeditore*). Second: prefixes, which are productively combined only with free words, may appear to be combined with bound stems (e.g. Engl. *reduce, produce, remit, commit*). Third: some prefixes are used as free words (e.g. It. *super* meaning "gasoline, petrol").

The third point concerns idiosyncratic behaviours of an extremely small number of prefixes. Rather than prefixes, they can actually be seen as truncated forms. The word-like meaning these self-standing prefixes express is that of the former complex word which underwent a truncation process. The second point stems from diachronic causes. These are words with a Latin origin, not formed according to productive word formation rules. The first point deserves our consideration since it is the kind of formation process involving cf's which most resembles prefixation.

5. The need for a definition of the prefix

We have seen that, in order to distinguish between prefixes and icf's, and to single out a set of homogeneous elements which can be defined as prefixes, the property of syntactic autonomy is not sufficient. Neither is it helpful to work out another criterion that might serve the same purpose on its own, for example, a criterion of etymological or semantic nature. (Reasons of space do not allow us to discuss these matters here).

The reason for this failure is that attempts were made to identify icf's by comparing them with prefixes without having an adequate definition of prefix. Whenever it has not been possible to identify a cf by showing that it may occur as the first and last element of a word with no difference in meaning, an attempt was made to prove that it is a bound element other than a prefix. But this reasoning demands independent criteria for "real" prefixes, as well as a comparison between prefixation and compounding.

In order to establish whether we are dealing with two different processes and kinds of lexical elements, or with one process and one kind of element, it is therefore necessary to draw up the set of properties defining a prototypical model of prefix and of

prefixed word. It is then necessary to compare the behaviour of bound initial elements according to these properties (taking into account tendencies in usage and quantitative and distributional data), so as to see which of those elements conform to the proposed model and thus can be defined "real prefixes". On the other hand, this will enable us to see which elements having behaviours and properties typical of compounding, gradually depart from it, and lean towards the definition of prefix.

In paragraphs 6 to 9 I will suggest a definition of prefixation and of compounding, and I will discuss how combining forms relate to the two formative processes. I have adopted this way of reasoning rather than an intensional definition of cf's because cf's are not a natural class in languages. They are bound stem-like elements combined in compounds belonging to technical and scientific registers. Many of them do not become part of common language, and stay apart from the word formation system. Some cf's become part of the language through grammaticalization and lexicalization processes, and their behaviour range from that of autonomous words to that of derivational affixes (See table in paragraph 8).

6. The defining criteria of prefixes

The following list suggests a set of properties identifying the prefixes productively used in the formation of words of common usage in the Italian language. (It can be extended to the other Romance languages with slight modifications).

1) They are affixes (= bound elements) without a syntactic category (= N, A, V) that attach to the left of a lexical base in order to modify its conceptual meaning.
2) They form new words by attaching to words; they do not productively form words if combined with affixes or other bound elements.
3) They can play a role in the formation of parasynthetic verbs, that is, verbs like the Italian *arrivare* "to arrive", *ingabbiare* "to cage", for which the corresponding non-prefixed verbs — **rivare*, **gabbiare* — do not exist.
4) They occur only in the initial position of a word (a prefix can be preposed to an other prefix within certain restrictions).
5) They act as determiners (they are not in coordinative relation with the word which they are attached to, nor with other prefixes).
6) They select bases mainly according to semantic criteria. They can thus violate the Unitary Base Hypothesis, even in the modified version proposed by Scalise (1984).
7) They express functional-relational meanings (as opposed to both lexical and grammatical ones); pragmatic or encyclopedic knowledge is not necessarily required to interpret the compositional meaning of prefixed words.
8) Their meaning is generally known to all speakers.
9) They form endocentric words of which they are not the head: the category, the gender and other inherent features of prefixed words (e.g. inflection, animacy) remain the same as those of the base word.
10) They can affect all the meanings of the base word or only part of them; in this second case, the characteristics inherited by the prefixed word are a subset of those proper to the base and match the selected meanings.
11) Some prefixes can be used recursively and have (like all prefixes) their semantic and syntactic scope over the whole complex word.

12) They do not move the primary stress of the base.
13) They do not modify the word onset of the base.
14) They can undergo slight phonotactic modifications.
15) They have restrictions in length.

7. Compounding

Let us now look briefly at the definition of compounding. (Based on Italian primary compounding).

- Compounding combines two stem-like elements.
- The elements of compounds have an independent lexical meaning.
- The semantic relation between the elements is not explicitly signalled; consequently, the meaning of compounds allows for several possible interpretations.
- A head-element can be a non-head in another compound.
- The typical relation between constituents is subordinative, but coordinative or exocentric constructions are also possible.

8. Different kinds of combining forms

By comparing the properties of compounding with those of affixal derivation, it is possible to work out a set of criteria which help us to identify cf's and their behaviour with respect to word formation processes.

The criteria of comparison I have used between compounding and affixal derivation are as follows:

1) Position: the affix has a fixed position, the constituent of compound does not.
2) Combinatorial capacity: the affix does not combine with other affixes; the constituent of compound combines with derivational and/or flexional affixes.
3) Relation between constituents: derivation only allows for a subordinative relationship, compounding also allows for a coordinative relationship.
4) Head position:
   a) derivatives are always right-headed, productive compounds are left-headed;
   b) both derivatives and compounds are typically endocentric.
5) Semantics: affixes have a relational meaning which requires a complementing semantic feature, and they clearly indicate the semantic relation between the elements of the derived word. The constituents of compounds have a denotative lexical meaning, which is semantically autonomous; the relation between the constituents is not clearly expressed; as a consequence, the meaning of compounds is much less precisely determined compared to that of derivatives.
6) Formal criterion: the affix is a bound element; the constituents of compounds are free elements.

On the basis of the above criteria, it is possible to draw the following table (worked out on the basis of Masseroli 1994:283) which depicts the behaviour of cf's as compared to the two poles consisting of prototypical derivational affixes and prototypical constituents of compounds.

Criteria	1	2	3	4a	4b	5	6
Prototyp. Affix	a	a	a	a	e	a	a
1) endo-, oligo-	a	a	a	a	e/x	a	a
2) crio-, igro-	a	a	a	a	e	c	a
3) agro-, bio-	a	a	c	a	e	c	a
4) lipo-, melano-	a	c	a	a	e	c	a
5) ippo-, mio-	a	c	c	a	e/x	c	a
6) -teca, -cida	a	a	a	a	e	a/c	a
7) -crazia, -machia	a	a	a	a	e	c/a	a
8) -antropo-, -geo-	c	c/a	c	a	e/x	c	a
9) -blasto-, -teo-	c	c	a	a	e/x	c	a
10) -termo-, -crono-	c	c	c	a	e/x	c	a
11) -socio-, -anglo-	c/a	c	c	a/c	e/x	c	a
12) -auto-, -foto-	c	c	c	a/c	e/x	c	c
Prototyp. Comp.	c	c	c	c	e	c	c

In this table, "a" indicates a characteristic typical of derivational affixes, "c" indicates a characteristic typical of constituents of compounds, except for column 4b, where "e" indicates endocentric formation and "x" exocentric formation. Where "c/a", "a/c", "e/x" appear, the first characteristic prevails over the second.

9. A list of prefixes

For reasons of space, my comments on the table will be limited. The table shows clearly the multiple behaviours possible in the whole set of cf's. Many cf's, though they are bound-forms and occupy a fixed position, have the typical behaviour of constituents of compounds, particularly with regard to semantics and the kind of relationships they have with the other members of a complex word. The conclusion we may draw is that the properties of some cf's cannot be applied to them all, and that cf's as a whole cannot be equated to affixes.

Restricting my comments to icf's, and returning to the issue of their distinction from prefixes, the data in the table show us that the icf's most similar to prefixes are those in group 1).

In addition to the properties included in the table, there are two more that help distinguish icf's from prefixes:

- Whether it is possible for them to combine with words and/or with cf's.
- Which register they belong to: everyday language as opposed to technical and scientific languages.

The great majority of words formed with cf's result from a combination of two cf's and belong to the technical and scientific register. In most cases, the meaning of cf's is only known to the people specialized in a particular field. When a similar word is used in everyday language (e.g. It. *telefono* "telephone", *frigorifero* "refrigerator", *idrogeno* "hydrogen"), the speakers are not usually able to distinguish the elements that make it up, and to give them their proper meaning. However, when a cf is attached to the left of several different free words (e.g. It. *ecoalimento, ecomercato, ecoprodotto, ecoterrorismo; protobarocco, protolingua, protoscienza, protoindustriale*) it is more

probable that speakers will be able to segment this kind of complex word. This is because they will recognize at least one familiar unit that is the word, and that is moreover the head of the formation. The compositional meaning of the "Icf + Word" type is easier to understand than the type consisting of two cf's. Since the base word is easier to understand, it will also be easier to interpret the meaning of the cf. So, the more an icf is used before words in formation of common usage, the more likely it is to become part of the competence of speakers. The structural analogy between this kind of formation and prefixation helps to associate (and in some cases identify) the icf's involved and prefixes.

Keeping these last two points in mind, the margin of uncertainty between prefixes and icf's narrows. And now the most probables candidates to prefix status are (within those depicted in group number 1) the icf's which mainly combine with words, which are known and used by the majority of speakers, and express a functional and relational meaning. Assigning these elements to the category of prefixes or to that of cf's requires some arbitrary judgement. (However, this only concerns a small number of items). Among these are a few which share all the defining properties of prefixes, even though sometimes they are used in combination with cf's to coin technical and scientific terms. Therefore, in a synchronic perspective, the following items can be considered as prefixes:

auto-, macro-, mega-, micro-, multi-, neo-, paleo-, para-, pluri-, poli-, semi-.

On the other hand, the other icf's belonging to group 1), do not fully match the defining properties of prefixes, even though they have many characteristics in common with prefixes. These icf's differ from prefixes in that they usually combine with other cf's (as well as words), they are not as general and diffuse in meaning as prefixes are, and they have not been mastered by the majority of speakers. They are:

archeo-, ecto-, emi-, endo-, equi-, eso-, etero-, filo-, meso-, oligo-, proto-, pseudo-, tauto-, vetero-.

The full list of Italian prefixes, i.e. of those elements which comply with the 15 characteristics shown in paragraph 6, is therefore as follows:

a(n)-, ad-, ante-, anti-[1] (e.g. *antieroe*), anti-[2] (e.g. *antivigilia*), arci-, auto- (e.g. *autogestione*), avan-, circum-, cis-, co-, con-, contro-, de-, dis-, ex-, extra-, in-[1] (e.g. *ingiallire*), in-[2] (e.g. *ingiusto*), infra-, inter-, intra-, iper-, ipo-, macro-, maxi-, mega-, meta-, micro-, mini-, multi-, neo-, oltre-, paleo-, para-, pluri-, poli-, post-, pre-, pro-, ra-, re-, retro-, ri-, rin-, s-, semi-, sopra-/vra-, sotto-, stra-, sub-, super-, sur-, trans-, ultra-, vice-.

Besides these, we should mention those prefixes which are no longer used to form productively words of common usage, but which are still present and easily detectable in many Italian words. (Some of them, however, are employed to coin technical and scientific terms). A full list follows:

ab-, ambi-, anfi-, archi-, bis-, circon-, citra-, contra-, di-, dia-, e-, es-, estra-, estro-, fra-, giusta-, intro-, mis-, ob-, per-, se-, so-, sor-, su-, tra-, tras-

10. Conclusion

In this talk I have argued that prefixation and composition are two separate processes involving elements of different kinds.

Cf's form a heterogeneous set which can interact both with the lexical elements and the word formation processes of a language. Their presence in compounds and their bound status are favourable conditions for their grammaticalization. The icf's expressing recurring components of meaning in paradigmatic relation with the meaning expressed by prefixes, and that are employed several times as determiners preposed to a word, can become so similar to prefixes as to fully identify with them.

An accurate definition of the notion of prefix, the comparison between such definition and the one identifying the constituent of compound, and a detailed analysis of the behaviour of initial bound-elements have made it possible

- to define the set of Italian prefixes
- to identify, motivate and circumscribe the inevitable overlapping between processes and formative elements that are actually different.

REFERENCES

Allen, M.R. (1978). *Morphological investigations*, Doctoral Dissertation, University of Connecticut.

Aronoff, M. (1976). *Word formation in generative grammar*, Cambridge (Mass.), MIT Press.

Aronoff, M. (1994). *Morphology by itself: stems and inflectional classes*, Cambridge (Mass.), MIT Press.

Booij, G.E. (1977). *Dutch morphology. A study of word formation in generative grammar*, Lisse, Peter de Ridder Press.

Corbin, D. (1987). *Morphologie dérivationelle et structuration du lexique*, Tübingen, Niemeyer.

Iacobini, C. (1992). *La prefissazione nell'italiano contemporaneo*, Unpublished Doctoral Dissertation, University of Rome "La Sapienza".

Lieber, R. (1981). *On the organization of the lexicon*, Doctoral Dissertation, MIT, Cambridge (Mass.), reproduced by the Indiana University Linguistic Club.

Lieber, R. (1992). *Deconstructing morphology*, Chicago, The University of Chicago Press.

Masseroli, A. (1994). *Le semiparole nella lingua italiana contemporanea*, Unpublished Doctoral Dissertation, University of Pavia.

Scalise, S. (1984). *Generative morphology*, Dordrecht, Foris.

Selkirk, E.O. (1982). *The syntax of words*, Linguistic Inquiry, Monograph Seven, Cambridge (Mass.), MIT Press.

Siegel, D. (1974). *Topics in English morphology*, Doctoral Dissertation, MIT, Cambridge (Mass.).

Takae Tsujioka and Yoshiho Shibuya
Georgetown University
tsujiokt@gusun.georgetown.edu, shibuyay@gusun.georgetown.edu

INALIENABLE OBJECT CONSTRUCTION IN JAPANESE*

This paper discusses the Inalienable Object (IO) construction in Japanese and argues that an adequate analysis of this construction necessitates postulating post-lexical incorporation and adjectivization. It will be shown that, while the IO construction in Japanese ostensibly resembles ordinary SOV sentences, it exhibits a number of properties which depart from those of the canonical verbal construction. The analysis which captures all the properties of the IO construction crucially depends on the modular conception of word formation found in Shibatani and Kageyama (1988), thereby constituting a piece of evidence for the theory.

1. Introduction

The locus of word formation has been an issue of great controversy in morphological theories. Positions vary from a strong lexicalist view, which takes the lexicon to be the only place of word formation (Jensen and Stong-Jensen, 1984; and Di Sciullo and Williams, 1987 among others), to a modular theory of word formation, which argues that word formation takes place both at lexical and post-lexical levels[1] (Shibatani and Kageyama, 1988; Kageyama, 1993).

The strong lexicalist hypothesis maintains that phrasal syntax cannot have access to the internal structure of words. Thus, word formation rules such as compounding and conversion can never occur during the course of syntactic derivation whether such derivation is in the overt syntax or at LF. On the other hand, the modularist theory allows the morphological rules to access multiple grammatical components, presumably all of the following: the lexicon, syntax (overt or covert), and phonology.

This paper presents supportive evidence for the latter position, a modular theory of word formation. The particular construction we discuss is the Inalienable Object (IO) construction in Japanese. The IO construction may best be understood as a phrasal counterpart of the English Possessional Adjective such as *blue-eyed* and *pretty-faced*. The analysis called for by the Japanese IO construction casts doubt on the strong lexicalist hypothesis, since it involves word formation outside of the lexical component, in the course of syntactic derivation. On the other hand, the proposed

* We gratefully acknowledge help and comments of Héctor Campos, Donna Lardiere, Paul Portner, Mitsuhiko Ota, Mary Rose, Raffaella Zanuttini, and Lisa Zsiga. Needless to say, all errors are our own.

[1] Here, the term "lexical" is used to refer to forms created in the lexicon, whereas "post-lexical" is used to refer to forms created in other components such as syntax or phonology.

analysis of the IO construction is compatible with the modularist conception of word formation.[2]

In Section 2, we will first introduce the general structure of the IO construction. In Section 3, we will show that, while the IO construction in Japanese at first glance resembles ordinary SOV sentences, it exhibits many properties which differentiate it from canonical verbal construction. In Section 4, we will then argue that the only plausible analysis which captures all the properties of the IO construction crucially depends on the modular conception of word formation. Specifically, we claim that the IO construction is a type of light verb construction, and it instantiates cases of incorporation and adjectivization that are not LEXICAL but POST-LEXICAL. Section 5 serves as a summary and conclusion.

2. Inalienable Object (IO) construction

Examples of the Japanese IO construction are given in (1).[3]

(1) a. ao-i me-o si-ta syoozyo
blue-AI eye-ACC do-PART girl
'blue-eyed girl'

b. syoozyo-ga ao-i me-o si-te-i-ru
girl-NOM blue-AI eye-ACC do-PART-be-PRS
'The girl is blue-eyed.'

The IO construction can be construed in two ways: pre-nominal modification and predicative function. When it appears pre-nominally, it has the general pattern shown in (2a), where NP_1-ACC followed by *si-ta* is an inalienable property (e.g., attribute or body-part noun) of N_2. When the IO construction serves a predicative function, the pattern in (2b) is used, in which NP_2 is the subject and the predicate consists of NP_1-ACC followed by *si-te-i-ru*. Here, again, NP_1 denotes an inalienable property of NP_2.

(2) a. Pre-nominal IO: $[NP_1$ modifier+$N_1]$-ACC si-ta N_2
b. Predicative IO: NP_2-NOM $[NP_1$ modifier+$N_1]$-ACC si-te-i-ru

Note that the citation form for both *si-ta* and *si-te-i-ru* is generally considered to be *suru* 'do.' We must stress that this citation form never occurs in the relevant construction. Nonetheless, we will refer to *si-ta* and *si-te-i-ru* collectively as *suru* for expository purposes, and we will postpone our discussion of the actual forms *si-ta* and *si-te-i-ru* until later (See Section 3.5.).

The IO construction in Japanese is semantically similar to the English Possessional Adjectives such as *blue-eyed* and *pretty-faced*. The entire construction modifies an entity by defining some inalienable property of that entity. That the property in question must be inalienable is illustrated by the ungrammaticality of (3).

(3) a. *[siro-i ie]-o sita onnanoko
white-AI house-ACC do-PART girl
'girl with a white house'

b. *[white-house]-d girl

3. Properties of the IO construction

3.1. *Suru* is a light verb

To our knowledge, little attention has been given to the Japanese IO construction in the literature, and existing studies of it are superficial. These equate the IO construction with ordinary SOV sentences. For example, Kageyama (1993) and Uchida and Nakayama (1993) make no distinction between (4a), which we call the IO construction, and (4b), which exemplifies an ordinary SOV sentence.

(4) a. Mary-ga ao-i me-o si-te-i-ru. (IO construction)
M.-NOM blue-AI eye-ACC do-PART-be-PRS
'Mary is blue-eyed.'

b. Mary-ga kirei-na iyaringu-o si-te-i-ru. (Ordinary SOV)
M.-NOM pretty-AI earrings-ACC do-PART-be-PRS
'Mary is wearing pretty earrings.'

Note that *suru* in (4b) has the sense of "wear," and it assigns the theta-roles (AGENT, THEME) to its arguments. In this sense, *suru* in (4b) is a regular full-fledged verb. However, characterization that *suru* in the IO construction is a full-fledged verb is counter-intuitive. Rather, it seems that the subject NP gets a THEME theta-role from the accusative-marked NP, and *suru* is void of meaning. This becomes clear when we try to substitute *suru* with a content verb which has a similar meaning. As the ungrammaticality of (5) indicates, no content verb can replace *suru* in (5). On the other hand, (6) shows that a heavy verb *suru* can be substituted by a content verb.

(5) *Mary-ga ao-i me-o mot/tuke -te-i-ru.
M.-NOM blue-AI eye-ACC have/wear -PART-be-PRS
'Mary has blue eyes/is blue-eyed.'

(6) Mary-ga kirei-na iyaringu-o mot/tuke/hame -te-i-ru.
M.-NOM pretty-AI earrings-ACC have/wear/put-on -PART-be-PRS
'Mary had/was wearing/put on pretty earrings.'

The contrast between (5) and (6) indicates that *suru* in the IO construction, unlike *suru* in ordinary SOV sentences, neither assigns theta-roles nor has a specified meaning on its own.

Traditionally, this type of *suru* which is void of meaning and does not have its own argument structure has been called a "light verb" in Japanese. Discussions of light verbs are particularly prevalent in the literature on the Verbal Noun (VN) construction in Japanese, which we will come back to in Section 4.1. Here, let us note that the presence of the light verb *suru* requires a special mechanism of theta-role assignment and any analysis of the IO construction should take it into account.

---

[2] It must be noted that the type of phrasal compounds we discuss in this paper differs in shape from the "post-syntactic compounds" introduced in Shibatani and Kageyama (1988) given below:

(i) [Amerika:hoomon] -no ori cf. Amerika-o hoomon-no ori
America visit GEN occasion ACC
'an occasion of visiting America'

For Shibatani and Kageyama (1988), the exclusion of case particles is an important characteristic of post-syntactic compounds, as opposed to purely syntactic constructions. We do not share this view. The use of inflectional affixes 'frozen' inside words is widely attested (as in German Fugenelementen (Beard, 1995)). According to our analysis, the accusative case particle, which has lost its grammatical function, does seem to appear inside phrasal compounds. Nonetheless, we are in agreement with Shibatani and Kageyama in their general grammatical model.

[3] Gloss abbreviations: AI = adjectival inflection, PART = participle, NOM = nominative case, ACC = accusative case, GEN = genitive case, COP = copula, TOP = topic, PRS = present tense, PST = past tense, COMP = complementizer.

3.2. NP-ACC and *suru* are inseparable

We now turn to the second property of the IO construction which differentiates it from ordinary SOV sentences. That is, the accusative-marked NP in the IO construction cannot be separated from the light verb *suru*. For example, the accusative-marked NP is invisible to a syntactic movement operation such as scrambling, a common phenomenon in Japanese, by which the constituents of a sentence are freely rearranged via IP adjunction (Saito, 1985).[4] However, scrambling cannot be applied to the accusative-marked NP in the IO construction. Compare (7) and (8).

(7) *ao-i me-o$_i$ Mary-ga t$_i$ si-te-i-ru.
blue-AI eye-ACC M.-NOM do-PART-be-PRS
'Mary has blue eyes/is blue-eyed.'

(8) kirei-na iyaringu-o$_i$ Mary-ga t$_i$ si-te-i-ru.
pretty-AI earrings-ACC M.-NOM do-PART-be-PRS
'Mary is wearing pretty earrings.'

The ungrammaticality of (7) indicates that scrambling of the accusative-marked NP is disallowed in the IO construction. On the other hand, scrambling of the object NP is possible in an ordinary SOV sentence, as in (8). Any analysis of the IO construction should address the reason why the accusative-marked NP in the IO construction does not undergo movement.

3.3. ECM compatibility

A third fact peculiar to the IO construction is that it can serve as a complement clause in the so-called ECM construction. In Japanese, the subject of an embedded clause can be marked exceptionally as accusative. However, the embedded predicates that allow this possibility are limited to adjectives and nominal + copula *da* (Kuno, 1973). Hence (9a) and (9b) are grammatical, whereas, (9c), in which a verb appears as the complement predicate, is ungrammatical.

(9) a. John-wa [Mary-**o** kawai-i] to omot-ta. (adjective)
J.-TOP M.-ACC pretty-AI COMP think-PST
'John thought Mary to be pretty.'

b. John-wa [Mary-**o** tensai-da] to omot-ta. (nominal + copula *da*)
J.-TOP M.-ACC genius-COP COMP think-PST
'John thought Mary to be a genius.'

c. *John-wa [Mary-**o** gakkoo-ni ki-ta] to omot-ta. (verb)
J.-TOP M.-ACC school-to come-PST COMP think-PST
'John thought Mary to have come to school.'

Now, observe the asymmetry between (10) and (11).

(10) John-ga [Mary-**o** ao-i me-o si-te-ir-u] to omot-ta.
J.-NOM M.-ACC blue-AI eye-ACC do-PART-be-PRS COMP think-PST
'John thought Mary to be blue-eyed.'

(11) *John-ga [Mary-**o** kirei-na iyaringu-o si-te-ir-u] to omot-ta.
J.-NOM M.-ACC pretty-AI earrings-ACC do-PART-be-PRS COMP think-PST
'John thought Mary to be wearing pretty earrings.'

In (10), the IO construction appears as the complement clause of the ECM construction, whereas in (11) an ordinary SOV sentence is embedded in the ECM construction. The grammaticality of (10) and the ungrammaticality of (11) show that the IO construction, but not an ordinary SOV sentence, is compatible with ECM. That the IO construction is compatible with ECM is, in fact, surprising for the following two reasons: (i) (10) contains two accusative marked NPs, which violates the "Double *o* Constraint" (Harada, 1973), a strict ban against two accusative-marked NPs in a single clause, and (ii) the complement predicate in (10) does not appear to be an adjective nor a nominal + copula. A satisfactory analysis of the IO construction should address both the reason why the IO construction can override the Double-*o* constraint and why its predicate behaves similarly to adjectives or nominal + copula.

[4] Except for the verb, which must be placed at the end of the sentence.

3.4. Modification by degree adverbs

The forth property unique to the IO construction concerns adverbial modification. Consider (12).

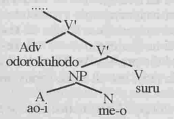
(12) a. Mary-ga odorokuhodo ao-i me-o si-te-i-ru.
M.-NOM to a surprising degree blue-AI eye-ACC do-PART-be-PRS
Reading 1: 'Mary is very blue-eyed.'
Reading 2: 'Mary has very blue eyes.'

b. odorokuhodo Mary-ga ao-i me-o si-te-i-ru.
to a surprising degree M.-NOM blue-AI eye-ACC do-PART-be-PRS
Reading 1: 'Mary is very blue-eyed.'

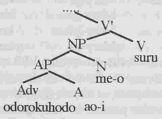
The sentence in (12a) is ambiguous. In the first reading, the adverb *odorokuhodo* 'to a surprising degree' has scope over the entire predicate, adjoined to the projection of *suru*. In the second reading, it takes a narrower scope which covers only the pre-nominal adjective *ao-i* 'blue.' In the latter case, the adverb is adjoined to the projection of the adjective, rather than to the projection of *suru*. The relevant structural configurations are illustrated in (13a) and (13b), respectively.[5]

(13)

a. Reading 1: wide scope



b. Reading 2: narrow scope



Notice that when *odorokuhodo* appears sentence initially, the second reading disappears, as in (12b). The unavailability of the second reading can be attributed to the fact that *odorokuhodo* is more deeply embedded for this particular reading, as shown in (13b). Extraction from that position is not licensed, as it breaks a constituent (i.e., NP).

That the structure in (13a) is possible is a property specific to the IO construction. The relevant class of degree adverbs modify mainly adjectives and adverbs, but not verbs. Consider (14) for an illustration of this point.

(14) a. Mary-ga odorokuhodo kirei-na iyaringu-o si-te-i-ru.
M.-NOM to a surprising degree pretty-AI earrings-ACC do-PART-be-PRS
'Mary is wearing very pretty earrings.' (= Reading 2: narrow scope)

b. *odorokuhodo Mary-ga kirei-na iyaringu-o si-te-i-ru.

[5] We assume a minimal tree structure following Chomsky (1995).

to a surprising degree M.-NOM pretty-A1 earrings-ACC do-PART-be-PRS

That (14a) is grammatical is trivial since its structure should parallel the one in (13b), not (13a). Notice that there is only one reading available for (14a), namely, the narrow scope reading. When the adverb is fronted and the narrow reading is no longer available, the sentence becomes ungrammatical as in (14b). Thus, we can conclude that the wide-scope reading of degree adverbs is not available for ordinary SOV sentences. Why then is modification by the same adverb possible for the IO construction? An analysis of the IO construction should answer this question.

3.5. *Suru* is obligatorily in participial form

Finally, let us turn to a discussion of the form of *suru* in the IO construction. The possible and impossible forms of *suru* in the two types of IO construction are shown in (15a) and (15b), respectively. Compare the patterns in (15) with those in (16).

(15) *suru* in IO construction

a. Pre-nominal IO:

ao-i me-o	***su-ru***	Mary
ao-i me-o	**si-ta**	Mary
ao-i me-o	**si-te-i-ru**	Mary
ao-i me-o	**si-te-i-ta**	Mary

b. Predicative IO:

Mary-ga ao-i me-o ***su-ru***
Mary-ga ao-i me-o ***si-ta***
Mary-ga ao-i me-o **si-te-i-ru**
Mary-ga ao-i me-o **si-te-i-ta**

(16) *suru* in ordinary SOV

a. Pre-nominal:

kirei-na iyaringu-o	**su-ru**	Mary
kirei-na iyaringu-o	**si-ta**	Mary
kirei-na iyaringu-o	**si-te-i-ru**	Mary
kirei-na iyaringu-o	**si-te-i-ta**	Mary

b. Predicative:

Mary-ga	kirei-na iyaringu-o	**su-ru**
Mary-ga	kirei-na iyaringu-o	**si-ta**
Mary-ga	kirei-na iyaringu-o	**si-te-i-ru**
Mary-ga	kirei-na iyaringu-o	**si-te-i-ta**

Notice that *su-ru* in both the pre-nominal IO and the predicative IO, as well as *si-ta* in the predicative IO are not available,[6] as shown in (15), but all these forms are possible in ordinary SOV sentences, as shown in (16).

Let us examine the pattern in (15) more closely. First, notice that *si-ta* in pre-nominal IO is possible, whereas *si-ta* in the predicative IO is not, though in both cases the form *si-ta* is used. Second, notice that the grammatical *si-ta* in the pre-nominal IO does not have the past tense interpretation which is normally associated with the function of *-ta* in Japanese. Third, the *su-ru* form is excluded in both the pre-nominal IO and the predicative IO. Fourth, in the predicative IO, the presence of *-i-ru* /*-i-ta* 'be' is obligatory. Finally, the pre-nominal IO allows the *si-ta* form, as well as the *si-te-i-ru/ta* form, while the same does not hold for the predicative IO.

These observations can be explained if we postulate that *suru* in the IO construction must take the participle forms, *si-ta* ~ *si-te*, assuming that *-ta* , similarly to English *-ed*, is both the past tense suffix and the participle suffix. First, that *si-ta* is allowed in the pre-nominal IO but not in the predicative IO receives a natural account: bare participles

[6] The pattern is much more rigid than those of the so-called Type 4 verbs (Kindaichi, 1976)

may appear in a reduced relative clause, but not in a matrix clause. Second, the ungrammaticality of *su-ru* follows, as the form *su-ru* is not a participle form.[7] If we assume that *-te* before *-i* 'be' is an allomorphic variant of *-ta*,[8] the reason that the forms *-i-ru* /*-i-ta*'be' are obligatory in the predicative IO is clear: in the predicative IO, *su-ru* in the participle form must be supported by a tensed auxiliary. Finally, we can now properly characterize the two types of pre-nominal IO: *si-ta* appears in a reduced relative clause, whereas *-i-ru* /*-i-ta* 'be' appear in full (i.e., tensed) relative clauses.

We have argued that the various restrictions on the form of *suru* which are specific to the IO construction amount to the fact that a participle form is obligatory for *suru* in this construction. In Section 4.2, we will discuss why this should be the case.

3.6. Summary to Section 3

The following summarizes all the properties of the IO construction we have discussed so far.

Properties of the IO construction

A. *Suru* is a light verb.
B. Accusative-marked NP is not extractable.
C. The IO construction overrides the Double *-o* constraint.
D. The IO construction overrides the restriction on the complement predicate in ECM
E. Degree adverbs can modify *suru*.
F. *Suru* must take a participial form.

4. Analysis

We have so far demonstrated that the IO construction departs from canonical SOV sentences in a number of ways. It is our task now to propose an analysis which can account for the observed properties. We believe that the properties of the IO construction listed in Section 3.6 can be divided into the following two groups: (i) those which can be explained if we postulate that the NP-ACC and *suru* form a unit tighter than ordinary object-verb pair, and (ii) those which follow if we posit that the IO predicate is an adjective. The properties A, B, and C above constitute the first group, and D, E, F, the second group. We will argue that the two groups of properties, follow from the postulation of two post-lexical morphological operations, INCORPORATION and ADJECTIVIZATION, respectively. We will show that both operations are required for theta-role discharge, and thus, can independently be motivated. In accordance with the modularist view, but contradicting the strong lexicalist view, we will conclude that the proposed incorporation and subsequent adjectivization necessarily occur at a post-lexical level since the incorporated nominal in this case is a phrasal compound of the type Adjective + Noun.

[7] It seems that the distinction between the present participle form *-ing* and the past participle form *-ed* is neutralized in Japanese, and *-ta* is used uniformly in both cases. In passing, *-ta* does not mark "passive participles" since Japanese has an independent passive morpheme *-rare*.

[8] One supportive fact for this treatment of *-ta*~*-te* variation is that the verbal stem form is identical in two cases.

4.1. Incorporation of A + N compound

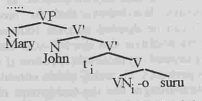
We argue that the first group of properties can be captured if we postulate that the light verb *suru* induces an incorporation of the theta-role-bearing nominal. In order to facilitate our discussion, let us briefly refer to a parallel analysis proposed for another type of the light verb construction: the Verbal Noun (VN) construction. Verbal nouns, similar to English deverbal nominals such as *destruction* and *examination*, can take arguments and assign theta-roles to the arguments. An example of VN construction is given in (17).

(17) Mary-ga John-to kaiwa-o su-ru
M. NOM J.-with conversation-ACC do-PRS
'Mary talks with John.'

In (17), the theta-roles assigned to the NPs *Mary* and *John* come from the VN *kaiwa* 'conversation,' not the light verb *suru*. *Suru*, therefore, does not assign any theta-role of its own.

Assuming a strict locality condition for theta-role assignment (Chomsky, 1981) such that the arguments must appear inside the projection of the head in order to receive a theta-role, (17) is puzzling, at first glance. In (17), *Mary* and *John* seem to be base-generated inside the projection of the light verb *suru*, not the VN. There must be an operation which enables VN to transfer its theta-roles to the (otherwise empty) argument structure of the light verb. In recent works under the Minimalist Program (Chomsky, 1995), the process has been identified with Baker's (1988) incorporation. Saito and Hoshi (1994), as well as Dubinsky (1994), argue that VN in (17) undergoes LF incorporation into *suru*. This process is illustrated in (18).

(18)



We do not have space to reproduce the detail of the LF incorporation analyses. Here, we want to stress that if incorporation is tied to the requirement of theta-role discharge, it must be the case that it is always triggered in the light verb construction. As we have observed earlier, *suru* in the IO construction, too, is a light verb. The THEME argument, which is assigned to the external argument in the predicative IO construction, comes from the inalienable property NP, not from the verb *suru*. *Suru* in the IO construction incorporates the inalienable property nominal. The resulting complex predicate inherits the argument structure of the nominal in much the same way *suru* inherits the argument structure of verbal nouns via incorporation.

The extension of the incorporation analysis to the IO construction enables us to capture the properties A, B, and C of the IO construction given in Section 3.6. First, that *suru* is a light verb (property A) motivates the incorporation analysis. Second, that the accusative marked nominal is not extractable (property B) results from the fact that the incorporated nominal is invisible to movement operations. Third, that the Double *o* constraint does not hold (property C) receives an explanation as the accusative case

marker, attached to the incorporated nominal, can be said to be "frozen" and lose its ordinary syntactic function.[9]

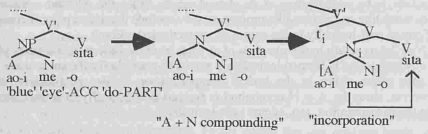
There are two questions we need to address when we apply the incorporation analysis to the IO construction: Q1: What exactly incorporates into the light verb *suru*? Q2: At what level does the incorporation take place? Let us first address Q1. Incorporation is a head-to-head operation. In the case of VN construction, what incorporates into the light verb *suru* is the verbal noun itself, and nothing else. In the case of the IO construction, too, we assume that a nominal head incorporates into *suru*. However, there is evidence that the incorporated nominal in this case is a compound of the type Adjective + Noun (A + N). Backward gapping in Japanese is known to respect morphological integrity but not syntactic constituency and, therefore, can serve as a strong "wordhood" test (Kageyama, 1989). Deleting a noun is normally allowed without deleting the adjective which modifies it. However, the same does not hold in the IO construction. Hence, the contrast between (19) and (20).[10]

(19) *Mary-wa [ao-i ~~me-o~~ ~~si-te-i-ru-si~~, (A + N in IO)
M.-TOP [blue-AI eye-ACC] do-PART-be-PRS-and
Jane-wa [kuro-i me-o] si-te-i-ru
J.-TOP [black-AI eye-ACC] do-PART-be-PRS
'Mary has blue eyes, and Jane, black ones'

(20) Mary-wa kirei-na ~~iyaringu-o si-te-i-ru-si~~, (Regular A and N)
M.-TOP pretty-AI earrings-ACC do-PART-be-PRS-and
Jane-wa kawai-i iyaringu-o si-te-i-ru
J.-TOP cute-AI earrings-ACC do-PART-be-PRS
'Mary is wearing pretty earrings, and Jane, cute ones'

The *Backward Gapping* test indicates that A + N in the IO construction form a morphological unit, that is, a compound. We thus propose that what undergoes incorporation is this A + N compound. (21) demonstrates the relevant derivations.

(21)



Since the A + N compound is the input of incorporation, we need to determine when it is formed before we can address Q2, that is, when the incorporation takes place. There are various facts which suggest that the relevant compound formation takes place post-lexically, operating on phrases. We mention two such properties here.

First, the phonological properties of the A + N compound in the IO construction is very different from those of lexical compounds. Lexical compounds in Japanese often

[9] cf. footnote 2.

[10] To ensure a backward gapping reading, a pause is required between the adjective and the noun in the second conjunct. It is with this pause that (19) is unacceptable. Without the pause, the sentence is grammatical but the resulting reading is "Mary is blue, and Jane is black-eyed."

undergo "sequential voicing" or *Rendaku* (Kubozono, 1993, 1995; Vance, 1987). Rendaku is a process where the first sound of the second element of the compound, generally a native Japanese word, becomes voiced, as illustrated in (22). As shown in (23), however, this process does not take place in the case of the A + N compound in the IO construction.

(22) maru + kao --> [marugao] (k > g)
'round' 'face' 'round-face'

(23) maru-i + kao-o --> [maru-i-kao-o] (*maru-i-gao-o)
'round' 'face-ACC' 'round-face-ACC'

Second, while lexical compounds are subject to the Anaphoric Island Constraint (AIC) which disallows anaphoric and deictic pro-forms to appear word-internally (Postal, 1969), the A + N compounds in the IO construction are not. Observe the difference between (24) and (25).

(24) a. [sara-arai]-wa muzukasi-i.
dish-washing-TOP difficult-AI
'Dish-washing is difficult.'

b. *[sore¡-arai]-wa muzukasi-i.
it-washing-TOP difficult-AI
'*It-washing is difficult.'

(25) a. uti-no kabe-wa [ao-i-iro-o] si-te-ir-u.
home-GEN wall-TOP blue-AI-color-ACC do-PART-be-PRS
'The wall of my house is blue-colored.'

b. uti-no kabe-wa [konna-iro-o] si-te-ir-u.
home-GEN wall-TOP 'like this'-color-ACC do-PART-be-PRS
Lit. 'The wall of my house is 'like-this' colored.'
'The wall of my house has a color like this.'

In (25b), the deictic pro-AP *konna* can appear inside the IO construction, contra the AIC. We take the behavior of the A + C compound with respect to the AIC, as well as *Rendaku*, to be an indication of its phrasal origin. Since the relevant compound shows both word-like and phrase-like properties, we conclude that it is a phrasal compound that is formed post-lexically in the course of syntactic derivation.

Now let us return to the question of at what level the incorporation takes place. If the A + N compound is formed lexically, it would be possible for the subsequent operations to take place either in the lexical component or in the post-lexical component. However, since the A + N compound is formed post-lexically, it immediately follows that the incorporation, too, occurs post-lexically.

4.2. Adjectivization

Let us now turn to the explanation for the second group of properties. Our assumption has been that the A + N compound in the IO construction intrinsically bears an internal theta-role, THEME. After the incorporation, *suru* inherits this internal theta-role. Recall that it is the subject NP of the predicative IO which eventually realizes this THEME theta-role. This means that there still is a step missing in our derivation in (21). We somehow need to assign the internal THEME theta-role to the subject NP. One may postulate that the subject NP is base-generated as an internal argument and undergoes movement. However, this option is not tenable since *suru* clearly assigns accusative case. Burzio's generalization (Burzio, 1986) states that a verb assigns accusative case if and only if it has an external theta-role, and vice versa. As

suru assigns accusative case, it should follow that it also assigns an external theta-role. How is it possible for *suru* to "externalize" the inherited internal THEME theta-role?

We propose that the answer to this question is a morphological operation, adjectivization. Specifically, we claim that, when the category of the IO predicate is changed from verb to adjective, its thematic structure is altered in such a way that its internal theta-role gets externalized. This characterization of "adjectivization" is due to the insight of Levin and Rappaport (1986). They argued that the externalization of the internal theta-role observed in the English adjectival passive formation is a by-product of the adjectivization of passive participles, resulting from the interaction between the properties of passive morpheme and the general property of adjectives such that they obligatorily assign external theta-roles.

The postulation of adjectivization can, at the same time, account for the properties D, E, and F of the IO construction in Section 3.6. That the IO construction can appear as the complement clause of the ECM construction (property D) is no longer puzzling, as the IO predicate is an adjective, in accordance with the restriction posed on the embedded ECM predicate. That *suru* in the IO construction can be modified by adjective/adverb-oriented degree adverbs (property E) is no longer surprising for the same reason. Finally, that *suru* in the IO construction obligatorily takes the participle form (property F) receives a natural account if we hypothesize that the input of adjectivization must be a participle. Since de-verbal adjectives often take the participial form (e.g., English participial adjectives *broken heart, bent stick, surprising fact*), such a proposal seems to be in accord with the facts in other languages.

We have argued earlier that preceding two operations, A + N compounding and incorporation both take place at post-lexical levels. Again, as a logical consequence, adjectivization must be post-lexical, too.

5. Summary and conclusion

In this paper, we have discussed the Japanese Inalienable Object construction. We have analyzed the relevant construction by postulating two post-lexical operations: noun incorporation and adjectivization. The A + N compounds in the IO construction exhibit both word-like and phrase-like properties, suggesting that they are a phrasal compound, post-lexically formed in the course of syntactic derivation. As a logical consequence, the subsequent two operations, incorporation and adjectivization, also take place post-lexically. Since the end-product is an adjective, the IO construction can be said to be a phrasal counterpart of the English Possessional Adjective (e.g., *blue-eyed*).

The analysis which is necessitated by the IO construction has an important implication for the theory of word formation. To the extent that our analysis holds, we have shown evidence against the strong lexicalist hypothesis, which denies the occurrence of word formation outside the lexical component. On the other hand, our analysis is compatible with the modular theory of word formation (Shibatani and Kageyama, 1988; Kageyama, 1993). It is crucial for our analysis of the IO construction that the morphological rules can operate outside the lexical component. One can restate the often-asked question "Does syntax have access to morphology?" as "Does morphology have access to syntax?" Our answer is clearly yes.

References

Baker, M. (1988) *Incorporation: A Theory of Grammatical Function Changing.* University of Chicago Press.

Beard, R. (1995) *Lexeme-Morpheme Base Morphology.* Albany, NY: SUNY Press.

Bresnan, J. and Mchombo, S. A. (1995) The Lexical Integrity Principle: Evidence from Bantu. *Natural Language and Linguistic Theory, 13*, 181-252.
Burzio, L. (1986) *Italian Syntax: A Government-Binding Approach*. Dordrecht: Kluwer academic Publishers.
Chomsky, N. (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, N. (1995) *The Minimalist Program*. Cambridge, Mass.: MIT Press.
Di Sciullo, A. M., and Williams, E. *On the definition of word*. Cambridge, Mass.: MIT Press.
Dubinsky, S. (1989) Compound *suru* verbs and evidence for unaccusativity in Japanese. *CLS 25*, 98-111.
Dubinsky (1994) Syntactic underspecification: A minimalist approach to light verbs. *MIT Working Papers in Linguistics, 24*, 61-81.
Harada, S.-I. (1973) Counter Equi NP Deletion. *Annual Bulletin 7*, Research Institute of Logopedics and Phoniatrics, University of Tokyo, 113-147.
Jensen, J. and Stone-Jensen, M. (1984) Morphology is in the Lexicon! *Linguistic Inquiry, 15*, 474-498.
Kageyama, T. (1977) Incorporation and sino-Japanese verbs. *Papers in Japanese Linguistics 5*, 117-155.
Kageyama, T. (1982) Word Formation in Japanese. *Lingua, 57*, 215-258.
Kageyama, T. (1989) The place of morphology in the grammar: verb-verb compounds in Japanese. *Yearbook of Morphology, 2*, 73-94.
Kageyama, T. (1993) *Bunpoo to Gokeisei*. Tokyo: Hitsuji Shoboo.
Kuno, S. (1973) *The structure of the Japanese language*. Cambridge, Mass.: MIT Press.
Kubozono, H. (1993) *The organization of Japanese Prosody*. Tokyo: Kuroshio Publishers.
Kubozono, H. (1995) *Go-keesee to Onin-koozoo* (Word Formation and Phonological Structure). Tokyo: Kuroshio Publishers.
Levin, B. and Rappaport, M. (1986) The Formation of Adjectival Passives. *Linguistic Inquiry, 17*, 623-662.
Lieber, R. (1992) *Deconstructing Morphology: Word Formation in Syntactic Theory*. Chicago: University of Chicago Press.
Matsumoto, Y. (1991) A syntactic account of light verb phenomena in Japanese. *Journal of East Asian Linguistics, 5*, 107-149.
Miyagawa, S. (1987) Lexical categories in Japanese. *Lingua, 73*, 29-51.
Postal, P. M. (1969) Anaphoric Islands. *CLS, 5*, 205-239.
Saito, M. (1985) *Some asymmetries in Japanese and their theoretical consequences*. PhD dissertation, MIT.
Saito, M. and Hoshi, H. (1994) Japanese light verb construction and the minimalist program. A chapter in the research report of the 1991-1993 project on Japanese grammar at the University of Connecticut, submitted to BBN Systems and Technologies.
Shibatani, M. and Kageyama, T. (1988) Word formation in modular theory of grammar: Post-syntactic compounds in Japanese. *Language, 64*, 451-484.
Teramura, H. (1984) *Nihongo-no sintakkusu to imi Vol 2*. Tokyo: Kurosio-shuppan.
Uchida, Y. and Nakayama, M. (1993) Japanese verbal noun constructions. *Linguistics, 31*, 623-666.
Vance, T. J. (1987) *An Introduction to Japanese Phonology*. Albany, NY: State University of New York Press.

Hitay Yükseker
University of Toronto, Queen's University, York University

TURKISH POSSESSIVE COMPOUNDS *

ABSTRACT

In this paper, I propose an analysis to account for the syntactic and semantic properties of possessive compounds in Turkish. These constructions have lexical heads that are affixed with the third person singular possessive marker. Unlike syntactic possessive constructions, the non-heads of the possessive compounds do not carry genitive case. I propose two functional projections, a Poss(essive) P(hrase) and a Genitive P(hrase). I show that the syntactic properties of the possessive compounds are a consequence of the interaction between these projections, and the semantic properties of these constructions is a reflex of the semantics of "third person".

1. Introduction: In this paper, I examine Turkish possessive compounds which have lexical heads that are affixed with the third person singular possessive marker. These compounds and syntactic possessives share the property of having lexical heads to which the possessive marker is attached. The difference between possessive compounds and syntactic possessives is the presence of the genitive marker on the non-head of the syntactic possessives. Examples of possessive compounds and syntactic possessives are shown in (1) and (2) below. The data in (1a) and (2a) also show that in the absence of the possessive marker the structures are interpreted as attributive-head, i.e. root, compounds. In the examples in (1) below, the head is a non-derived noun, kutu 'box' and in (2) the head is a deverbal nominal, kapak 'lid':

(1) a. Root Compound:
oyuncak kutu
toy box
'a box which is a toy'
*'a box in which toys are stored'

b. Possessive Compound:
oyuncak kutu-su
toy box-poss
'a box in whichtoys are stored'
*'a box which is a toy'

c. Syntactic Possessive:
oyuncağ-ın kutu-su
toy-3sggen box-poss
'the box in which a particular toy is stored'

(2) a. Root Compound:
silindir kapa-k
cylinder close-instr
'a lid shaped as a cylinder'
*'a lid designed to be used with cylinder shaped containers'b.

b. Possessive Compound:
silindir kapa-ğ-ı
cylinder close-instr-poss
'a lid designed to be used with cylinder shaped containers'
*'a lid shaped as a cylinder'

c. Syntactic Possessive:
silindir-in kapa-ğ-ı
cylinder-3sggen close-instr-poss
'the lid of the cylinder shaped container'

The structural difference between the possessive compounds in (1b) and (2b) and the syntactic possessives in (1c) and (2c) have a semantic reflex. The possessive compounds are non-referential and non-specific, i.e. generic and the syntactic possessives are referential and specific.

There are some additional facts that need to be considered. These are shown in (3) and (4). (3) shows an ambiguous string, which can either be an example of the interaction between a syntactic possessive and a possessive compound or an example of the interaction between a syntactic possessive and a root compound. In either interaction, the genitive marked nominal has to precede the generic non-head as in (3a). The opposite ordering in (3b) is unacceptable:

(3) a. Hitay-ın oyuncak kutu-su
-3sggen toy box-poss
(i) 'Hitay's box in which toys are stored'
(ii)'Hitay's toy which is a box'

b. *oyuncak Hitay-ın kutu-su
toy -3sggen box-poss
(i) 'Hitay's box in which toys are stored'
(ii) 'Hitay's toy which is a box'

The data in (4) show the effects of subject pro-drop in syntactic possessives. (4a) shows that a first person singular marked possessive marker allows the subject pro to drop without any syntactic/semantic consequence. In other words, the structure will always be interpreted as a syntactic possessive. On the other hand, (4b) and (4c) show that, if the possessive marker is third person singular, dropping the subject pronoun will yield only the possessive compound interpretation. Genitive marked third person singular pronoun has to be present for the structure to be interpreted as a syntactic possessive:

(4) a. (ben-im) oyuncak kutu-m
(I-1sggen) toy box-1poss
(i) 'my box in which toys are stored'
(ii) 'my toy which is a box'
(iii) *'a box in which toys are stored'

b. oyuncak kutu-su
toy box-poss
(i) *'his/her box in which toys are stored'
(ii) *'his/her toy which is a box'
(iii) 'a box in which toys are stored'

c. on-un oyuncak kutu-su
3sg-gen toy box-poss
(i) 'his/her box in which toys are stored'
(ii) 'his/her toy which is a box'
(iii) *'a box in which toys are stored'

The purpose of this paper is to answer the questions in (5):

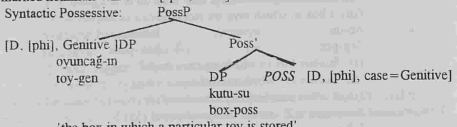
(5) a. What is the structure of syntactic possessives?
b. What is the structure of possessive compounds?
c. How do we account for the relative ordering shown in (3)?
d. What is the cause of the ambiguity in (3a)?
e. What does the pro-drop phenomenon shown in (4) have to say about the structure and derivation of possessive compounds and syntactic possessives?
f. Are possessive compounds syntactic or lexical?

2. Analysis: The analysis is based on the principle of morphology which allows to have both abstract and concrete versions of a given morpheme (Chomsky 1993). The analysis also assumes that the concrete versions of the morphemes attach to their lexical heads in the morphology component of the grammar and words enter syntax fully inflected. The analysis also assumes feature checking as outlined in the Minimalist Program (Chomsky 1995).

2.1. Syntactic Possessives and Possessive Compounds: I propose that in order to understand the properties of possessive compounds, we need to formulate an account of syntactic possessives. The proposed structure of syntactic possessives is shown in (6). The head of the structure is the abstract morpheme, *POSS* , which selects a DP complement. This abstract morpheme is [+N] and it also has a [D] feature, [ ] and case features. The [+N] feature of the abstract head is checked by the concrete morpheme which has the same feature. The lexical possessive morpheme is attached to the lexical head in the morphology component of the grammar. This lexical head undergoes head movement to check the [+N] category feature of the abstract head. The abstract head projects a Poss(essive) P(hrase) whose specifier position is where the [D], [phi] and [Case] features of the head are checked in a [spec-head] relation. The structure is derived by lexical insertion, i.e merge. The spec of PossP is for the genitive case to be checked. Along with Legate and Smallwood (1997), I assume that structural case and morphological case are distinct. Therefore, any nominal with any case can be picked up during enumeration to fill the [spec,PossP] position. But unless the nominal that merges to that position has the

genitive morpheme the cases of the position and the nominal that fills that position will clash, case feature checking will fail, and the derivation will crash. Therefore, only genitive marked nominals will occur [spec,PossP].

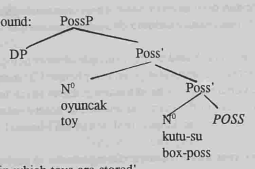
(6) Syntactic Possessive:



'the box in which a particular toy is stored'

Now consider the derivation of possessive compounds. The generic interpretation of the string should follow from the structure of the possessive compound. We might suggest that the structure results from the abstract head POSS selecting either a DP (as in (6) above) or an N^0 as complement, and that another N^0 may head adjoin to Poss' as shown in (7) below:

(7) Possessive compound:

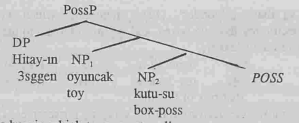


'a box in which toys are stored'

The problem with this proposal is that the structure above is not permissible. Since neither nominal in (7) projects further, according to Bare Phrase Structure Principles (Chomsky 1994) their category must be maximal.

Let us consider another possibility. In possessive compounds the non-head is in a thematic relation with the lexical head. Ediskun (1985) refers to this type of relation as an abstract relation of possession which does not have any reference to reality at the moment of speaking, i.e non-referential and non-specific. From (1c) and (3), we know that the lexical head can also be in a referential and specific possession relation with the non-head. These referential and non-referential possession possibilities can be structurally represented as the abstract *POSS* morpheme projecting two specifiers, one referential and specific possession (DP) and one non-referential and nonspecific possession (NP_1) as shown in (8) below. The structure that will result is the interaction between a syntactic possessive (DP specifier) and a possessive compound (NP specifier), i.e. the representation of (3).

(8)



'Hitay's box in which toys are stored'

One major criticism of the structure in (8) is that the relative positions of the specifiers, though providing a grammatical string (shown in (3)) do not follow from any independent principle. In order to address this issue, I propose to develop a classification of Turkish nominal phrases. Consider the different types of nominal phrases occurring in direct object position, as exemplified in (9). (9a) has a [determiner+N] direct object and this direct object has an overt case marker. In (9b) we have a [number+N] direct object. This direct object does not have an overt case marker. In (9c) we have an [N] direct object. This direct object does not carry an overt case marker either:

(9) a. Hitay- Ø bu kitab-ı oku-du- Ø
Hitay-nom this book-acc read-pa-3sg
'Hitay read this book'

b. Hitay- Ø iki kitap- Ø oku-du- Ø
Hitay-nom two book-acc read-pa-3sg
'Hitay read two books'

c. Hitay- Ø kitap oku-du- Ø
Hitay-nom book read-pa-3sg
'Hitay did book reading'

I propose that the direct object in (9a) is referential and specific; the one in (9b) is referential and non-specific; and the one in (9c) is non-referential and non-specific, i.e. generic. As (10) shows, it is not necessary to have an overt determiner for a nominal phrase to be referential and specific. The semantic information is encoded in the portmanteau case marker:

(10) Hitay- Ø kitab-ı oku-du- Ø
Hitay-nom book-acc read-pa-3sg
'Hitay read the book.'

(11) summarizes the semantic properties of nominal phrases in Turkish:

(11)

DP	NumberP	NP
referential	referential	non-referential
specific	non-specific	non-specific, i.e. generic

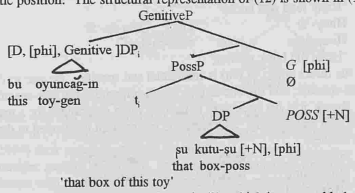
Now let us reconsider the structure of syntactic possessives. (12) shows that both the genitive and possessive markers are affixed to N's of DP's. Turkish has three demonstrative pronouns describing the closeness of the object in question to the speaker:

bu 'this'; şu 'that'; o 'yonder'. (12) is an utterance which would be used in the context when a specific toy has two boxes, and we are referring to the one which is in some distance from the speaker:

(12) bu oyuncağ-ın şu kutu-su
this toy-3sggen that box-poss
'that box of this toy'

To account for the derivation of the syntactic possessives, I propose to separate theta assignment (possession theta role) and the checking features (of case and referentiality & specificity). There are two functional projections: a PossP and a Genitive P(hrase). The possession theta role is assigned to [spec, PossP] and referentiality and specificity and genitive case are checked in [spec, GenitiveP]. That is to say, there is an abstract head *POSS* which bears the feature [+N] and takes a DP complement, and projects a specifier to which the possession theta role is assigned. The PossP is the complement of an abstract head *G*(enitive) which is phonologically null in all instances. The [spec, GenitiveP] is referential and specific ([D]), along with having the features [Genitive Case] and [phi]. The referential and specific DP in [spec, PossP] moves to [spec, GenitiveP] to check the strong referentiality and specificity features. The lexical possessive morpheme undergoes head movement to check the [+N] feature of the abstract *POSS*. The same lexical head, which also carries [phi] features, moves to the abstract *G* head. The [phi] features of [spec GenitiveP] are checked via spec-head relation. [Spec, PossP] is merely a thematic position. The structural representation of (12) is shown in (13):

(13)



'that box of this toy'

Recall the ambiguous structure shown in (3), which is repeated below as (14):

(14) Hitay-ın oyuncak kutu-su
-3sggen toy box-poss
a.'Hitay's box in which toys are stored'
b. 'Hitay's toy which is a box'

For the time being let us put aside the ambiguity issue and address the interaction between syntactic possessive and the possessive compound, which is mainly the ordering of the genitive marked nominal phrase and the generic nominal phrase. The string in (14) has one possessive head, and one referential and specific possessor (Hitay-in 'Hitay's') and one generic possessor (oyuncak 'toy'); and one lexical head kutu 'box'. To account for these facts, I proposed two specifiers for one functional head. Recall that the relative ordering of the constituents in (8) was not motivated. Consider the following examples which show that if the complement of the abstract *POSS* is referential and specific (DP), the specifier of PossP also has to be referential and specific (DP, see (15a&b)). On the other hand, if the complement of abstract *POSS* is non-referential and nonspecific (NP), the specifier can be either non-referential and nonspecific (NP see (15c)) or referential and specific (DP see (15d)) or both (15e):

(15) a. bu oyuncağ-ın şu kutu-su
this toy-3sggen that box-poss
'this toy's that box'

b. *oyuncak su kutu-su
toy that box-poss
'toy that box'

c. oyuncak kutu-su
toy box-poss
'a box in which toys are stored'

d. oyuncag-in kutu-su
toy-3sggen box-poss
'the box in which a particular toy is stored'

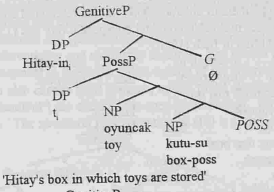
e. Hitay-ın oyuncak kutu-su
-3sggen toy box-poss
(i) 'Hitay's box in which toys are stored'
(ii) 'Hitay's toy which is a box'

The facts shown in (15a) and (15b) cannot be accounted for structurally. The only possible explanation, then, is a semantic one. It seems that a referential and specific possessed can only have a referential and specific possessor, a form of semantic incompatibility (Ghomeshi & Massam 1994; Yükseker 1995). On the other hand a generic possessed can have either a specific and referential possessor or a generic possessor.

Now let us turn to the structural representation of the interaction between a syntactic possessive and a possessive compound. There are two possible derivations. These are

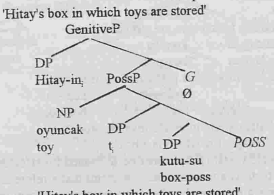
shown in (16). In both structures two specifier positions are created. One for non-referential and nonspecific possession (NP), and one for referential and specific possession (DP). (16a) and (16b) show different orders of merge. In (16a) first an NP merges and then a DP. In (16b) the order is reversed. Either one can represent the derivation of the string in (14). In (16a), as already proposed, DP moves to [spec, GenitiveP] to check referentiality and specificity. In (16b) even if the NP is closer to the specifier position of the GenitiveP, and even if it moves to [spec, GenitiveP], it does not have the appropriate features to check the features of referentiality and specificity of the specifier position, causing the derivation to crash. Therefore, in (16b) given that the DP is the only constituent which has the appropriate features of referentiality and specificity to check the features of the head G , it is the constituent to move to [spec,GenitiveP]. Therefore, the observed ordering facts follow from the principles and the mechanisms of the theory.

(16) a.



'Hitay's box in which toys are stored'

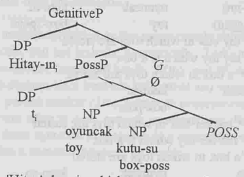
b.



'Hitay's box in which toys are stored'

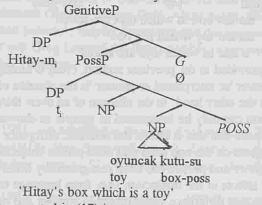
Before I address the last question listed in (5), I account for the ambiguity found in the string (3), which is repeated as (14) above. The representations in (17) below provide the two structures that account for the ambiguous string. In (17a) we see an example of the interaction between a syntactic possessive and a possessive compound. In (17b) we see a syntactic possessive and a root compound interaction.

(17) a.



'Hitay's box in which toys are stored'

b.



'Hitay's box which is a toy'

The ambiguity represented in (17) is an expected consequence of the model proposed in this paper, which also assumes that root compounds are created in the morphology.

To conclude, I have shown that the structural and semantic properties of syntactic possessives and possessive compounds can be accounted for by proposing two functional projections, a PossP and a GenitiveP. I also separated the assignment of the possession thematic role, and checking of the [case], [D] and [phi] features, and assigned each function to the domain of a different projection.

Now I turn to the subject pro-drop properties of syntactic possessives and possessive compounds.

2.2. Pro-drop; syntactic possessives and possessive compounds: Recall the strings in (4), which are repeated below :

(18) a. (ben-im) oyuncak kutu-m
(I-1sggen) toy box-1poss
(i) 'my box in which toys are stored'
(ii) 'my toy which is a box'
(iii) *'a box in which toys are stored'

b. oyuncak kutu-su
toy box-poss
(i) *'his/her box in which toys are stored'
(ii) *'his/her toy which is a box'
(iii) 'a box in which toys are stored'

c. on-un oyuncak kutu-su
3sg-gen toy box-poss
(i) 'his/her box in which toys are stored'
(ii) 'his/her toy which is a box'
(iii) *'a box in which toys are stored'

Given the analysis provided in the previous section, (18a) is ambiguous as we expect it to be. But (18a) can never be interpreted as generic in the absence of an overt pronoun in [spec, GenitiveP]. On the other hand, in the absence of an overt third person genitive pronoun the construction can only be interpreted as generic as shown by the variation in (18b) and (18c). I propose that this is due to the semantic difference between the first and second person subjects on the one hand and third person subjects on the other. This difference between number and person has been recognized by different researchers over the years although the effects of the distinction vary from language to language. For instance, in Athapaskan this difference is expressed in number and person marking for the first and second persons, and only number for the third person (Rice and Saxon 1994). For Labrador Inuttut the choice of indicative vs participial mood is determined by the person agreement markers, where the first and the second persons pattern together, and third person has a different behaviour (Johns 1993). In Turkish this difference is semantic, i.e. referentiality and specificity. The first and the second persons are referential and specific, the third person is generic. This distinction is limited to person morphemes only. The same is not true for pronouns. It appears that the pragmatics of pronouns require them to be referential. Therefore, in the absence of an overt referential and specific genitive 3rd person (singular) subject pronoun which is needed to check the referentiality and specificity features of [spec, GenitiveP], a structure with the third person possessive marker will be interpreted as generic. Words are usually considered to be generic (Di Sciullo and Williams 1987). Therefore, generic third person singular syntactic possessives are interpreted as words. Compounds are words that have complex internal structures. Therefore, syntactic possessives with 3rd. person singular heads without overt possessor subjects are interpreted (possessive) compounds. That is to say, Turkish possessive compounds are not lexically derived, but they are one type of

syntactic possessive which are interpreted as generic due to the lack of a referential and specific subjects. Thus they are syntactic.

3. Conclusion: The analysis provided in this paper clearly shows that Turkish possessive compounds are syntactic, and not lexical. Syntactically, Turkish possessive compounds are projections of two distinct functional heads, *POSS* and *G*. At the same time, semantically, Turkish distinguishes between referential and specific person markers (1st and 2nd) and generic person marker (3rd). Syntactic possessives which lack a referential and specific third person subjects have generic interpretation. The "word, i.e. compound" status of these syntactic possessives is a consequence of the semantics of genericity which is also a semantic property of words.

*I would like to thank the members of University of Toronto, Department of Linguistics, Syntax Project (Elizabeth Cowper, Elaine Gold, Alana Johns, Päivi Koskinen, Niina Ning Zhang) for their patience, helpful comments and discussions. I would also like to thank Barry Miller and Greg Lessard for their valuable suggestions. Also my thanks to my informant, Erkul Yükseker.

References:

Chomsky, Noam. 1993. 'A Minimalist Program for Linguistic Theory'. In K. Hale and S. J. Keyser (eds)*The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* ,. pp. 1-52. Cambridge, Mass. The MIT Press.

Chomsky, Noam. 1994. 'Bare Phrase Structure'. *MIT Occational Papers in Linguistics 5*. Department of Linguistics and Philosophy, MIT.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass. The MIT Press.

Di Sciullio Anna-Marie, & Edwin Williams. 1987. *On the Definition of Word* Linguistic Inquiry Monograph #14. Cambridge, Mass. The MIT Press.

Ediskun, Haydar. 1985. *Türk Dilbilgisi* . Remzi Kitabevi, Istanbul.

Ghomeshi, Jila and Diane Massam. 1994. 'Lexical/Syntactic Relations without Projection'. *Linguistic Analysis*. vol. 24. pp 175-217.

Johns, Alana. 1993. 'Symmetry in Labrador Inuttut'. *Papers on Case and Agreement. MIT Working Papers in Linguistics*. vol. 18. pp. 43-57.

Legate, Julie Anne and Carolyn Smallwood. 1997. 'The Case Filter Meets the Minimalist Program: Evidence for Strong Case Features'. Paper presented at OLF 1997, University of Ottawa, March 1997.

Rice, Keren and Leslie Saxon. 1994. 'The Subject Position in Athapaskan Languages'. In H. Harley and C. Phillips (eds) *The Morphology-Syntax Connection*. MIT Working Papers in Linguistics, volume 22,173-196.

Yükseker, Hitay. 1995. 'How Many -ler/lar's are there in Turkish?'. In Päivi Koskinen, (ed) *Proceedings of the 1995 Annual Conference of Canadian Linguistics Association. Toronto Working Papers in Linguistics*, University of Toronto, pp. 639-650.

Hitay Yükseker
hy2@qsilver.queensu.ca
yukseker@yorku.ca

Section IV
Inflection

Bozena Cetnarowska
Institute of English
University of Silesia, Katowice
e-mail: cetnarow@us.edu.pl

INHERENT INFLECTION IN POLISH AS INPUT TO MORPHOLOGICAL DERIVATION

The present paper examines the phenomenon of inflection feeding derivation in modern Polish. Although markers of contextual inflection in Polish do not appear word-internally, markers of inherent inflection can be followed by derivational suffixes, especially by highly productive ones. Only a small section of lexemes containing word-internal inflectional morphemes in Polish can be analyzed as derived from lexicalized (irregular) word forms. Inflectional morphemes occurring „inside" derivational morphemes in Polish include stem-forming inflectional morphemes found in comparative or superlative degree forms of adjectives, in past participles, present (active) participles and passive participles. Polish participles are also shown to undergo conversion into adjectives and adverbs.

1. Inflectional categories in Polish

Recent coursebooks for students of Polish morphology (e.g. Laskowski 1984, Strutynski 1996) list the following morphological categories as being inflectional in Polish: case, number, gender, tense, person, mood, voice, aspect[1] and degree. A distinction is drawn between inflectional endings (markers of agreement) and stem-forming inflectional morphemes. The latter are employed to build extended (secondary, derived) stems of lexemes, such as for instance the comparative and superlative degree stems of adjectives. Conjugational paradigms of verbs include, apart from the basic verb stem, four extended stems, namely the imperative verb stem, the past tense stem and two participial stems. Stem-forming inflectional morphemes in Polish can be regarded as markers of inherent inflection (in the sense of the term proposed in Booij 1996). As will be illustrated below, inherent inflection can feed derivation in modern Polish.

2. Markers of comparative and superlative degree occurring word-internally

Derivatives considered in this section exhibit formal and semantic affinity to comparative or superlative forms of adjectives. The morpheme *-sz-/-ejsz-* is the regular comparative marker while the prefix *naj-* marks the superlative degree of adjectives. The data in (1-2) demonstrate that irregular inflectional forms (which exhibit stem allomorphy or suppletion) can function as input to derivation since they are entered in the lexicon[2].

[1] The status of the category of aspect is controversial. In contrast to Laskowski (1984), Grzegorczykowa (1997) regards aspect as a lexical category. Consequently, I shall refrain from discussing lexemes which are derived from secondary (derived) imperfectives.

(1) deadjectival prefixed verbs
- a. *podwyzszyc* 'to heighten' (cf. *wyzszy* 'higher' from *wysoki* 'high, tall')
- b. *pogorszyc* 'to worsen' (cf. *gorszy* 'worse' from *zly* 'bad')
- c. *przewyzszyc* 'to surpass' (cf. *wyzszy* 'higher' from *wysoki*)
- d. *ulepszyc* 'to better, to improve' (cf. *lepszy* 'better' from *dobry* 'good')
- e. *zmniejszyc* 'to lessen' (cf. *mniejszy* 'smaller' from *maly* 'small, little')
- f. *zwiekszyc* 'to increase' (cf. *wiekszy* 'bigger, larger' from *wielki* 'big, large')

(2) *-osc* nouns (names of qualities - Nomina Essendi)
- a. *blizszosc* 'the quality of being closer in time or space' (cf. *blizszy* 'closer' from *bliski* 'close')
- b. *lepszosc* 'the quality of being better' (cf. *lepszy* 'better' from *dobry* 'good')
- c. *mniejszosc* 'minority' (cf. *mniejszy* 'smaller, minor' in 1e)
- d. *najlepszosc* 'the quality of being the best' (cf. *najlepszy* 'best' from *lepszy* 'better')
- e. *najwyzszosc* 'the quality of being the highest in rank or size' (cf. *najwyzszy* 'highest' from *wyzszy* 'higher' in 1c)
- f. *nizszosc* 'the quality of being lower' (cf. *nizszy* 'lower' from *niski* 'low')
- g. *wiekszosc* 'majority' (cf. *wiekszy* 'larger, major' in 1f)
- h. *wyzszosc* 'superiority' (cf. *wyzszy* 'higher, superior' in 1c)

The word-form *starszy* 'older, elderly, senior' (the comparative degree of *stary* 'old'), which gives rise to the derivatives given in (3), should also be regarded as lexicalized because it exhibits a semantic shift.

(3) miscellaneous nouns and adjectives derived from *starszy* 'older, senior'
- a. *starszawy* 'elderly'
- b. *starszak* 'older child in kindergarten'
- c. *starszenstwo* 'seniority'
- d. *starszosc* 'the quality of being older'
- e. *starszyzna* 'the seniors, the officers of high rank'

The data in (4) show that highly productive derivational suffixes, e.g. *-osc,* can attach to regular comparative forms and produce names of qualities (NE).

[2] All the examples from Polish given below occur in a slightly simplified spelling since Polish diacritic marks indicating palatalization of consonants and nasalization of vowels are not available in the international version of the word processor employed here. In accordance with standard Polish orthography, I use the digraph 'ch' to represent a velar fricative. The digraphs 'cz', 'dz' stand for post-alveolar affricates while 'sz' is a post-alveolar fricative.

(4) *-osc* Nomina Essendi (NE)
- a. *bielszosc* 'the quality of being whiter' (cf. *bielszy* 'whiter')
- b. *jasniejszosc* 'the quality of being fairer or clearer" (cf. *jasniejszy* 'fairer, clearer')
- c. *mlodszosc* 'the quality of being younger' (cf. *mlodszy* 'younger')
- d. *piekniejszosc* 'the quality of being more beautiful' (cf. *piekniejszy* 'more beautiful')
- e. *wczesniejszosc* 'the quality of being earlier' (cf. *wczesniejszy* 'earlier')
- f. *weselszosc* 'the quality of being more cheerful' (cf. *weselszy* 'more cheerful')
- g. *wrazliwszosc* 'the quality of being more sensitive' (cf. *wrazliwszy* 'more sensitive')
- h. *zywszosc* 'the quality of being more agile' (cf. *zywszy* 'more agile')

Some of the names of qualities in (4) are felt to be non-institutionalized[3] (e.g. 4e) and the list of NE can be easily extended with novel formations.

3. Derivatives from past tense forms (or past tense stems)

The lexemes discussed in this section appear to contain the past tense marker *-l-*. The nouns in (5) belong to the realm of expressive morphology. Creation of nonce-formations exemplifying the patterns in (5) is highly probable, especially when such nouns denote people.

(5) *-ak, -uch, -ek, -ec* nouns
- a. *opuchlak* 'swelling' (*opuchl* '(he/it) became swollen' and *opuchly* 'swollen')
- b. *spróchnialec* 'one that has grown rotten or carious' (cf. *spróchnial* '(it) decayed, grew carious' and *spróchnialy* 'rotten, carious')
- c. *umarlak* (colloq.) 'dead fellow' (cf. *umarl* '(he) died' and *umarly* 'dead')
- d. *zdechlak* (colloq.) 'weakling' (cf. *zdechl* '(it) died' and *zdechly* '(of animals) dead')
- e. *zmarzlak* (colloq.) 'chilly fellow' (cf. *zmarzl* '(he) felt chilly' and *zmarzly* 'chilly')
- f. *zmarzluch* (colloq.) 'chilly person' (cf. 5a above)
- g. *zgnilek* (colloq.) 'rascal' (cf. *zgnil* '(he/it) decayed, went rotten, putrid ' and *zgnily* 'rotten, decayed, putrid')
- h. *zgnilec* '(of bees) foul brood')
- i. *zgnilki* (pl.) 'rotten fruit'
- j. *zgorzknialec* 'sour person' (cf. *zgorzknial* '(he/it) went sour' and *zgorzknialy* 'sour, bitter')
- k. *zniewiescialec* 'one that has grown effeminate' (cf. *zniewiescial* '(he) grew effeminate' and *zniewiescialy* 'effeminate')

[3] I follow Bauer (1983) in distinguishing between nonce-formations, non-institutionalized lexemes and lexicalized formations.

The nouns in (6) below are best analyzed as lexicalized (as there are few derivatives exemplifying such patterns and some of them are obsolete). The formations in (7), in contrast, represent productive patterns of derivation and many of them may be felt to be nonce-formations.

(6) *-izna, -ina* nouns
- a. *odumarlina* (obsolete) 'possessions of a man who died leaving no successors' (cf. *odumarl* (rare) '(he) died and left someone behind' and *odumarly* (rare) 'left behind at someone's death')
- b. *opuchlizna, opuchlina* 'swelling' (cf. *opuchl* in 5a)
- c. *padlina* 'carcass, the body of a dead animal' (cf. *padl* '(he/it) fell, died' and *padly* 'fallen, dead')
- d. *stechlizna* 'fustiness, mustiness' (cf. *stechl* '(it) grew fusty or musty' and *stechly* 'fusty, musty, frowsty')
- e. *zgnilizna* 'rot, putridity' (cf. *zgnil* in 5g)
- f. *zmarzlina* 'frozen layer of earth in the Far North that unfreezes during a short summer' (cf. *zmarzl* in 5e)

(7) *-osc* derivatives (Nomina Essendi) and *-awy* attenuative adjectives
- a. *dojrzalosc* 'ripeness, maturity' (cf. *dojrzal* '(he/it) grew ripe or mature' and *dojrzaly* 'ripe, mature')
- b. *sflaczalosc, sflaczenie* 'flabbiness, limpness' (cf. *sflaczal* '(he/it) became flabby or limb' and *sflaczaly* 'falbby, limp. flaccid')
- c. *spróchnialosc* 'rot, decay' (cf. *sprochnial* in 5b)
- d. *stezalosc* '(of solutions) solidification, concentration' (cf. *stezal* '(it) concentrated, coagulated' and *stezaly* 'concentrated, solidified')
- e. *zdziczalosc* 'savageness' (cf. *zdziczal* '(he) grew wild, savage' and *zdziczaly* 'savage')
- f. *zgnilosc* 'rot, putridness' (cf. *zgnil* in 5g)
- g. *zgorzknialosc* 'bitterness, acrimony' (cf. *zgorzknial* in 5j)
- h. *zmurszalosc* 'mustiness' (cf. *zmurszal* '(it) mouldered' and *zmurszaly* 'mouldy, musty')
- i. *zniewiescialosc* 'effeminacy' (cf. *zniewiescial* in 5k)
- j. *zwiedlosc* 'fadedness, withered state' (cf. *zwiedl* '(it) withered'and *zwiedly* 'faded, withered')
- k. *sprochnialawy* 'somewhat rotten, decayed' (cf. *sprochnial* in 5b)
- l. *stechlawy* 'somewhat fusty' (cf. *stechl* in 6d)
- m. *zgnilawy* 'somewhat rotten' (cf. *zgnil* in 5g)

The formations listed in (5-7) above cannot be derived directly from past tense verb forms since the suffixes *-ak, -uch, -ek, -ec, -osc, -izna, -ina* and *-awy* are subcategorized to attach to adjectival, not to verbal, bases. Consequently, it is the adjectives terminating in the sequence *-ly* which must serve as the immediate input to the suffixation processes in question. However, I shall argue below that the data in (5-7) instantiate the

phenomenon of inflection feeding derivation since the resultative *-l* - adjectives in question are derived from past tense stems through conversion (adjectivization).

An argument supporting supporting such an analysis is the formal identity of past tense stems and corresponding resultative adjectives. When past tense forms are morphologically irregular, resultative adjectives exhibit the same kind of allomorphy. This happens in the case of the verb *opuchnac* 'to swell, inf.' - *opuchnie* '(it) swell-FUTURE' - *opuchl* '(it) swelled' - *opuchly* 'swollen' (note the lack of the thematic suffix *-na-* in past tense forms and in the resultative adjective) or the verb *umrzec* 'to die, inf.' - *umiera* '(it) dies' - *umarl* '(he/it) died' - *umarly* 'dead' (observe that the quality of the root vowel is the same in the past tense form and the resultative adjective).

It can be argued, moreover, that resultative *-ly* adjectives show semantic affinity to past tense forms and exhibit the internal syntax of finite verb forms.

Firstly, they convey the idea of the past, which is emphasized by the presence of time expressions such as *wczoraj* 'yesterday', *przed chwila* 'a moment ago' or *jesienia* 'in autumn' in the phrases *rozkwitle wczoraj kwiaty* 'the flowers that opened yesterday', *przybyly przed chwila czlowiek* 'the man that arrived a moment ago', and *opadle jesienia liscie* 'leaves that fell down in autumn'. This is not possible with other deverbal adjectives, as shown in **interesujacy wczoraj wyklad* 'interesting yesterday lecture' or **ozywcze przed chwila powietrze* 'refreshing a moment ago (breath of) air'.

Secondly, resultative adjectives resemble finite verb forms in allowing modifiers denoting reason (cause), e.g. *opuchle z glodu nogi* 'legs swollen from hunger', *pozólkle ze starosci listy* 'letters turned yellow with age'.

Thirdly, the negative particle *nie-* does not attach to *-ly* adjectives (as happens in the case of other adjectives, e.g. *niemily* 'not nice' or *nieuprzejmy* 'impolite') but is spelled as a separate word, e.g. *nie sprochnialy* 'not rotten' and *nie rozkwitly* 'that has not blossomed yet'. Consequently, the rules of Polish orthography interpret *-ly* adjectives as verb forms. Tokarski (1951) analyzes such adjectives as respresenting the category of 'participium staticum'.

Finally, there is some cross-linguistic evidence suggesting that Polish resultative adjectives derive from past tense stems (which are analyzed as past participles in Tokarski 1951). It is only unaccusative (ergative) verbs in Polish which have related resultative *-ly* adjectives. Hoekstra (1984) postulates the participle-adjective conversion as one of the tests employed to distinguish ergative and unergative verbs. Ergative verbs in Dutch allow their past participles to be used as adjectives. The same phenomenon can be observed in English, where the past participles *fallen* and *withered* (but not *come* or *swum*) can modify nouns. Markantonatou (1995) demonstrates that also in Modern Greek past participles of ergative verbs convert to adjectives.

4. Suffixal derivatives from passive participles

This section of the paper deals with derivatives of passive participles in Polish, i.e. with forms containing word-internally the inflectional affix *-an-* or its phonologically conditioned allomorphs *-on-* and *-t-*.

The nouns listed in (8-9) exhibit vivid semantic (and formal) affinity to passive participles since they denote personal Patients and, less commonly, inanimate affected ob-

jects. Their semantic interpretation contains an element of passivity. The departicipial formations in (8-9) terminate in productive nominalizing affixes, although novel formations of this type cannot be formed as freely as Nomina Essendi in *-osc* .

(8)*-ec, -nik, -ek, -ka* names of Patients
- a. *obrzezaniec* 'circumcised man' (from *obrzezany* 'circumcised')
- b. *opetaniec* 'one possessed of evil' (from *opetany* 'possessed of evil')
- c. *pomazaniec* 'annointed man' (from *pomazany* 'annointed, smeared')
- d. *przesiedleniec* 'emigrant, displaced person' (from *przesiedlony* 'displaced, rehoused')
- e. *skazaniec* 'man condemned to death' (from *skazany* 'condemned')
- f. *wybraniec* 'the privileged man' (from *wybrany* 'chosen')
- g. *wybranka* 'the girl of one's choice' (from *wybrany* 'chosen')
- h. *wygnaniec* 'exile, outcast' (from *wygnany* 'expelled, banished')
- i. *wyzwoleniec* 'freedman' (from *wyzwolony* 'freed, liberated')
- j. *wychowanek* 'alumnus', *wychowanka* (fem.) (from *wychowany* 'brought up')

(9) *-ec, -ek, -ka* object nouns
- a. *nabytek* 'acquisition' (from *nabyty* 'acquired')
- b. *odbitka* (phot.) 'proof' (from *odbity* 'reflected')
- c. *roztrzepaniec* 'sour milk that has been beaten up' (from *roztrzepany* 'beaten up')
- d. *wyjatek* 'exception, excerpt' (from *wyjety* 'extracted, removed')
- e. *wziatka* (colloq.) 'trick (in cards)' (from *wziety* 'taken')

Formations listed in (10) below exemplify a very productive pattern of coining names of objects, typical of specialized vocabulary.

(10) *-ka* derivatives
- a. *bitka* 'cutlet' (cf. *bity* 'beaten, crushed')
- b. *dlubanka* 'canoe' (cf. *dlubany* 'hollowed out')
- c. *kiszonka* 'silage' (cf. *kiszony* 'fermented, pickled')
- d. *krajanka* 'a kind of cake, pasta or cheese (sold sliced or cut into portions)' (cf. *krajany* 'cut, sliced')
- e. *kraszanka* '(dial.) Easter egg' (cf. *kraszany* (dial.) 'painted')
- f. *lepianka* 'mud-built cabin' (cf. *lepiony* 'moulded, built out of sth')
- g. *mieszanka* 'mixture' (cf. *mieszany* 'mixed')
- h. *malowanka* 'painted figure, folkloric ornament' (cf. *malowany* 'painted')
- i *mrozonka* 'chilled fruit, vegetables or other food products (cf. *mrozony* 'frozen, chilled')
- j. *palonka* '(in ceramics) grog' (cf. *palony* 'burnt')
- k. *prazonka* 'roasted ore' (cf. *prazony* 'roasted')
- l. *warzonka* 'table salt' (cf. *warzony* 'vaporized (in order to get rid of dross)')
- m. *wedzonka* 'smoked bacon' (cf. *wedzony* 'smoked')

n. *wycinanka* 'decorative paper cut-out adorning walls of peasant cottages' (cf. *wycinany* 'cut out')
o. *zapiekanka* 'dish baked in oven' (cf. *zapiekany* 'baked')

There is a notable difference between the nouns in (8-9) and those in (10). The formations in (8) are formally and semantically related to passsive participles derived from perfective (prefixed) verbs. The nouns in (10), although exhibiting a resultative (telic) interpretation, are formally related to passives of imperfective verbs. I would like to suggest that the names of objects in (10) are derived from adjectival passives. Adjectival passives are, in turn, derived though conversion from verbal passives[4] . Departicipial passive *any/-ony/-ty* adjectives given in (10) in brackets can be characterized as stative, resultative and perfective, no matter whether they are derived from imperfective or from perfective verb forms.

5. Conversion of participles into nouns or adverbs

Both passive participles and present (active) participles in Polish can undergo conversion into nouns (i.e. substantivization), as shown in (11-12) below. Active (present) participles can be formed from imperfective verb stems only and they contain the inflectional morpheme *-ac-*. The formations given in (11) are semantically lexicalized.

(11) a. *chowany* 'hide and seek' (cf. *chowany* 'hidden-impf.')
b. *dane* (pl.) 'data' (cf. *dany* 'given-pf.')
c. *oskarzony* 'accused' (cf. *oskarzony* 'accused-pf.')
d. *przewodniczacy* 'chairman' (cf. *przewodniczacy* 'presiding (over sth)')
e. *sluzacy* 'servant' (cf. *sluzacy* 'serving')
f. *smazone* 'fried food' (cf. *smazony* 'fried - impf.')
g. *wygrana* 'winnings' (cf. *wygrany* 'won-pf.')

Substantivized passive and active (present) participles in (12) can be formed in a fairly productive manner and they have exclusively human reference. They preserve the internal syntax of verbs and can occur with complements and adjuncts characteristic of finite verb form and verbal participles.

(12) a. *pokrzywdzony przez los* 'one who suffered a bad fate' (cf. *pokrzywdzony* 'wronged')
b. *powracajacy z zagranicy* 'one who is returning (or has returned) from abroad' (cf. *powracajacy* 'returning')
c. *uczacy jezyka angielskiego* 'one who teaches English' (cf. *uczacy* 'teaching')
d. *umierajacy z glodu i pragnienia* 'one who is dying out of starvation and thirst' (cf. *umierajacy* 'dying')

[4] I advance this hypothesis also in Pasternak-Cetnarowska (1986). When discussing English verbal passive participles, Strauss (1987) makes a similar assumption that they undergo conversion into adjectives.

e. *zasypany przez lawine* 'one who has been buried by the avalanche' (cf. *zasypany* 'covered up')

The data in (13) appear to indicate that verbal present participles undergo conversion into adverbs. However, while verbal active *-acy* participles denote a process in progress, e.g. *rozjasniajaca teraz mrok pokoju lampa* 'a lamp which is now lightening the darkness of the room', the forms terminating in *-acy* which give rise to the adverbs in (13) denote a property of an object, e.g. *proszek wybielajacy* 'the whitening powder', as is characteristic of adjectives.

(13) a. *chlodzaco* 'with the cooling effect' (cf. *chlodzacy* 'cooling')
b. *kruszaco* '(about explosives) breaking up the rocks' (cf. *kruszacy* 'breaking up rocks')
c. *lagodzaco* 'soothingly' (cf. *lagodzacy* 'palliative, soothing')
d. *odurzajaco* 'dizzingly' (cf. *odurzajacy* 'dizzing, stupefying')
e. *rozgrzewajaco* 'with the warming up effect' (cf. *rozgrzewajacy* 'warming up')
f. *rozjasniajaco* 'with the brightening effect' (cf. *rozjasniajacy* 'brightening, bleaching')
g. *wybielajaco* 'producing the bleaching effect' (cf. *wybielajacy* 'bleaching')
h. *wzmacniajaco* 'with the strengthening effect' (cf. *wzmacniajacy* 'strenghtening')

It seems justified to assume that the adverbs in (13) are related to departicipial *-acy* adjectives. Such adjectives are derived from verbal present participles through conversion (adjectivization) and therefore show closer semantic/syntactic relatedness to verbal participles than deverbal *-acy* adjectives do (such as *znaczacy* 'telling' cognate to the verb *znaczyc* 'to mean, to mark').

6. Conversion as syntactic word-class exchange or as semantic extension

There is no agreement within the linguistic literature as to the proper treatment of conversion. The phenomenon in question can be treated as purely syntactic word-class exchange or as a derivational process. This is partly due to the variety of processes subsumed under the label of conversion. Marchand (1968), for instance, treats noun-to-verb or verb-to-noun conversion in English as an instance of morphological derivation (i.e. zero-derivation) while the cases of partial conversion, such as the occurrence of the noun phrases *the wealthy* and *the poor,* are analyzed as involving functional transposition of words (i.e. a temporary change of their syntactic function).[5] Within the framework of Cognitive Linguistics, Twardzisz (1996) analyzes zero-derivation in English as

[5] In a similar vein, Strauss (1987) regards the English noun phrases *the oldest* and *the very poor* as instantiating syntactic conversion of adjectives into nouns while the nouns *the untouchables* and *sexists* are examples of morphological conversion. A critical overview of theoretical approaches to conversion is offered in, among others, Cetnarowska (1993), Don (1993), and Twardzisz (1997).

a process of semantic extension and hence being of of essentially the same nature as metonymy and metaphor.

Various types of conversion have been proposed in this paper to occur in Polish.

Derivation of resultative adjectives (such as *zwiedly* 'faded, withered' discussed in section 3 above) from past tense stems can be regarded as a non-affixal morphological process, involving a change of inflectional paradigm, syntactic word-class and semantic properties of the derivational base.

Conversion of Polish verbal passive participles into departicipial adjectives (proposed in section 4) constitutes a special case of semantic extension. It results in a modification of the meaning of participles but does not seriously affect their syntactic properties and involves no change in their inflectional paradigm. Both adjectival resultative passives and verbal passives agree in case, number and gender with their heads. They cannot be inflected for degree , they form no Nomina Essendi and cannot take the negative prefix *nie-*. Moreover, departicipial resultative adjectives and verbal passives can occur with agentive adjuncts, complements and adverbials of manner, as in the sentence *Przez cale lato pola byly calkowicie zalane woda* 'All the summer the fields were flooded with water completely', *Przywiozlam ogórki kiszone w zeszlym roku przez moja mame* 'I've brought cucumbers pickled by my mother last year.' Adjectival passives which function as derivational bases for the nouns in (10) differ from corresponding verbal passives in their inability to follow the copula verb *zostac* 'to become' and ability to take the copular verb *byc* 'to be'.

Similarly, present (active) *-acy* participles and departicipial adjectives share their inflectional paradigms and the choice of complements but differ mainly in their semantic interpretation.

Formation of adverbs terminating in *-aco* (listed in 13 above) involves mainly a change of syntactic category and inflectional paradigm and can be treated as functional shift. I assume, however, that just as the remaining types of conversion mentioned earlier, adverbialization is a morphological process.

Substantivization of present and passive participles illustrated in (12) in section 5 instantiates a temporary use of verbal participles in the function of nouns (or noun phrases). It exhibits similarity to partial conversion in English and can be regarded as a syntactic (not a morphological) process.

7. Inflection/derivation continuum. Class-changing inflection

It is notable that none of the derivatives discussed in the present paper contain inflectional endings (i.e. markers of agreement) embedded word-internally and preceding derivational affixes. In other words, contextual inflection cannot feed morphological derivation in Polish while some types of inherent inflection can.

In view of Dressler's (1989) criteria for distinguishing prototypical from non-prototypical inflection, most types of inherent inflection in Polish do not constitute prototypical instances of inflection. This is the reason why the inflectional status of passive participles or comparative degree forms is still a controversial issue among Polish morphologists.

The idea that inflectional processes cannot be separated from derivational processes in a non-arbitrary fashion was expressed by Adam Heinz , who shared Jerzy Kurylowicz's belief in essentially homogenous nature of all morphological operations. Heinz (1961) regards semantic regularity, predictability of existence ('categorial-ness') and the capability of forming a closed set of items (a paradigm) as indices of grammaticality of morphological categories. Inflectional and derivational categories differ in their degree of grammaticality, hence there is a gradual transition between them. According to Heinz (1961), comparative and superlative degree markers and passive participles belong to the transitional zone between inflection proper and derivation proper.

There is an interesting overlap between the ideas expressed in Heinz (1961) and the more recent proposals of Haspelmath (1996). Haspelmath (1996) points out that the inflection/derivation distinction allows for gradience and fuzzy borders. He says (p. 47): „formations are inflectional to the extent that they are regular, general and productive; formations are derivational to the extent that they are irregular, defective and unproductive". Consequently, he treats German adjectival participles and Lezgian masdars (i.e. verbal nouns) as inflectional forms due to their productivity and regularity. These are instances of word-class-changing inflection. He also mentions, among others, the possibility of treating English *-ly* adverbs as inflectional forms of adjectives and Upper-Sorbian possessive adjectives as inflectional forms of nouns, i.e. as cases of transpositional (class-changing) inflection

Within the Polish tradition of morphology, the recognition of word-class-changing inflection has been proposed in Tokarski (1973), where action nouns in *-nie/-cie* are included into the verb paradigms as gerunds while adverbs are regarded as inflectional forms of corresponding adjectives. It is suggested in Tokarski (1973) that deadjectival *-osc* nouns, such as *pieknosc* 'beauty', *taniosc* 'cheapness', are to be regarded as nominal forms of adjectives, i.e. as additional cases in declensional paradigms of adjectives. Tokarski's inclination toward expanding he domain of inflection is driven mainly by his desire to facilitate the task of the lexicographer. Tokarski assumes that derived words should appear in the lexicon as independent entries while the burden of predicting the occurrence of inflectional forms should rest entirely with the grammar.

It is worth noting that the majority of suffixes appearing "outside" markers of inherent inflection in Polish lexemes (such as *-osc, -awy, -ak, -ka*) can be incorporated into the "transitional zone" between prototypical inflection and prototypical derivation since they exhibit relatively high productivity. These affixes can also be reanalyzed as inflectional morphemes, if one allows for the occurrence of word-class changing inflection[6].

[6] A potential problem with the postulation of word-class changing inflection is that it results in blurring the distinction between properties of word-forms and of derived lexemes. Firstly, derivatives undergo semantic drift when institutionalized (e.g. *-osc* abstract nouns develop concrete senses) while word-forms are not expected to do so. Secondly, while instances of parallel word-forms are rare (e.g. the alternative forms *dyrektorzy, dyrektorowie* 'directors, nom.pl'), examples of parallel derivation are numerous in Polish, as shown in the analysis of Polish diminutives in Malicka-Kleparska (1985).

8. Conclusion

The Polish data analyzed in this paper indicate that inherent and non-prototypical (i.e. stem-forming) inflection can feed non-prototypical (i.e. highly productive) suffixal derivation. Moreover, stem-forming verbal inflection can constitute an input to non-affixal (hence also non-prototypical) morphological processes, namely adjectivization and adverbialization of verbal participles.

REFERENCES

Bauer, Laurie (1983) *English Word-Formation*, CUP, Cambridge.

Booij, Geert (1996) "Inherent versus contextual inflection and the split morphology hypothesis", *Yearbook of Morphology 1995*, 1-16.

Cetnarowska, Bozena (1993) *The Syntax, Semantics and Derivation of Bare Nominalisations in English*, Wydawnictwo Uniwersytetu Slaskiego, Katowice.

Don, Jan (1993) *Morphological Conversion*, doctoral dissertation, OTS, Utrecht University.

Dressler, Wolfgang U. (1989) „Prototypical differences between inflection and derivation", *Zeitschrift fur Phonetik, Sprachwissenschaft und Kommunikationsforschung* 42, 3-10.

Grzegorczykowa, Renata (1997) „Nowe spojrzenie na kategorie aspektu w perspektywie semantyki kognitywnej" [A new approach to verbal aspect in terms of cognitive semantics], in Grzegorczykowa, Renata and Zofia Zaron (eds.) *Semantyczna struktura slownictwa i wypowiedzi* [The semantic structure of vocabulary and of the utterance], Wydawnictwa Uniwersytetu Warszawskiego, Warszawa, 25-38.

Haspelmath, Martin (1996) "Word-class changing inflection and morphological theory", *Yearbook of Morphology 1995*, 43-66.

Heinz, Adam (1961) "Fleksja a derywacja" [Inflection and derivation], *Jezyk Polski* 41, 343-354.

Hoekstra, Teun (1984) *Transitivity: Grammatical Relations in Government Binding Theory*, Foris, Dordrecht.

Laskowski, Roman (1984) „Podstawowe pojecia morfologii" [Basic morphological concepts], in Grzegorczykowa, Renata, Laskowski, Roman and Henryk Wróbel (eds.) *Gramatyka wspolczesnego jezyka polskiego. Morfologia* [A grammar of present-day Polish. Morphology], PWN, Warszawa, 9-57.

Malicka-Kleparska, Anna (1985) „Parallel derivation and lexicalist morphology: the case of Polish diminutivization", in Gussmann, Edmund (ed.) *Phono-morphology. Studies in the Interaction of Phonology and Morphology*, Redakcja Wydawnictw KUL, Lublin, 95-112.

Marchand, Hans (1969) *The Categories and Types of Present-Day English Word-Formation,* C.H. Beck, Munchen.

Markantonatou, Stella (1995) „Modern Greek deverbal nominals: an LMT approach", *Journal of Linguistics* 31, 267-299.

Pasternak-Cetnarowska, Bozena (1986) „Derivatives from adjectival participles in Polish", *Polonica XII*, 173-188.

Strauss, Steven L. (1987) „Word-formation with lexical insertion and filters", in Gussmann, Edmund (ed.) *Rules and the Lexicon. Studies in Word-Formation*, Redakcja Wydawnictw KUL, Lublin, 137-168.

Strutynski, Janusz (1996) *Gramatyka polska. Wprowadzenie. Fonetyka. Fonologia. Morfologia.* [Polish grammar. Introduction. Phonetics. Phonology. Morphology], Ksiegarnia Akademicka, Kraków.

Tokarski, Jan (1951) *Czasowniki polskie* [Polish verbs], PWN, Warszawa..

Tokarski , Jan (1973) *Fleksja polska* [Polish inflection], PWN, Warszawa.

Twardzisz, Piotr (1997) *Zero Derivation in English. A Cognitive Grammar Approach,* Wydawnictwo UMCS, Lublin.

Greville G. Corbett
g.corbett@surrey.ac.uk

HYPOTHESES ON THE STATUS OF NUMBER[1]

Abstract
Number is often cited as an obvious example of an inflectional category. Yet there are considerable problems. Nine hypotheses about number are considered in turn, and all but one are shown to be false. For instance, not all languages have number; in those that do it is not necessarily inflectional. When we consider the distinction between inherent and contextual inflection we see that the number values for these two may not match (evidence from Miya). Thus rather than being a textbook example of inflection, number proves to be a specially interesting category.

Introduction

A category often held to be prototypically inflectional, namely number, proves less uniform in its status cross-linguistically than was once thought (Booij 1993, 1996; van Marle 1996).[2] We shall consider nine relevant hypotheses in turn, disproving most of them, and thereby showing that number is more complex than is generally recognised.

Number as an inflectional category

In discussions of inflectional morphology, the category chosen for illustration tends to be number. It is used by Bloomfield (1933: 222-224), Stump (1990: 98) and Matthews (1991: 53), to name just three. And it is, after all, one of the phi-features. But what does it mean when we talk of a particular category, in this case number, as inflectional? We approach this question by investigating a list of reasonable hypotheses, the sort of hypotheses that writers may have had in mind when they chose number as the category for illustrating inflectional morphology. That is to say, we examine various possible interpretations of the claim that number is inflectional. Rather than adding to the discussion of the meaning of 'inflectional',[3] which is not our primary concern in this brief paper, we shall choose the cases discussed so that they are as far as possible

[1] The support of the British Academy and of the ESRC (grant R000236063) is gratefully acknowledged. I also wish to thank Norman Fraser and Andrew Hippisley for helpful comments on a draft. A version was read at the First Mediterranean Morphology Meeting (Mytiline 1997); I am grateful to those present for useful discussion.

[2] Others who have discussed the status of number include Kuryłowicz (1964: 16-17), who makes a distinction within inflectional forms of the same word between those which vary only in syntactic value (as with case) and those which differ semantically (as with number), and Beard (1982), who adopts the opposite position to the common one, in arguing against an inflectional interpretation of number. Interestingly, in a brief discussion Dressler treats it as not prototypically inflectional (1989: 6). A recent psycholinguistic perspective is provided by Baayen, Lieber & Schreuder (1997).

[3] See Scalise (1988), Plank (1994) and references there.

'consensus cases', where different definitions of inflectional (in the broad sense to cover both inherent and contextual inflection) converge on the same result.

Hypothesis 1: All languages have the category of number and it is inflectional

At least the first part of this claim is widely accepted. For instance:

> 'All languages have pronominal categories involving at least three persons and two numbers.' Greenberg (1963: Universal 42).

This reasonable claim appears to be incorrect. Let us consider Pirahã, the only remaining member of the Mura family, spoken in 1997 by some 220 people along the Maici River (Amazonas, Brazil). It has been described by Everett (1986) on the basis of fourteen months of intensive contact with the Pirahã, updated (1997) after five years of fieldwork. He states (1986: 217): 'there are no plural forms in Pirahã'. This holds even for pronouns, whose free forms are as follows (1986: 280):

first person	ti
second person	gíxai
third person	hiapióxio

Table 1: Personal pronouns in Pirahã

'There are no special plural forms for these pronouns.' This means that *hiapióxio* (third person) can be plural or singular, as this example shows (1986: 282):

(1) hiapióxio soxóá xo-ó-xio
3RD already jungle-LOC-DIR
(i) 'He already went to the jungle' or
(ii) 'They already went to the jungle'

There are ways of expressing what in other languages would be plurality, by conjoining, for instance (1986: 281):

(2) ti gíxai pí-o ahá-p-i-í
1ST 2ND also-OBL go-IMPRF-PROX-COMPLETE.CERT
'You and I will go (i.e. we will go)'
[abbreviations: OBLique, PROXimate, CERTainty]

There are other means for expressing the notion of plurality:[4] the associative/comitative postposition *xigí* and various quantifiers. But this does not mean that the language has a number category; after all, English can express duality through the use of *two* and *both*, but this does not mean that English has a dual. The grammar of English does not need to

[4] More generally, in our discussions of whether a particular language has number, and for which word classes, we should bear in mind that number may be expressed indirectly, for example through distributivity, in order not to be misled by phenomena of this type.

refer to a value 'dual'. Similarly in Pirahã, from Everett's description, the grammar has no need to refer to a value 'plural'. We conclude that Pirahã has no number category.

Kawi (Old Javanese) is reported to have been similar to Pirahã in this respect, in not having plural nouns or pronouns, but marking number by conjoining pronouns or by quantifiers such as 'many' and 'all' (Becker & Oka 1974: 232).

From now on we shall consider only languages with a number category, and assume this in our hypotheses.

Hypothesis 2: Where number is found it will be inflectional

This weakening of the original claim allows for there to be languages without number, but claims that where it is found, number will always be inflectional. It too is false, there is no universal list specifying that categories must be of a particular type. In fact genuine verbal number (rather than nominal number found on verbs by agreement) is typically derivational (Durie 1986; Mithun 1988a, 1988b). Verbal number has been claimed to exist in many languages. It is particularly widespread in North America; it is also found in the South Central Dravidian group of languages of southern India (Steever 1987) and in many languages of Africa (Brooks 1991), the Chadic group being particularly well documented (Newman 1990: 53-87). A major analysis of the subject is that of Durie (1986); Frajzyngier (1985) was a forerunner and Mithun (1988a) gives a diachronic perspective.

The meaning of verbal number is still not well researched; and the difficulty is compounded by the fact that the terminology is not standardized. For example, Eulenberg discussing a reduplicated verb in Hausa says that it represents:

> 'a derivational category widespread among Nilo-Saharan and Afro-Asiatic languages, though rather marginal in Niger-Congo. This category is variously known as the *intensive*, *habitative*, *frequentative*, *repetitive*, or *plural* verb. ... it has the general meaning of a repeated action, an action simultaneously performed by several agents, an action performed on more than one object, or various combinations of these "plural" meanings.'
>
> Eulenberg (1971: 73)

There are two main types of verbal number: **event number** and **participant number**. We will consider an example of event number here, and an example of participant number (from Georgian) below. Event number can be illustrated from Hausa (a Chadic language, Chadic being one of the branches of Afro-Asiatic); the data are from Eulenberg (1971: 73-74):

(3) naa aikee su[5]
I.COMPL send them

(4) naa a"aikee su
I.COMPL send.PL them

[5] *naa* 'I' is in a form marking completive aspect (COMPL); the verb is *aikaa* 'to send' but the *-aa* ending changes to *-ee* because of the presence of a pronominal object.

Note that both have a singular subject and a plural object. Example (3) has a simple verb, but (4) has a verb with partial reduplication, which marks it as 'intensive' or 'plural'. Example (3) can be used with the meaning 'I sent them at the same time to the same place' and (4) would not be appropriate there. Both examples could be used with the following meanings:

(i) I sent them at the same time to different places
(ii) I sent them at different times to the same place
(iii) I sent them at different times to different places

Thus the plural verb *a"aikee* indicates that the sending was not simple; rather it involved more than one time or more than one place - more than one 'sending-event'. Its use is not obligatory, however. The important thing is that the use of the 'plural' verb here indicates the number of sendings; it is an instance of verbal number.

Hypothesis 3: For a given language, number will be either inflectional or not inflectional (but not both)

This claim is false: there is not necessarily one answer for a given language: number may be both inflectional and derivational. We can conveniently show this, together with disproving the following, weaker claim.

Hypothesis 4: For a given lexical class, number will be either inflectional or not inflectional (but not both)

Surprisingly, perhaps, inflectional and derivational number may co-occur on a single item. We shall see this in examples involving the participant type of verbal number, in the South Caucasian (or Kartvelian) language Georgian (Aronson 1982: 243, 406-407, quoted in Durie 1986):

(5) ivane še-mo-vid-a da da-ǰd-a
John PRV-PRV-enter-AOR.3.SG and PRV-sit.SG-AOR.3.SG
'John entered and sat down'(PRV = preverb, AOR = aorist)

(6) čem-i mšobl-eb-i še-mo-vid-nen da
my-AG parent-PL-NOM PRV-PRV-enter-AOR.3.PL and

da-sxd-nen
PRV-sit.PL-AOR.3.PL
'My parents entered and sat down'

[AG indicates an agreement marker; the ending *-i* is syncretic, covering nominative singular and plural, and genitive singular and plural]

The verbs agree in number in a straightforward way. This is nominal number expressed on the verb by agreement. It is inflectional. Additionally, though, the verb 'sit' (unlike the verb 'enter') is one of those which has different derived forms according to whether one person sits (*daǰd-*), or more than one (*dasxd-*). The choice can be seen as a case of

verbal number, determined by semantic considerations. Now consider what happens when there is a numeral phrase. Numerals require a singular noun (*megobari* 'friend', the plural would be *megobr-eb-i)* and the resulting phrase controls singular agreement :

(7) čem-i sam-i megobar-i še-mo-vid-a da
my-AG three-AG friend.SG-NOM PRV-PRV-enter-AOR.3.SG and

da-sxd-a
PRV-sit.PL-AOR.3.SG
'My three friends entered and sat down'

Singular agreement is found on both verbs. Yet the second, which has two forms according to the number of participants, shows the plural verbal form *dasxd-*, since more than one participant is involved in the action. In other words, the verb is plural in terms of verbal number, but this does not determine the agreement, which is singular. Thus in Georgian we have derivational and inflectional number together. And they can take different values.

Our rejection of hypotheses 2-4 has depended on the notion of verbal number. Some might not accept that the verbal opposition in the Hausa and Georgian examples above is an instance of the category of number. It could be argued that this was a case of aspect. Repeated versus non-repeated action is a classic aspectual distinction. There is a clear link between aspect and nominal number: if a language marks repeated action in some way, this is much more likely to be found when plurality is involved than without it (in the real world, a single person is, for instance, unlikely to send a single package repeatedly). Alternatively we might analyze the Hausa example as showing distributivity. The examples of participant number (as in Georgian) are perhaps harder to discount. However, for those who would restrict number to nominal number (including nominal number expressed on the verb by agreement), it still does not follow that hypotheses 2-4 hold. They will be disproved using different evidence along with hypothesis 5.

Hypothesis 5: At least for the nominals in a given language, taking them together, number will be either inflectional or not inflectional (but not both)

We might expect that if there is inflectional nominal number, it will occur throughout the nominals. This claim too, is false. There can be splits within the nominals. This has been known for some time, but the theoretical consequences have generally not been thought through. There are several examples; we will take a less usual one, namely Marind, which belongs to the family of the same name and has about 7000 speakers in southern Irian Jaya. The data, originally from Drabbe (1955: 19-20), are presented in Foley (1986: 78, 82-83).[6] Marind has four genders (which we designate I-IV in the examples), and nouns are assigned to them as follows: gender I is for male humans, gender II for female humans and animals, gender III is mainly for plants and trees, while the semantic residue makes up gender IV. First we see examples of genders I and II:

[6] As yet I have unfortunately been unable to gain access to a copy of the original.

(8) e-pe anem e-pe akek ka
I-the male.person I-the light.I be
'the man is light'

(9) u-pe anum u-pe akuk ka
II-the female.person II-the light.II be
'the woman is light'

(10) u-pe ŋgat u-pe akuk ka
II-the dog II-the light.II be
'the dog is light'

The agreement is prefixed on *-pe* 'the' but infixed in the adjective *ak-k* 'light'. In the plural, the forms are these:

(11) i-pe anim i-pe akik ka
PL-the person.PL PL-the light.PL be
'the people are light'

(12) i-pe ŋgat i-pe akik ka
PL-the dog PL-the light.PL be
'the dogs are light'

There is just one plural agreement form for genders I and II. *Anum* 'man' has the plural *anim*; while *ŋgat* 'dog/dogs' does not change morphologically. There are, however, nouns denoting animals which mark number, for instance *namakud* 'animal' has the plural *namakid*. Though not marking number itself, *ŋgat* when plural takes plural agreements. For genders III and IV, the forms are these:

(13) e-pe de e-pe akak ka
III-the wood III-the light.III be
'the wood is light'

(14) i-pe behaw i-pe akik ka
IV-the pole IV-the light.IV be
'the pole is light'

Nouns of genders III and IV, those which are 'below' animals, have no distinct plural forms and no plural agreement forms. (Note that the gender IV marker is the same as the plural marker for genders I and II.)

This is one instance of a more general claimed regularity. Smith-Stark (1974) proposed this version of the Animacy Hierarchy:

speaker > addressee > kin > rational > human > animate > inanimate
(1st person pronouns) (2nd person pronouns)

Figure 1: The Smith-Stark (Animacy) Hierarchy

He claimed that when plurality 'splits' a language, some top segment of the hierarchy will be involved in plural marking. For some languages there is a relatively clear split within the nominals, for others it is much less clear, with optional marking available at some positions on the hierarchy. What matters here, however, is that it is quite normal for nominals at different points on the hierarchy to behave differently with regard to number.

It is tempting to claim that number is inflectional for the count nouns of a language. However, this use of 'count' leads to circularity, if it means no more than the nouns which have inflectional number. To avoid circularity we would need to show that items denoted by nouns below the count noun threshold of the particular language, are not counted. This is certainly not the case for the Miya examples discussed below (see especially example (16)).

The Marind data suggest new hypotheses, in that the examples include marking of number both on the noun and through agreement. The first is an instance of 'inherent' inflection, while agreement shows that the number of the noun (through the noun phrase of which it is the head) also has a role in contextual inflection (Booij 1996: 28). There are at least two hypotheses to consider with respect to this distinction, one leading to the other.

Hypothesis 6: For all the nominals in a given language, number will be a category of inherent inflection or it will have a role in contextual inflection

The Marind data are sufficient to disprove this hypothesis. Nominals below the animacy threshold are outside the number system, both in terms of marking number and in terms of agreement. We are not dealing with isolated exceptions but with a substantial proportion of the noun inventory.

This suggests a further hypothesis:

Hypothesis 7: For each use of each nominal the value of the number category for inherent inflection must match the value for its role in contextual inflection

This makes the reasonable claim that those nominals for which number marking is available will match those which can head noun phrases controlling number agreement. Thus there will be a single cut-off point on the Animacy Hierarchy. If it were true, it would mean that for investigating number in nominals the inherent/contextual distinction was not relevant. However, we shall see that even this claim does not hold.

We might think of British English *committee* type nouns here, since they allow plural agreement while standing in the singular. However, these are a special case in that their agreements need not be consistent (*this committee, after long deliberation, have decided to ...*).

There is a more clear-cut counter-example. The relevant data are found in the West Chadic language Miya (Schuh 1989); the split involves obligatory/optional number marking and obligatory/excluded agreement. Number is involved in agreement and hence

is relevant to syntax; furthermore: 'Potentially, any noun may be pluralized morphologically.' (Schuh 1989: 173). Hence by almost any definition the language has inflectional number. Let us look at its distribution. Nouns are of two genders, masculine and feminine; males are masculine, females feminine, and non-sex differentiables can be either. Agreement targets (and many different items agree) have three agreement forms: masculine singular, feminine singular and plural. This may be illustrated by one of the demonstrative pronouns:

	singular	plural
masculine	nákɔ́n	níykín
feminine	tákɔ́n	

Table 2: The demonstrative 'this' in Miya (Schuh 1989: 172, 176)

In addition there is an animate/inanimate distinction: the animate nouns are those which denote 'all humans, most, if not all, domestic animals and fowl, and some large wild animals.' Large wild animals are the 'grey area'. The remaining nouns are inanimate (1989: 175). This distinction is relevant for number marking in that animate nouns must be marked for plurality when appropriate:

(15) tə̀vàm tsɔ́r
woman.PL two
'two women'

cf: *'ám tsɔ́r
woman.SG two
*'two women'

For inanimates on the other hand marking is optional:

(16) zə̀kìyáyàw vàatlə̀
stone.PL five
'five stones'

cf: zə̀kìy vàatlə̀
stone.SG five
'five stones'

Animate plural nouns take plural agreements:

(17) níykín tə̀vàm
this.PL woman.PL
'these women'

Inanimate nouns, however, even if they are marked as plural, do not take plural agreement; they take agreement according to their gender in the singular:

(18) nákɔ́n viyáyuwáwàw
this.M.SG fireplace.PL
'these fireplaces'

(*vìyàyúw* 'fireplace' is masculine)

(19) tákə́n tlə́rkayáyàw
this.F.SG calabash.PL (*tlə́rkay* 'calabash' is feminine)
'these calabashes'

Thus the status of number is different for animate and inanimate nouns. Marking of number is obligatory for animates but optional for inanimates. Number is syntactically relevant, since it is an agreement category; however, while agreement in number with animates is obligatory, plural agreement with inanimates is impossible. And, most interestingly, agreement with inanimate plurals does occur, but in gender and not in number. This shows that there is an agreement rule for inanimates where we might have expected to find number agreement, but where the latter fails to occur. Thus inanimate nouns have inherent number, marked optionally, but this number does not have a role in contextual inflection. The value of the number category for inherent inflection need not match the value for its role in contextual inflection and hypothesis 7 is shown to be false.

At least, we might think, the mismatch will always be this way:

Hypothesis 8: For the nominals in a given language, where the role of the number category differs for inherent inflection and contextual inflection, the role of inherent inflection will extend lower down the Animacy Hierarchy than that of contextual inflection.

This proves to be another reasonable but false supposition. Consider Merlan's (1983) account of Ngalakan, a language of the Gunwinjguan group, which had around 25 speakers in the late 1970's, at Bulman and Ngukurr in Arnhem Land, Australia. Here too, marking of number on the verb is sensitive to position on the hierarchy:

> '... in Ngalakan explicit non-singular marking on the noun is limited; nouns not explicitly marked as non-singular can be cross-referenced as non-singular, but this possibility is limited almost entirely to human and sometimes animate nouns. Non-singular reference of inanimate NPs is generally not explicitly marked in the verb, and is largely to be understood from the larger context of discourse.' (Merlan 1983: 90)

The implication of the interaction of number with the Animacy Hierarchy is that the status of number as an inflectional category is much less straightforward than generally imagined. It really is not a simple inflectional feature (+/- plural) available to play a role in the syntax. It is also worth mentioning here that, to keep things simple, discussion has been restricted to singular and plural. Other values of the number category add whole layers of complexity: it is not that case that, for example, in a singular-dual-plural system what is true for the plural will be true for the dual. They can vary independently.

However, after several hypotheses which have been proved false, it is time to suggest a new one, which it is hoped will prove correct:

Hypothesis 9: For the nominals in a given language, where the role of the number category differs for inherent inflection and contextual inflection, there may be counter-examples to the requirement of the Animacy Hierarchy in terms of inherent inflection but not in terms of contextual inflection.

We can illustrate the effect of this constraint from English, where the number split is very low on the hierarchy, being found within the inanimates. Nouns like *sheep* are therefore exceptional in terms of number marking:

(20) This sheep has been cloned.
(21) These sheep have been cloned.

Since sheep are animate, the noun would be expected to mark number (as indeed it once did). The noun is irregular in terms of inherent inflection, but regular in terms of its role in contextual inflection (it takes plural agreement when plural). Imagine a new lexical item *peesh* (a cloned sheep). It could not be the grammatical reversal of *sheep*:

(22) This peesh has been fed. [Hypothetical: singular]
(23) This peeshes has been fed. [Hypothetical: plural: claimed impossible]

At first sight, the hypothetical system which is claimed to be impossible looks rather like that which is found in Miya. The difference is that in Miya there are two splits, different for noun marking and agreement, but both in accord with the Animacy Hierarchy. English *sheep* is not part of a regular split but is a lexical exception. Exceptions of this type are allowed, while the converse, like the hypothetical *peesh*, are not.

Conclusion

Number, which is taken so readily as an illustrative case of inflectional morphology, is a category whose status is hard to determine. We have seen a list of reasonable hypotheses which have been proved false, and just one which appears promising. The status of number is clearly worth pursuing further.[7]

References

Aronson, Howard I. 1982. *Georgian: a Reading Grammar.* Columbus, Ohio: Slavica.

Baayen, R. Harald; Lieber, Rochelle & Schreuder, Robert 1997. The morphological complexity of simplex nouns. *Linguistics* 35.861-877.

Beard, Robert 1982. The plural as a lexical derivation. *Glossa* 16.133-48.

Becker, A. L. & Oka, I. Gusti Ngurah 1974. Person in Kawi: exploration of an elementary semantic dimension. *Oceanic Linguistics* 13.229-255.

Bloomfield, Leonard 1933. *Language.* New York: Holt, Rinehart & Winston.

Booij, Geert 1993. Against split morphology. In: Geert Booij & Jaap van Marle (eds) *Yearbook of Morphology 1993,* 27-49. Dordrecht: Kluwer.

Booij, Geert 1996. Inherent versus contextual inflection and the split morphology hypothesis. In: Geert Booij & Jaap van Marle (eds) *Yearbook of Morphology 1995,* 1-15. Dordrecht: Kluwer.

Brooks, Bryan 1991. Pluractional verbs in African languages. *Afrikanistische Arbeitspapiere* (Cologne) 28.157-68.

Corbett, Greville G. forthcoming. Prototypical inflection: implications for typology. To appear in: Geert Booij & Jaap van Marle (eds) *Yearbook of Morphology 1998.* Dordrecht: Kluwer.

Corbett, Greville G. in preparation. *Number.* Cambridge: Cambridge University Press.

[7] This research is continued in Corbett (forthcoming, in preparation).

Drabbe, P. 1955. Spraakkunst von het Marind. *Studia Instituti Anthropos* 11. [Cited from Foley 1986.]
Dressler, Wolfgang U. 1989. Prototypical differences between inflection and derivation. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 42.3-10.
Durie, Mark 1986. The grammaticization of number as a verbal category. In: Vassiliki Nikiforidou, Mary VanClay, Mary Niepokuj & Deborah Feder (eds) *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society: February 15-17, 1986,* 355-70. Berkeley, California: B. L. S., University of California.
Eulenberg, John B. 1971. Conjunction reduction and reduplication in African languages. In: Chin-Wu Kim & Herbert Stahlke (eds) *Papers in African Linguistics* (Current Inquiry into Language and Linguistics 1), 71-80. Edmonton: Linguistic Research.
Everett, Daniel 1986. Pirahã. In: Desmond C. Derbyshire & Geoffrey K. Pullum (eds) *Handbook of Amazonian Languages: I,* 200-325. Berlin: Mouton de Gruyter.
Everett, Daniel 1997. Pirahã. (ESRC Research Seminar Series "Challenges for Inflectional Description" - III, University of Surrey, 13.5.1997). Unpublished presentation.
Foley, William A. 1986. *The Papuan Languages of New Guinea.* Cambridge: Cambridge University Press.
Frajzyngier, Zygmunt 1985. Ergativity, number, and agreement. In: Mary Niepokuj, Mary Van Clay, Vassiliki Nikiforidou & Deborah Feder (eds), *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society, February 16-18, 1985,* 96–106. Berkeley, California: B. L. S., University of California.
Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In: Joseph H. Greenberg (ed.) *Universals of Language,* 73-113. Cambridge, Massachusetts: MIT Press.
Jespersen, Otto 1924. *The Philosophy of Grammar*. London: Allen & Unwin.
Kuryłowicz, Jerzy 1964. *The Inflectional Categories of Indo-European.* Heidelberg: Carl Winter.
Marle, Jaap van 1996. The unity of morphology: on the interwovenness of the derivational and inflectional dimension of the word. In: Geert Booij & Jaap van Marle (eds) *Yearbook of Morphology 1995,* 67-82. Dordrecht: Kluwer.
Matthews, Peter H. 1991. *Morphology*. Cambridge: Cambridge University Press. Second edition [first edition 1974]
Merlan, Francesca 1983. *Ngalakan Grammar, Texts and Vocabulary* (Pacific Linguistics, series B, no. 89). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
Mithun, Marianne 1988a. Lexical categories and the evolution of number marking. In: Michael Hammond & Michael Noonan (eds) *Theoretical Morphology: Approaches in Modern Linguistics,* 211-34. San Diego: Academic Press.
Mithun, Marianne 1988b. Lexical categories and number in Central Pomo. In: William Shipley (ed.) *In Honor of Mary Haas: From the Haas Festival Conference on Native American Linguistics,* 517-37. Berlin: Mouton de Gruyter.
Newman, Paul 1990. *Nominal and Verbal Plurality in Chadic* (Publications in African Languages and Linguistics 12). Dordrecht: Foris.
Plank, Frans 1994. Inflection and derivation. In R. E. Asher (ed.) *Encyclopedia of Language and Linguistics* volume 3, 1671-1678. Oxford: Pergamon Press.
Scalise, Sergio 1988. Inflection and derivation. *Linguistics* 26.561-581.

Schuh, Russell G. 1989. Number and gender in Miya. In: Zygmunt Frajzyngier (ed.) *Current Progress in Chadic Linguistics: Proceedings of the International Symposium on Chadic Linguistics: Boulder, Colorado, 1-2 May, 1987* (Current Issues in Linguistic Theory 62), 171-181. Amsterdam: Benjamins.
Smith-Stark, T. Cedric 1974. The plurality split. In: Michael W. La Galy, Robert A. Fox & Anthony Bruck (eds) *Papers from the Tenth Regional Meeting, Chicago Linguistic Society, April 19-21, 1974,* 657-71. Chicago: Chicago Linguistic Society.
Steever, Sanford B. 1987. The roots of the plural action verb in the Dravidian languages. *Journal of the American Oriental Society* 107.581-604.
Stump, Gregory T. 1990. Breton inflection and the split morphology hypothesis. In: Randall Hendrick (ed.) *The Syntax of the Modern Celtic Languages* (Syntax and Semantics 20), 97-119. San Diego: Academic Press.

Ash Göksel, Bosphorus University and University of London
gokselas@boun.edu.tr & ag7@soas.ac.uk

WORD LENGTH

Abstract
In this paper it is argued that the process of affixation is subject to a morphological constraint which determines the size of a word. Languages vary with respect to the affix space they have, that is, the number of affixes in a word depends on the number of slots available for affixes in that language. This hypothesis is based on certain morphological and syntactic differences between two Turkic languages, Turkish and Yakut. The morphological form of main and embedded verbs, the presence and behaviour of buffer stems, the interaction of affixation and auxiliaries, the cross-linguistic asymmetries in the presence of complementisers and double nominal constructions (possessive phrases and relative clauses) indicate that Turkish has a maximum number of three slots for inflectional affixes whereas Yakut has two. According to this proposal morphological and syntactic processes relating to inflectional morphology are not coextensive, and morphology is a separate component which contains restrictions on word length.

Introduction
One of the central issues regarding morphology is its status within the structure of grammar. This issue becomes particularly relevant when approaching inflectional morphology. In a number of studies, inflection is characterised purely in syntactic terms, by means of representing inflectional affixes in syntax (Pollock 1989, Chomsky 1993). The part morphology plays is reduced to lexical features and their role in deriving syntactic representations. The implicit assumption is that there is not much to be said about inflectional morphemes other than their syntactic manifestation. An alternative route of research highlights the differences between word structure and phrase structure (Di Sciullo and Williams 1987, Anderson 1992, Bresnan and Mchombo 1995, Sells 1995, Di Sciullo 1996, Bach 1996, among others). One common factor in these studies is that, although there may be overlaps, morphological structure is not an extension of syntactic structure, and vice versa.

The proposal that I put forward in this paper supports the view that morphology is a separate component. Its architecture contains, among other things, a condition on word length characterised in terms of morphological slots. In particular, I propose that the number of slots available determines the upper limit of affixation. Based on the differences between verbal forms in main clauses and embedded clauses in Turkic languages, it will be argued that some of the differences in the inflectional paradigm of these languages can only be explained in tems of a cross-linguistic variation in morphological space. In addition, this hypothesis explains certain variations in the clause structure of Turkic languages, it accounts for the defective nature of the tense/aspect paradigm in relative clauses and sheds light on the placement of agreement

morphology, the occurrence of complementisers, the availability of multiple nominal constructions and the presence of buffer stems. It also explains why certain forms which are syntactically allowed do not exist.

Parts 1 and 2 focus on main verbs and embedded verbs in Turkish and Yakut, a Turkic language spoken in north-eastern Siberia. A comparison of the two languages leads to the proposal concerning morphological templates in the third section. The fourth section gives further evidence for morphological templates. Section 5 provides a discussion of morphological slots, followed by the implications of the present hypothesis for syntax in the final section.

1.Turkish

1.1 Main verbs
Inflected verbs in Turkish are minimally composed of a stem followed by a tense suffix and by a subject agreement marker:

(1)	Ben	balığ-ı	ye-di-m.
	I	fish-ACC	eat-PAST-1
	I ate the fish.		

Other suffixes which may occur in the place occupied by the past tense marker are the folowing:

(2)	Ben	balığ-ı	yi-(y)eceğ-im	FUTURE: I will eat the fish.
	Ben	balığ-ı	ye-r-im.	HABITUAL: I eat fish.
	Ben	balığ-ı	yi-yor-um.	PROGRESSIVE: I am eating fish.
	Ben	balığ-ı	ye-se-m	CONDITIONAL: If I eat the fish
	Ben	balığ-ı	ye-miş-im.	HEARSAY: I must have eaten the fish.

An inflected verb may have two temporal/aspectual suffixes:

(3)	a.Ye-r-di-m.	b.Yi-yor-muş-um.	c.Yi-(y)ecek-se-m...
	eat-AOR-PAST-AGR	eat-PROG-HS-AGR	eat-FUT-COND-AGR
	I used to eat.	I appear to have been eating.	If I am going to eat...

The differences between temporal and aspectual suffixes do not have a bearing on the present analysis and will not be discussed here. For practical purposes, I will refer to all tense and aspect suffixes as T/A, and the part of the verb containing T/A suffixes and agreement as the inflectional domain. The inflectional domain follows grammatical function changing suffixes, modals and negation. The data above indicates that the morphological structure of the main verb with respect to inflection is:
(4) V-T/A-(T/A)-AGR

1.2 Embedded verbs
Verbs in complement clauses and object relative clauses are also inflected for T/A and agreement[1]:

[1]Other types of embedded clause such as the subject relative clause which is not inflected for T/A or person will nor be discussed here.

(5) a. *Complement clause*

[Ben-im balığ-ı ye-diğ-im] ortada.
I-GEN fish-ACC eat-PAST-C-AGR is obvious
It is obvious that I eat/ate the fish.

b. *Object relative clause*

[Ben-im ye-diğ-im balık] barbunya.
I-GEN eat-PAST-C-AGR fish red mullet
The fish that I eat/ate is red mullet.

-diğ often denotes non-future events. It is morphologically replaceable only by one other form, *-(y)eceğ* , which denotes future events:

(6) a. *Complement clause*

[Ben-im balığ-ı yi-yeceğ-im] biliniyor.
I-GEN fish-ACC eat-FUT-C-AGR is known
It is known that I will eat the fish.

b. *Object relative clause*

[Ben-im yi-yeceğ-im balık] barbunya.
I-GEN eat-FUT-C-AGR fish red mullet
The fish that I will eat is red mullet.

In grammars of Turkish both *-diğ* and *-(y)eceğ* are treated as single morphemes, as they are in the generative literature (Özsoy 1988, Kennelly 1992), with the exception of Kural (1993) who analyses *-ğ* in *-(y)eceğ* as a separate morpheme. I suggested elsewhere (Göksel 1997) that this analysis was preferable for morphological reasons and extended it to include *-diğ*, hence *-di+ğ*, and represented *-ğ* as the complementiser in Turkish. The motivation for this analysis is based on the fact that a syntactically discrete structure (subordination) is marked phonologically (*-ğ* being vowel length), shown below:

(7)	Main clause	Subordinate clause
non-future	/yedim/ (I ate)	/yedi:m/ (that I ate)
future	/yiycæm/ (I will eat)	/yiycæ:m/ (that I will eat)[2]

There are a number of differences between embedded verbs and matrix verbs. Embedded verbs are only marked for future versus non-future specification, and they lack the array of T/A suffixes which occur in matrix verbs. Embedded verbs cannot occur with any of the markers in (2) (other than *-(y)eceğ*). Then, they cannot have more than one T/A which is possible in matrix verbs:

[2]See Göksel 1997 for a discussion of the dialectal variations of this form.

(8) a. Ye-r-di-m
eat-AOR-PAST-AGR
I used to eat

b. *ye-r-di-ğ-im balık
eat-AOR-PAST-C-AGR NP
(Intended. interpretation: the fish I used to eat)

Finally, embedded clauses are nominal constructions, which is indicated by the fact that they occur with agreement suffixes from the nominal paradigm and that they are inflected for case. It then seems to be the case that the complementiser C has the dual function of marking nominalisation and subordination. The structure of the embedded verb is:

(9) V-T/A-(*T/A)-C-AGR

A comparison of the structure of main verbs and embedded verbs indicates that the position occupied by an optional T/A marker in the main verb is taken up by the complementiser in the embedded verb:

(10) Main verbs V - T/A - (T/A) - AGR
Embedded verbs V - T/A - C - AGR

2. Yakut

2.1 Main verbs

As in Turkish, verbs in Yakut main clauses are made up of a stem followed by a T/A marker and a person marker:

(11) a. Min bahı-ı si-ebit-im.
I fish-ACC eat-PAST-AGR
I ate the fish.
b. Min balı-ı si-eγ-im. FUTURE: I will eat the fish.
c. Min balı-ı si-y-bin. HABITUAL: I eat fish .

However, unlike Turkish, Yakut verbs can only have one T/A, illustrated below:

(12) *si-y-bit-im
eat-AOR-PAST-AGR

indicating that the structure of the matrix verbs is:

(13) V-T/A-(*T/A)-AGR

2.2 Embedded verbs

In object relative clauses in Yakut the agreement marker attaches to the head noun:

(14) a. Min si-ebit balı-ım sobo.
I eat-PAST fish is swordfish
The fish that I ate is swordfish.

and it cannot occur on the verb as the ungrammaticality of (15a) illustrates. As in Turkish, the embedded verb does not have the array of T/A markers found in main verbs, illustrated by the ungrammaticality of (15b):

(15) a. *min si-ebit-im balık (Intended interpretation: the fish that I ate)
b. *min sie-γ balı-ım (Intended interpretation: the fish that I will eat)

The morphological structure of the embedded clause is:

(16) V-T N-AGR

3. A comparison of Yakut and Turkish morphology

Yakut and Turkish are similar in terms of the unavailability of some T/A affixes in embedded clauses. However, there are non-trivial differences between the two. The inflectional domain of the main verb in Turkish contains three suffixes as opposed to two in Yakut. The number of suffixes is again three in Turkish embedded verbs whereas in Yakut embedded constructions there is one suffix each on the verb and on the head of the nominal construction, summarised below:

(17)	*Turkish*	*Yakut*
Main verbs	V-T/A- (T/A) - AGR	V-T/A-AGR
Embedded verbs	V-T/A- C -AGR	V-T/A N-AGR
Complementiser	✓	×

I suggest that these dissimilarities are the manifestation of a difference in the morphological structure of Turkish and Yakut. Turkish has three slots available within the inflectional domain whereas Yakut has two slots, illustrated below:

(18)	*Turkish*	*Yakut*
	V.... X^1 X^2 X^3	V..... X^1 X^2

where X'stands for any inflectional morpheme. The difference is thus reducible to the space that is available in each language: Turkish has a morphological template that has three slots for inflectional affixes, whereas the morphological template in Yakut has two slots.

The existence of morphological templates explains the absence of a second T/A marker in embedded verbs in Turkish, this position being occupied by C. It also explains why the agreement marker in Yakut has to occur on the noun rather than on the verb in an embedded clause. If it were to occur on the verb it would require a nominaliser, as it is a nominal agreement marker. However, the presence of a nominaliser would itself have the effect of filling up the morphological space. An additional suffix, in this case agreement, would violate the constraint that no more than two suffixes are allowed in Yakut.

4. Further evidence for the presence of morphological templates

The presence of morphological templates explains a number of constructions in Turkish and Yakut, such as the occurrence of buffer stems, the structure of auxiliaries and possessive constructions.

4.1 Buffer stems

Both Turkish and Yakut use buffer stems as a strategy for further affixation. As explained in 1.2, the embedded verb in Turkish does not have the array of T/A affixes found in main verbs. A verbal form such as (19a) does not have a corresponding embedded form (19b), which is ungrammatical for the reasons explained above. Instead, (19c) corresponds to (19a).

(19) a. Oku-muş-tu-n
1 2 3
read-HS-PAST-AGR
You had read.

b. *oku-muş-tu-ğ-un kitap
read-HS-PAST-C-AGR book
1 2 3 4
(Intended interpretation: the book that you had read)

c. oku-muş ol-du-ğ-un kitap
read-HS AUX-PAST-C-AGR book
1 1 2 3
the book that you had read.

(19c) contains the copular verb *ol* 'be' which, here, serves no other purpose than to act as a buffer stem to which additional affixation can be attached. Yakut also uses a copular verb as a buffer stem. A form such as (20a) which is semantically possible does not exist, as the morphological template of Yakut has two slots for the inflectional domain. Instead the copular verb *olor* 'be' is used solely for reasons of morphology.

(20) a. *Si-i-bit-im.
1 2 3
eat-AOR-PAST-AGR
(Intended interpretation: I had been reading.)

b. Si-i olor-but-um.
eat-AOR AUX-PAST-AGR
1 1 2
I had been reading.

4.2 Modals

Modal auxiliaries exhibit similar properties. In constructions with the modal verb *buol* 'can/may' in Yakut, the two stems share the inflectional affixes. Again, as expected, neither stem has more than two suffixes, and the person marker can either appear on the main verb as in (21b) or on the auxiliary, as in (21c):

(21) a. *Si-i(r)-i-im buol
eat-AOR-FUT-AGR may
1 2 3
(Intended interpretation: I will probably eat.)

b. Si-i(r)-im buol-uox
eat-AOR-AGR may-FUT
1 2 1
I will probably eat.

c. Si-i(r) buol-uoy-um
eat-AOR may-FUT- AGR
1 1 2
I will probably eat.

4.3 Double possessives

Embedded clauses in Turkish and Yakut are nominal constructions which have similar properties to possessive phrases. In the possessive construction in Turkish, the possessor is in the genitive case, as is the subject of an embedded clause. Both constructions have nominal agreement:

(22)

embedded clause			*possessive phrase*	
ben-im	dik-ti-ğ-im	elbise	ben-im	elbise-m
I-GEN	sew-PAST-C-AGR	dress	I-GEN	dress-AGR
the dress that I sewed			my dress	

The position of the agreement marker in embedded clauses makes it possible to have double possessive constructions in Turkish, where the relative clause is embedded in the possessive construction:

(23)

ben-im	şu	sen-in	dik-ti-ğ-in	elbise-m
I-GEN	this	you-GEN	sew-PAST-C-AGR2	dress-AGR1
this dress of mine which you sewed				

In Yakut, embedded clauses are also similar to possessive phrases:

(24)

embedded clause			*possessive phrase*	
min	si-ebit	balɪ-tɪm	min	balɪ-tɪm
I(NOM)	eat-PAST	fish-AGR	I(NOM)	fish-AGR
the fish that I ate			my fish	

However, an embedded clause cannot occur within a possessive phrase:

(25)

a. *min	en	si-ebit	balɪ-tɪm	
I	you	eat-PAST	fish-AGR1	
b. *min	en	si-ebit-in	balɪ-tɪm	
I	you	eat-PAST-AGR2	fish-AGR1	
c. *min	en	si-ebit	balɪ-tɪm-ɪn	
I	you	eat-PAST	fish-AGR1-AGR2	

(Intended interpretation: the fish of mine which you ate)

Although morphologically well-formed, (25a) is ungrammatical because of the absence of an agreement marker. The occurrence of second person agreement marking on the embedded verb, as in (25b) results in ungrammaticality as well, this time for both morphological and syntactic reasons. Although complying with the requirements of the morphological template, it is ungrammatical as a result of the presence of nominal agreement on the verb, as AGR2 is a member of the nominal paradigm. If it were to appear, it would require a nominaliser. However, the presence of a would violate the template, as explained above, in section 3. The ungrammaticality of (25c) arises from

two agreement markers occurring one after the other, a violation of slot-(affix) type mismatches.

5. The nature of morphological slots

The discussion so far has centered around motivating morphological templates and how they provide the slots for hosting affixes. A question that arises at this point concerns the pairing of affixes and slots. The fact that the order of affixes is not free and that there are restrictions on the co-occurrence of certain affixes indicate that affixes and slots must match, an issue discussed in Sells (1995). There it is argued that the occurrence of certain affixes in certain positions, as well as the non co-occurrence of particular affixes are due to their being typed to occur in a particular slot.

The ungrammaticality of (25c) can then be explained in terms of slot-type mismatches, where, for two affixes of the same type (AGR1 and AGR2), there is only one slot. Similarly, there is no apparent reason for certain sequences in Turkish to be ungrammatical, except for slot-type mismatches, as in (26):

(26) *Oku-yor-acak-sın.
read-PROG-FUT-AGR
(Intended interpretation: You will be reading.)

The only explanation for the ungrammaticality of (26) is that the progressive and the future markers are typed to occur in the same slot. This suggestion is supported by the grammaticality of (27) where both affixes occur in the first slot after the stem:

(27) Oku-yor ol-acak-sın
read-PROG AUX-FUT-AGR
You will be reading.

6. Implications for syntax

Certain aspects of the proposal regarding morphological templates are relevant to the structure of syntax. Firstly, the present proposal assigns inflection to the morphological component. This is not, in itself, at variance with having a syntactic reflex for each inflectional affix. Morphological templates and slot-type matching requirements could determine certain aspects of inflectional morphology, and other aspects could be represented in syntax. The question is, of course, whether one component replicates the effects of the another component.

In current analyses in syntax, inflectional affixes are posited as heads of functional projections. Following Pollock (1989) the earlier version of this research programme involves generating each functional category as a head. Words are then formed in the syntax as a result of head movement. In the Minimalist Programme (Chomsky 1993) words are formed in the lexicon but they still have to go through the process of head movement for feature checking. In both versions, the order of the morphemes is determined by the selection properties of heads.

There is no a priori reason to object to such a design. In fact, it is even desirable to have a maximal projection for each functional head for its specifier and complement positions, if it can be shown that the presence of these positions account for the facts. In the literature there are a number of convincing arguments in favour of positing each functional category as a syntactic head (Pollock 1989, Ouhalla 1991, Bobaljik and Jonas

1996, among others). As is well known, one of the motivations behind the proliferation of functional projections within inflection is the contrast between word order variations in French and English, which is argued to be an outcome of a difference in the occurrence of adverbs in different specifier positions (Pollock 1989). Ouhalla (1991) derives the difference between VSO and SVO languages from the difference between the ordering of tense and agreement morphology, and crucially, from the specifier positions that are available because of the status of these morphemes as syntactic heads.

As for Turkish, there is compelling evidence regarding specificity that there are two specifier positions, one related to VP, the other to IP (Kennelly 1993). However, it has to be seen whether the proliferation of functional categories serves a purpose in Turkish. To my knowledge, there is no evidence that Turkish requires the specifier positions associated with other inflectional elements.[3] The role of the availability of various specifier positions in determining word order, as summarised above, does not apply to Turkish because these arguments are based on the interference of the position of the specifier with the path of head movement, both of which are on the left (for the languages mentioned in Pollock 1989 and Ouhalla 1991). As Turkish is an SOV language, it has rightward head movement, whereas specifier positions are on the left. Therefore arguments based on the interference of specifier positions with head movement are irrelevant for Turkish.[4] If the structure of a morphologically complex word is derivable by mechanisms specific to the morphological component, there is no reason to represent the complexity of the same word in syntax, unless there are independent reasons for doing so.

Conclusion

Certain differences in the morphology of Turkish and Yakut can be explained in terms of morphological templates which determine the length of a word. The data indicates that Turkish has three slots in its inflectional domain and Yakut has two. Language internal and cross-linguistic variations in verbal and nominal morphology which are a consequence of this constraint are summarised below:

[3]It has recently come to my attention that Tosun (forthcoming) suggests that scope properties of certain adverbs might require an analysis where they are generated in the specifier of TP. At the time of writing this paper there wasn't conclusive evidence for this suggestion.

[4] It might be tempting to argue, along the lines of the antisymmetry of syntax hypothesis (Kayne 1994) that Turkish is also left-headed, ie. SVO, in which case it would have to be shown that the arguments above apply to Turkish as well. Kelepir (1996) has shown that positing an underlying SVO structure for Turkish creates insurmountable problems for the representation of clause structure.

(28)	*Turkish*	*Yakut*
Main verb	V...T/A-T/A-AGR	V...T/A-AGR
Embedded verb	V...T/A-C-AGR	V...T/A
Embedded clause	V...T/A-C-AGR N	V...T/A N-AGR
Buffer stems/		
Auxiliaries	V...T/A V-T/A-(T/A)-AGR	V...T/A V-T/A-AGR
	V...T/A V-T/A- C -AGR	V...T/A-AGR V-T/A
Complementiser	✓	×
Double Possessives	V...T/A-C-AGRx N-AGRy	×
Structure	V..... X^1 X^2 X^3	V..... X^1 X^2

The hypothesis regarding morphological templates is a morpheme based approach which supports the presence of an independent morphological component. Although certain aspects of morphology overlap with phonology and syntax, and although certain properties of grammar may apply to all components, a morphological template is definable only in morphological terms. It does not seem to be reducible to phonological structure since morphemes which have one or more syllables behave similarly with respect to morphological slots. It is not reducible to syntax either, as there seem to be purely morphological reasons for some inflectional phenomena to be as they are.

One of the outcomes of this proposal is that it provides a theoretical explanation for morphological typology. If this proposal is correct, the descripton of the morphological structure of languages as isolating, synthetic and agglutinating can be recast partly in terms of the morphological space they have.

References

Anderson, S. (1992) *A-morphous Morphology*. Cambridge: CUP.

Bach, E. (1996) "On the Grammar of Complex Words", in A.-M. Di Sciullo (ed.) *Configurations; Essays on Structure and Interpretation*. Somerville: Cascadilla Press.

Bobaljik, J.D. & D. Jonas (1996) "Subject Positions and the Roles of TP." *Linguistic Inquiry* 27/2.

Bresnan, J. and S. Mchombo (1995) "The Lexical Integrity Principle: Evidence from Bantu." *Natural Language and Linguistic Theory* 13/2.

Chomsky, N. (1993) "A Minimalist Program for Linguistic Theory", in K. Hale and J. Keyser (eds.) *The View from Building 20*. Cambridge, Mass.: MIT Press.

Di Sciullo, A.-M. (1996) "Atomicity and Relatedness in Configurational Morphology", in A.-M. Di Sciullo (ed.) *Configurations; Essays on Structure and Interpretation*. Somerville: Cascadilla Press.

Di Sciullo, A.-M. and Williams, E. (1987) *On the Definition of Word*. Linguistic Inquiry Monograph 14. Cambridge: MIT Press.

Göksel, A. (1997) "Morphological Asymmetries between Turkish and Yakut", in K. İmer and N.E. Uzun (eds) *Proceedings of the VIIIth International Conference on Turkish Linguistics*. Ankara: Ankara Üniversitesi Basımevi.

Kayne, R. (1994) *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
Kelepir, M. (1996) The Implications of the Antisymetry Theory of Syntax on Turkish Word Order. Unpublished MA disertation, Bosphorus University.
Kennelly, S. (1992) "Turkish Subordination: /-CP, -Tense, +Case/. Talk presented at the 6th Conference on Turkish Linguistics, Eskişehir.
Kennelly, S. (1993) "Non-specificity in Turkish: Inherent Case Revisited". Ms.
Kural, M. (1993) "V-TO(I-TO)-C in Turkish." *UCLA Occasional Papers in Linguistics*, 11.
Ouhalla, J. (1991) *Functional Categories and Parametric Variation*. London: Routledge.
Özsoy, S. (1988) "On Complementaton in Turkish," in Koç, S. (ed.), *Studies on Turkish Linguistics*. Ankara: Middle East Technical University Press.
Pollock, Y.-J. (1989) "Verb Movement, Universal Grammar, and the Structure of IP." *Linguistic Inquiry* 20/3.
Sells, P. (1995) "Korean and Japanese Morphology from a Lexical Perspective". *Linguistic Inquiry* 26/3.
Tosun, G. (forthcoming) Functional Projections in Turkish.

This research has been funded by the School of Oriental and African Studies (grant XG15) and the British Academy (grant BA-AN1438/APN1654). I am grateful to Lena Lukchina, Vassily Klimovich Pavlov, Vladislav Terekov and Svetlana Yegorova who worked as Yakut informants in the project.

Marianne Mithun
University of California, Santa Barbara
mithun@humanitas.ucsb.edu

THE STATUS OF TENSE WITHIN INFLECTION

Tense is frequently cited as a prototypical example of inflectional morphology. In recent work by Anderson (1992) and by Booij (1994, 1996) inflection has been subdivided into types, four by Anderson (configurational, agreement, phrasal, and inherent) and two by Booij (contextual versus inherent). In general, Anderson's first three types are subsumed under Booij's first. European tense systems are classified in different ways, however, as contextual by Anderson but as inherent by Booij. Here it is shown that Booij's system predicts the otherwise unexpected constellation of characteristics of a tense system of a quite different kind, that of Central Alaskan Yup'ik Eskimo. Though the markers can be seen as paradigmatic and obligatory, they also interact in interesting ways with the elaborate derivational morphology.

One of the most frequently cited examples of a prototypical inflectional category is tense.* Yet the motivation for classifying tense as inflectional varies widely according to the criteria proposed to delineate inflection from derivation. Inflection has often been taken as a cluster concept composed of characteristics such as those detailed in Bauer 1983, Scalise 1988 and Plank 1994. Categories may thus be inflectional to varying degrees, depending on the number of pertinent characteristics they exhibit. Sometimes a single characteristic has been seen as criterial, such as obligatoriness or relevance to the syntax. The definition of syntactic relevance must of course depend in turn on the particular view of syntax assumed.

The status accorded tense within the domain of inflection has varied as well. Anderson (1992: 82-3) distinguishes four types of inflection:

a) configurational (case)
b) agreement (number concord on English verbs)
c) phrasal (genitive on English noun phrases, tense on verbs)
d) inherent (gender on Latin nouns).

Booij (1994, 1996) distinguishes just two:

a) contextual (number agreement on Dutch verbs)
b) inherent (number on Dutch nouns, tense on verbs)

Anderson's first three types, configurational, agreement, and phrasal (a-c), are subsumed under Booij's contextual type (a). The types they label inherent are essentially the same

*Mark Aronoff and Paul Kiparsky made helpful comments on several points discussed here. Work on Yup'ik was made possible by grants from the Academic Senate, University of California, Santa Barbara. Abbreviations are detailed in the Appendix.

(Booij 1994:28). Tense is accorded different positions within the two schemas, however. Anderson classifies tense as phrasal (c) because it is a property that is `assigned to a larger constituent within a structure' (the clause) but `realized on individual words' (verbs). Booij concurs that tense has scope over a whole clause, but classifies it as inherent, because `the tense of the verb is not determined by syntactic structure' (1994:30).

A significant feature of inherent inflection noted by Booij is the fact that it can interact with derivation, an observation that argues against split models of morphology. Booij's model also allows a more specific formulation of the nature of the boundary between inflection and derivation. Contextual inflection, defined as `that kind of inflection that is dictated by syntax' (1996:2), differs cleanly from derivation, while inherent inflection may differ from derivation to varying degrees.

For many languages, the various criteria for identifying inflection yield the same categorization of tense. For some, however, they do not, providing us a better vantage point from which to compare their utility. Such a situation will be illustrated here with material from Central Alaskan Yup'ik, a language of the Eskimo-Aleut family. It will be shown that Booij's schema accounts well for the sometimes surprising patterning of tense markers synchronically and diachronically in the language.

1. Tense in Central Alaskan Yup'ik

At first glance, Yup'ik appears to exhibit a regular paradigmatic inflectional tense system similar to those of many European languages. Examples are drawn here from the speech of the Charles family of Bethel Alaska, particularly Nick Charles, Elena Charles, George Charles, Elizabeth Charles Ali, and John Charles. (Additional descriptions of the system are in Mithun 1995, in press, and to appear, and Snyder 1996.) I am especially grateful to Elizabeth Ali and George Charles for their help in transcribing and discussing the material.

(1) Basic tense suffixes

ayagtua	*ayallruunga*
ayag-tu-a	*ayag-llru-u-nga*
go-IND.INTR-1SG	go-PAST-IND.INTR-1SG
`I'm going'	`I went'
ayakatartua	*ayaciqua*
ayag-qatar-tu-a	*ayag-ciqe-u-a*
go-IMMINENT.FUT-IND.INTR-1SG	go-FUT-IND.INTR-1SG
`I'm going to go'	`I'll go'

In much spontaneous speech, the use of the suffixes appears quite straightforward, essentially the same as in English. Present tense verbs are unmarked for tense, while those referring to past events carry the past suffix *-llru-* and those referring to future events carry the imminent future *-qatar-* `going to' or the general future *-ciqe-* `will'.

(2) Use of tense suffixes in conversation: Elizabeth Ali, speaker

Wiinga	*tang*	*kaikapailrianga.*
wiinga	*tang*	*kaig-qapiar-lria-nga*
I	see	be.hungry-very-PARTICIPIAL-1SG

`You see, I'm very hungry.

Atsalurpainek	*kiimek*	*tuai*	*nerellruunga.*
atsar-lugpiar-nek	*kii-mek*	*tuai*	*nere-llru-u-nga*
berry-authentic-ABL.PL	only-ABL	that.is	eat-PAST-IND.INTR-1SG

I only ate salmonberries.

Paluqatartua.
palu-qatar-tu-a
starve-IMMINENT.FUTURE-IND.INTR-1SG
I'm going to starve.

Carrakuinermek	*tauggaam*	*cikiquvnga*	*tuai*
carrar-kuiner-mek	*tauggaam*	*cikir-ku-vnga*	*tuai*
little.bit-small.amount-ABL	but	give-COND-2SG/1SG	well

But if you give me just a little bit,

quyapaircigua.
quya-pair-ciqe-u-a.
be.thankful-very-FUTURE-IND.INTR-1SG
I will be most grateful.'

On many occasions, however, it seems almost haphazard. Verbs relating past events often lack past tense suffixes. The passage in (3) came from a breakfast table conversation. Mrs. Charles, the mother of the family, is a gifted Yup'ik speaker.

(3) Apparent optionality: Elena Charles, speaker

Last fall-gguq	*maaten-gguq*
last fall=gguq	*maaten=gguq*
last fall=HEARSAY	when=HEARSAY
`Last fall when	

Frankynguk	*tekituk*
Franky-ngu-k	*tekite-u-k*
Franky-ASSOC-DU	arrive-IND.INTR-3.DU

Franky and his companion arrived (no tense)

campaput	*yungqellruyaaqelliniuq*
campaq-aput	*yuk-ngqerr-llru-yaaqe-llini-u-q*
camp-1PL/3SG	person-have-PAST-actually-apparently-IND.INTR-3SG

they realized that there had been (PAST) people at our camp.

upallrullinilutenq
upag-llru-llini-lu-teng
change.residence-PAST-apparently-SUB-3PL
They had moved (PAST)

carayiim *piateng.*
carayag-m *pi-a-ateng*
bear-ERGATIVE do-CONSEQUENTIAL-3SG/R.PL
because a bear was bothering them (no tense).

Franky-gguq *bother-neritellinilutek.*
Franky=gguq *bother-nrite-llini-lu-tek*
Franky=HEARSAY bother-not-apparently-SUB-3DU
But Franky said that they (he and companion) were not bothered (no tense).

The Yup'ik tense markers are not absolute, as in English, but relative. In English, the deictic center of the tense system is generally the moment of speech. Past tense markers indicate a time before the moment of speech, and future tense markers a time after it. In Yup'ik the deictic center may be the moment of speech, as in (2) above, but within narrative, even short anecdotes, the deictic center is the narrative time. Events happening along the timeline of the narrative are unmarked for tense. A past tense marker specifies a time before the current narrative moment, and a future tense marker a time after it. The reference tlme in (3) is Franky's visit to the camp. The past tense suffixes on `there had been people' and `they had moved' specify a time before his visit, before the narrative moment. The clause stating that Franky and his friend were not bothered by a bear is unmarked for tense, because it is simultaneous with the visit.

A relative future can be seen in (4). Most of the events related in the narrative took place along the narrative timeline, so they are unmarked for tense. The final line, however, `I would be squashed', projects an event after the narrative moment.

(4) Relative future: Elena Charles, speaker
[Those two moose there were looking at me.]

Wiinga-gg *tangvagkegka*
wiinga=gga *tangvag-ke-gka*
I=as.for watch-PARTICIPIAL-1SG/3DU
As for me, I was watching them (no tense).

Tuai *tuntuviik* *ukuk* *taingareskaggnek*
tuai *tuntivag-ek* *uku-k* *tai-ngarte-ku-agnek*
and moose-DU these.approach-DU come-suddenly-COND-3.DU
And if these two moose came suddenly (no tense) ...

tuai *yaavet* *qercigua*
tuai *yaavet* *qerte-ciqe-u-a*
so to.yonder squashed-FUTURE-IND.INTR
then I would be squashed (FUTURE).'

The pattern is reminiscent of the historical present in English. In Yup'ik however, narrative time must be assumed as the point of reference, because the tense system is a relative one. It is a matter of grammar. In English, the historical present is simply a stylistic option, a possible exploitation of an absolute tense system for stylistic effect. Of

course one may evolve from the other diachronically.

A shift in the reference point to narrative time is not necessarily signalled formally in Yup'ik. There need not be an overt past tense marker to shift it away from the speech time. Often, of course, time is established at the beginning of a narrative with adverbials or lengthier explanations. The passage in (4) above, for example, opened with `Last fall'. In the same way, shifts in the point of reference back to the moment of speech are not necessarily marked.

Even in the course of telling narratives, however, speakers do not always maintain a single point of temporal reference. Speakers often step out of the narrative world to add comments from their present vantage point. Such a shift can be seen below. As the narrative unfolded, there was no past tense marking. After a pause and a murmur from the audience, she made the statement in (5), this time with a past tense suffix.

(5) Shift in perspective: Elena Charles, speaker
`We went again (no tense) to see (no tense) Qitenguq. You see, we could not catch game (no tense). And those two accompanied us (no tense), those two from up there, Peter Aluska and another, travelling (no tense) with their own boat ... and Bob Qilang.' (Mmm).

Yunerillruuq *tauna.*
yunerir-llru-u-q *tauna*
die-PAST-IND.INTR-3SG that
`He died (PAST TENSE), that one.

Ayiimek *tuai* *mululutek.* ...
ayag-a-mek *tuai* *mulu-lu-tek*
go-CONSEQUENTIAL-3R.DU s o be.late-SUB-3.DU
The two left late [but at least they arrived, and the weather was good ...]'

The dying clearly did not take place prior to the narrative time, when the two men were coming along in their boat. Mrs. Charles left the narrative time to mention the death of Mr. Qilang; the past tense on `he died' situates his death prior to the moment of speaking, not the narrative. She then resumed the story with no signal.

Discussions of past habitual events generally exhibit systematic past tense marking on each clause. This is because there is no narrative timeline, no sequence of events.

(6) Past habituals: Elena Charles, speaker
Ayagllermegni *nunanirqelallruuq.*
ayag-ller-megni *nunanirqe-la-llru-u-q.*
go-CONTEMPORATIVE.PAST-1DU pleasant-HABITUAL-PAST-IND.INTR-3SG
`When we travelled, it used to be beautiful (PAST HABITUAL).

Ayakarrarrlemegni *qamani*
ayag-qarraar-ller-megni *qama-ni*
go-at.first-CONTEMPORATIVE.PAST-1DU upriver-LOC
When we first travelled in there

uitalallruukuk *qaivani* *Itiuliggegmi.*
uita-la-llru-u-kuk *qaiva-ni* *Itiuliggeg-mi*
stay-HAB-PAST-IND.INTR-1PL upriver-LOC Ituli-LOC
we would stay (PAST HABITUAL) far in there at Ituli.

Allaneq-am *ikanitengnaqlallruuq* *qikertarraremi*
allaner=am *ikani-te-ngnaqe-la-llru-uq* *qikertar-rrar-mi*
stranger=EMPH across.there-go.to-try-HAB-PAST-IND.INTR island-little-LOC
A stranger used to try to stay over there (PAST HABITUAL) on a little island.

Wiinga-am *tauna* *assikngamku*
wiinga=am *tauna* *assike-nga-mku*
I=EMPH that like-CONSEQUENTIAL-1SG/3SG
Because liked that place

tuantelallruukuk *kiigamegnuk*
tuan-te-la-llru-u-kuk *kiiga-megnuk*
there-go.to-HABITUAL-PAST-IND.INTR-1DU lone-1DU
we used to stay there (PAST HABITUAL) by ourselves.'

The clauses in this passage represent related comments around a theme, rather than the progression of a sequence of events.

Once the relative nature of the Yup'ik tense system is understood, it appears straightforwardly inflectional, according to most of the usual criteria for inflection. The tense suffixes do not appear to create new lexemes: verbs remain verbs. They affect none of the features cited by Scalise (1988:568) as alterable by derivation only: syntactic category, conjugation class, subcategorization features, or selectional features. They are fully productive. They are unconstrained by blocking, by which tense marking on certain stems would be avoided because of the prior existence of synonyms. Their allomorphy is regular, phonologically rather than lexically conditioned. Their semantic contribution is transparent and predictable, as well as abstract. The markers are also inflectional by Plank's criterion of relationality (1994: 1673), `specifying the temporal relation between the proposition and the speech act'. Tense qualifies as inflectional even by the more elusive commutability criterion: there are no monomorphemic stems in Yup'ik that could replace a stem plus tense suffix. Furthermore, there are no independent words that could replace a tense suffix in a clause, though of course there are adverbials that cooccur with tense markers. The classification of the Yup'ik tense suffixes as inflectional is buttressed by the fact that tense is a commonly recurring inflectional category cross-linguistically.

The Yup'ik system does raise interesting questions about a feature often considered definitive for the distinction between inflection and derivation: obligatoriness.

2. Obligatoriness

Among the characteristics of inflectional categories, the feature of obligatoriness has often been taken as criterial. Bybee (1985: 81) remarks, for example:

> One of the most persistent undefinables in morphology is the distinction between derivational and inflectional morphology. While linguists seem to have an intuitive understanding of the distinction, the objective criteria behind this intuition have proved difficult to find. The most successful criterion is *obligatoriness*, applied to the definition of derivation and inflection by Greenberg 1954. Obligatory categories force certain choices upon the speaker.

An analysis of Yup'ik tense as obligatory entails the recognition of a meaningful zero; the lack of a tense suffix must signal time as well, either present or, more precisely, `time simultaneous with the deictic center'. Such a characterization accords with what we have seen of the Yup'ik tense markers so far. Yet further examination of natural Yup'ik speech shows that verbs sometimes occur without tense marking when they represent events not simultaneous with the deictic center. The passage in (7) below describes a narrative sequence, appropriately unmarked for tense: getting up, drinking coffee, going down, stopping, shooting. Yet when Ayaginar spoke, we might have expected a past tense within his utterance.

(7) Unmarked tense: Elena Charles, speaker
`In the morning we woke up and it was raining. We had coffee and those two men, our companions, came up to have coffee too. Then your daddy said to them, "Now over there, to the side of us, dock at the edge of the lake and look to see if there is game." The two left and after some time they suddenly stopped, and they shot their guns. Ayaginar [the father] said:

Cakma *tuai* *tuntuturtuk*
cakma *tuai* *tuntu-tur-tu-k*
down.there.obscured so moose-catch-IND.INTR-3DU
"They must have caught a moose down there (unmarked tense)."'

He was making an observation, not telling a narrative, so we would anticipate that the deictic center for him would be the moment of speech. The sounds of the shots had already faded by the time he spoke. Yet his comment carried no tense marking.

An investigation of the use of such verbs without tense suffixes might suggest that the Yup'ik point of temporal reference covers a larger span of time than its English counterpart. The uses of the different tenses with the *nalkute-* `find' can be compared as follows. Mr. Charles reports that if he and a friend were out looking for a lost knife, and he suddenly spied it, he could use the unmarked (present) tense as he was bending over to pick it up: *nalkutaqa* `I'm finding it'. If he and his friend were some distance apart, so that after picking up the knife he had to make his way over to where the friend was searching, he could still use the same verb several minutes later to announce his good luck. If the two men were far from home and then spent most of the day returning, he could use the same unmarked verb to announce his success to his wife that evening. If his mother had been asleep when he returned, he could even use the unmarked verb to tell her the news the following morning. Mrs. Ali concurred, commenting, "To her, it's still lost until you tell her". Immediately after the announcement, the mother could turn to her own husband and use the past tense: *nalkutellruullinia* `he apparently found it'. The Yup'ik unmarked present tense thus seems appropriate for a span of time encompassing not only the moment of speech, but as long as the preceding day and night as well.

But the difference is more interesting. Scurrying around the kitchen preparing dinner, I

might realize that I have mislaid my knife. Discovering it a few moments later, Mrs. Ali notes that I could use the unmarked *nalkutaqa* just as I caught sight of it. Now if my husband had been on his way out when I began searching for the knife, but he returned 15 minutes later to find me engrossed in a book, having completed dinner preparations, I would use the past tense to announce my discovery: *nalkutellruaqa* `I found it'. This time the Yup'ik unmarked present tense seems to cover a span no longer than 15 minutes.

The unmarked tense category does not of course indicate a specific span of time. It is used to convey immediacy, for what is portrayed as immediate rather than displaced experience. What is included within the realm of immediate experience can vary to a certain extent with the situation and the desire of the speaker. When Ayaginar spoke in (7) above, he was portraying the shooting of the moose as part of the current situation. The same was true of Mr. Charles announcing the discovery of his knife even after a day had passed. English shows a somewhat similar use of the present progressive for imminent futures (*I'm leaving*) but the similarity does not extend to past events. Yup'ik speakers systematically use the unmarked present for past punctual events that have current relevance, as in `I find my knife', `they catch a moose'. In similar situations an English speaker could use a perfect: `I've found my knife', `they've caught a moose'. The fact that the unmarked present covers both past and future events indicates that it is neither a perfect nor an imminent future marker, but simply a marker of immediate relevance.

3. Paradigmaticity

Closely related to the issue of obligatoriness and meaningful zeroes is paradigmaticity. We expect inflectional categories to be expressed by a relatively small set of terms that constitute a closed class, and to be mutually exclusive. There are actually several more Yup'ik tense suffixes than those discussed so far. Among them are *-arkau-* `will eventually', `should', `is supposed to'; *-niar-*`will perhaps' *-ki-* `do later' (delayed imperative), *-ngaite-*, `will not', *-ngite-* `please do not', and *-niarar-*, `be going to soon'. A number of these have resulted from the compounding of adjacent suffixes, such as *-nrilki-* from the negative *-nrite-* + delayed future *-ki-*. The fact that new markers may enter the system is not problematic for a classification of tense as inflectional, however. All grammatical systems evolve over time. What is interesting is the extent to which the creation of a new category affects the system as a whole, since it is purportedly paradigmatic, and markers should be mutually exclusive. The creation of new markers by compounding is not disruptive, since verbs containing reanalyzed sequences still have just one tense marker. The tense markers are not, however, clearly mutually exclusive.

(8) Imminent future + past: George Charles, speaker
ayakatallruunga
ayag-qatar-llru-u-nga
go-IMMINENT.FUTURE-PAST-IND.INTR-1SG
`I was going to go'

(9) Past + future: George Charles, speaker
ayallrucigua
ayag-llru-ciqe-u-a
go-PAST-FUTURE-IND.INTR-1SG
`I will have gone'

Examples (8) and (9) each contain both a past and a future suffix. Semantic scope relations are reflected in the order of the suffixes. The first verb *ayaqatallruunga* `I was going to go' represents an imminent event (inner formation `about to go'), the whole set in past time (outer past suffix *-llru-*). The second verb *ayallruciqua* `I will have gone' represents a past event (inner formation `went') viewed from the future (outer future suffix *-ciqe-*). The existence of such forms does not necessarily constitute evidence against the paradigmaticity of the tense markers, however, if the complexes are analyzed as members of the system in their own right: *-qatallru-* and *-llruciqe-*. It is significant that the alternative orders are not possible: there is no **ayaciqellruunga* (go-FUTURE-PAST-IND.INTR-1SG) and no **ayallruqatartua* (go-PAST-IMMINENT.FUTURE-IND.INTR-1SG).

4. Relevance to the syntax: agreement

A frequently cited criterion for inflection is relevance to the syntax. This criterion has important implications for models of linguistic structure such as that of Anderson, in which inflection is accomplished by syntactic rules rather than by processes localized within a separate morphological component. Booij (1994, 1996) has proposed that not all categories that would be considered inflectional by other criteria are relevant to syntax, and that the distinction can be captured by recognizing two types of inflection, contextual inflection, `that kind of inflection that is dictated by syntax, such as person and number markers on verbs that agree with subjects and/or objects, agreement markers for adjectives, and structural case markers on nouns', and inherent inflection, `the kind of inflection that is not required by the syntactic context, although it may have syntactic relevance' (1996:2). He notes that `inherent inflection is more similar to derivation, and it may feed word formation, unlike contextual inflection, which is peripheral to inherent inflection. Language acquisition and language change also appear to reflect this distinction' (1996:3). For Anderson, tense is relevant to the syntax because it is a property `assigned to a larger constituent within a structure' (the clause) but `realized on individual words'. Booij concurs that tense has scope over a whole clause, but classifies it as inherent, because `the tense of the verb is not determined by syntactic structure' (1994:30).

If we were to find tense agreement, we would have a clear case of contextual inflection. Yup'ik appears to offer just such a system. Tense can be marked on nouns as well as verbs.

(10) Tense on nouns

uillra	*akutarkat*
ui-ller-a	*akutar-kar-t*
husband-PAST-3SG/3SG	Eskimo.ice.cream-FUTURE-PL
`her former husband'	`future Eskimo ice-cream'

Tense suffixes on nouns and verbs can cooccur within a sentence, suggesting the possibility of agreement.

(11) Cooccurrence of noun and verb tense

uillra	*sugtullruuq*
Ui-ller-a	*sugtu-llru-u-q*
husband-PAST-3SG/3SG	tall-PAST-INDICATIVE-3SG
`Her husband was tall.'	

A closer look reveals that the noun and verb suffixes operate in different domains. The verb suffixes situate events in time, while the noun suffixes situate referents. They need not match within a clause.

(12) No agreement: past tense Elizabeth Ali, speaker

Uillra	*sugtuuq*
ui-ller-a	*sugtu-u-q*
husband-PAST-3SG/3SG	tall-INDICATIVE-3SG

`Her former husband is tall.'

(13) No agreement: future tense: Elena Charles, speaker

Qallalluki	*piuratuaput*	*akutarkat*
qallate-lu-ki	*piurar-tu-a-put*	*akutar-kar-t*
boiling-SUB-R/3PL	continue-CUST-IND.TR-1PL/3PL	mixture-FUT-PL

`We bring them a boil, those (fish) that will be made into Eskimo ice cream.'

There is of course correspondence between sentence adverbials and tense. The cooccurrence could be taken as the result of either grammatical constraints or simply the fact that speakers say things that make sense. In any case, there is little formal evidence that Yup'ik tense should be considered contextual in Booij's sense.

5. Interaction with derivation

Booij has proposed that inherent inflection, unlike contextual inflection, can interact with derivation. Here the Yup'ik case becomes especially interesting. Yup'ik contains an unusually rich inventory of suffixes. They include some suffixes that affect argument structure, *-ni-* `claim that', *-yuke-* `think that', and *-nayuke-* `think that maybe'. They introduce a claimer or thinker. If the derived verb is inflected intransitively, it specifies that the person cast as the absolutive thinks something about himself or herself. If it is inflected transitively, it specifies that the person cast as the ergative thinks something about another, cast as the absolutive. The verbs in (14) are derived from *ayag-* `leave', as in *ayagtuq* `he's leaving'.

(14) Derivational suffix *-ni-* `say that': Elizabeth Ali, speaker

Ayagniuq	*Ayagniat*
ayag-ni-u-q	*ayag-ni-a-at*
leave-say-IND.INTR-3SG	leave-say-IND.TR-3PL/3SG
`He says he (himself) is leaving'	`They say he's leaving.'

Tense markers can appear with derived verbs of claiming and thinking. A past tense suffix, for example, may follow the derivational suffix of saying to put the entire claiming event expressed by the derived verb stem in the past, as in (15).

(15) Past claim: Elizabeth Ali, speaker
Ayagnillruat
ayag-ni-llru-u-at
leave-say-PAST-IND.TR-3PL/3SG
`They said he was leaving.'

Tense markers may also precede the derivational suffix of saying. The past tense suffix in (16) puts the event claimed in the past.

(16) Claim about the past: Elizabeth Ali, speaker
Ayallruniat
ayag-llru-ni-a-at
leave-PAST-say-IND.TR-3PL/3SG
`They say he left.'

Tense suffixes may even occur both before and after the derivational suffix of saying.

(17) Past claim about previous event: Elizabeth Ali, speaker
Ayallrunillruat
ayag-llru-ni-llru-a-at
leave-PAST-say-PAST-IND.TR-3PL/3SG
`They said he had left.'

The tense markers can and do interact with the derivational morphology.

The capacity of tensed verbs to serve as the input to derivational processes has consequences for related features considered characteristic of inflection. The tense markers are not always `outer' affixes, occurring at the margins of words. As we can see from examples (16) and (17), tense affixes can appear inside of derivational affixes, closer to the root. The tense suffixes could also be said not to have an invariant order with respect to other suffixes; as seen above, they occur sometimes before and sometimes after the derivational suffix *-ni-*. They could even be said to apply recursively, that is, to their own output, with the mediation of suffixes like *-ni-*.

The situation is actually just what would be predicted by Booij's scenario. Yup'ik tense would be classified as inherent derivation.

> Inherent inflection is the kind of inflection that is not required by the syntactic context, although it may have syntactic relevance. Examples are the category number for nouns, comparative and superlative degree of the adjective, and tense and aspect for verbs ... Inherent inflection is more similar to derivation, and it may feed word formation, unlike contextual inflection, which is peripheral to inherent inflection. Language acquisition and language change also appear to reflect this distinction. (Booij 1996:2-3)

Booij notes further that `contextual inflection tends to be peripheral with respect to inherent inflection' (1996:11). All nouns and verbs in Yup'ik consist of a base (root), any number of optional postbases, and one and only one obligatory ending. On nouns, the ending specifies number and case. If the noun is possessed, the ending encodes the possessor and possessed in a transitive pronominal suffix. On verbs, the ending consists of two parts: a mood marker and a pronominal suffix complex.

The endings would qualify as inflectional suffixes by any criteria. They are obligatory and paradigmatic, they comprise a closed set, they are fully productive and applicable to all stems, they show only phonologically conditioned allomorphy, and they contribute predictable meanings. Their order is invariant, and they are not recursive. They do not

feed derivation. They would generally be considered contextual inflection. On nouns they specify case and number in portmanteau forms, and case is obviously highly contextual syntactically. On verbs, the mood suffixes function relate clauses to the larger discourse event (indicative, interrogative, optative) or to each other (participial, subordinative, connectives). The pronominal suffixes specify the core argument of the clause. Yup'ik morphology thus shows a structure perfectly in accord with Booij's division of inflection into inherent and contextual. The contextual categories in Yup'ik are always word-final, and the inherent categories, particularly tense, are word-internal.

6. The shifting of categories over time

Yup'ik also shows us that morphological categories do not necessarily occupy fixed positions between derivation and inflection. The past tense suffix *-llru-* is not reconstructed for Proto-Eskimo (Fortescue, Jacobson, and Kaplan 1995). Jacobson 1984 derives it from a compounding of the nominal past tense suffix *-ller-* plus the verbalizing suffix *-u-* `be'. The suffix *-ller-* can be attached to either noun stems or verbs stems, but it always derives a noun stem: `former N', `the one that Ved'. It is thus a past nominalizer, always including a specification of past tense. (Uvular *r* automatically appears as the stop *q* word-finally.)

(18) Historic elements of past *-llru-*: *-ller-* + *-u-* Jacobson 1984: 491, 488,

angyara	`his boat'	*ayag-*	`to leave'
angyallra	`his former boat'	*ayalleq*	`the one who left'
angyaq	`boat'		
angyauguq	`it is a boat'		

The derivational leanings of the modern past tense suffix *-llru-* might be explicable in part as relics of its earlier source, literally `to be the one that V-ed'.

Markers may apparently slide along the continuum between derivation and inflection in either direction. The suffix *-ller-* also appears as an etymological element in a number of other suffixes, some highly derivational. It has been compounded with the suffix *-(ng)un-* `supply of', for example, to yield a new suffix *-nguteller-* `empty container which held N'.

(19) Element of new derivation: Jacobson 1984: 583

ciku	`ice'
cikuutelleq	`empty container which held ice'

Of special interest is the separate evolution of the past tense nominalizer *-ller-* into a modern inflectional suffix (ending), the past contemporative mood *-ller-* `when (in the past)'. Its use can be seen in example (6), repeated here in part.

(6) Past contemporative *-ller-* `when (in the past)': Elena Charles, speaker

Ayagllermegni	*nunanirqelallruuq.*
ayag-ller-megni	*nunanirqe-la-llru-u-q.*
go-CONTEMPORATIVE.PAST-1DU	be.pleasant-HABITUAL-PAST-IND.INTR-3SG

`When we travelled, it used to be beautiful.

The mechanism by which the derivational past nominalizer was reinterpreted as an inflectional connective mood is clear. As a nominalizer, it formed nouns that could be inflected for number, case, and possession, just like other nouns. The intransitive pronominal suffixes that appear with the modern inflectional past contemporative mood resemble those that appear with locative endings on verbs.

(20) Past contemporative mood: Elizabeth Ali, speaker

tangvagkai	*ayallratni.*
tangvag-ke-ai	*ayag-ller-atni*
watch-PARTICIPIAL-3SG/3PL	leave-PAST.CONTEMPORATIVE-3PL

`He watched them as they were leaving (in their leaving).'

(Compare *angyaatni* `in their boat'.) Contemporatives like *ayallratni* `as they were leaving' are no longer nouns, however. A nominal identifying those leaving is in the absolutive case (`leave' is intransitive), rather than the ergative (genitive) case.

(21) Past contemporative with absolutive noun: Elizabeth Ali, speaker

angun	*ayallrani*
angun	*ayag-ller-ani*
man.ABSOLUTIVE	leave-PAST.CONTEMPORATIVE-3SG

`as the man (ABSOLUTIVE) was leaving'

(Compare *angute-m angyaani* `in the man's (ERGATIVE) boat.') In transitive verbs, the traces of the nominal source are disappearing. The past contemporative mood is usually (though not always) followed by the same verbal transitive pronominal suffixes that appear with other connective moods.

7. Conclusion

Yup'ik tense marking provides us with an example of a system that would be perplexing for traditional accounts of inflection, but that is predicted by the proposal by Booij (1994, 1996) for separating contextual from inherent inflection. Once the relative nature of the system is understood, and it is seen that speakers exploit the unmarked present to convey a sense of immediacy, the system shows most marks of prototypical inflection. Tense suffixes do not create new lexemes: verbs remain verbs with essentially the same meanings, and syntactic category, conjugation class, subcategorization features, and selectional features remain intact. The markers are fully productive, and their semantic contributions are transparent and predictable. They are obligatory and paradigmatic. On the other hand, the tense suffixes can interact with derivation. This is just the constellation of features proposed by Booij to characterize inherent inflection. At the same time, a closer look at the shallow history of the suffixes themselves reminds us that the position of markers along a continuum from derivation to inflection is not necessarily fixed for all time. The suffix *-ller-* has been seen to evolve in several directions, from derivational to more derivational, to inherent inflection, and to contextual inflection.

Appendix

The transcription used here is in the practical orthography developed by the Alaska Native Language Center in Fairbanks. Abbreviations in glosses are as follows: ABL=ABLATIVE, ABS=ABSOLUTIVE, ASSOC=ASSOCIATIVE, DU=DUAL, CNTP=CONTEMPORATIVE, COND=CONDITIONAL, EMPH=EMPHATIC, FUT=FUTURE, HAB=HABITUAL, IND=INDICATIVE, INTR=INTRANSITIVE, LOC=LOCATIVE, PL=PLURAL, R=COREFERENTIAL, SG=SINGULAR, SUB=SUBORDINATIVE, TR=TRANSITIVE.

References

Anderson, Stephen 1992. *A-Morphous Morphology*. Cambridge.

Bauer, Laurie 1983. *English Word-formation*. Cambridge, Mass: MIT.

Booij, Geert 1994. Against split morphology. *Yearbook of Morphology 1993*. 27-49.

- 1996. Inherent versus contextual inflection and the split morphology hypothesis. *Yearbook of Morphology 1995*. Geert Booij & Jaap van Marle, eds. Dordrecht: Kluwer. 1-15.

Bybee, Joan 1985. *Morphology*. Amsterdam: John Benjamins.

Fortescue, Michael, Steven Jacobson, and Lawrence Kaplan 1995. *Comparative Eskimo Dictionary*. Fairbanks: Alaska Native Language Center.

Greenberg, Joseph. 1954. A quantitative approach to the morphological typology of language. *IJAL* 26: 178-194.

Jacobson, Steven 1984. *Yup'ik Eskimo Dictionary*. Alaska Native Language Center.

Mithun, Marianne 1995. The codification of time on the Pacific Rim of North America. Symposium on Time and Language, National Museum of Ethnology. Osaka, Japan.

- in press. The codification of time on the Pacific Rim of North America. *Time and Language*. Yasuhiko Nagano, ed. Osaka, Japan: National Museum of Ethnology.

- to appear. *The Languages of Native North America*. Cambridge.

Plank, Frans 1994. Inflection and derivation. *Encyclopedia of Language and Linguistics*. 3:1671-1678.

Scalise, Sergio 1988. Inflection and derivation. *Linguistics* 22:561-581.

Snyder, Jill 1996. A discourse-based analysis of tense in Central Alaskan Yupík. *Prosody, Grammar, and Discourse in Central Alaskan Yupík*. Marianne Mithun, ed. Santa Barbara Papers in Linguistics 7:55-63.

Irene Philippaki-Warburton & Vassilios Spyropoulos
The University of Reading
llswarbn@reading.ac.uk & llr96vs@reading.ac.uk

ON THE BOUNDARIES OF INFLECTION AND SYNTAX

Abstract

This paper examines the status of object clitic pronouns and preverbal particles in Greek. Evidence, both empirical and theoretical is presented, which shows that these elements exhibit two types of properties. In some ways they behave as independent syntactic units, but in others (both phonologically and syntactically) they seem attached as affixes to the grammatical word which they modify. We propose to capture this intermediate character of clitics and particles by treating them on the one hand as independent words in the Lexicon and, on the other, by having them undergo a merging operation, namely Move-Incorporate, within syntax. This rule brings them together with their host grammatical word, to form a new type of unit, which we call syntactic word, following Di Sciullo and Williams (1987).

1. Introduction

The precise definition of "word" becomes problematic when we consider the status of elements such as clitics and particles which behave like affixes in some ways but also like full words in others. We will try to show that the intermediate character of these elements is due to the fact that on the one hand they exist as full words stored in the lexicon, but on the other, during the derivation, they combine with other full *grammatical words* creating a new syntactic unit which we call *syntactic word* following Di Sciullo and Williams (1987). In other words we will argue that clitics and particles start as separate and independent lexical entries but end up as affixes in the syntactic component. We will try to support this position by focusing on the following specific questions:

a) What is the status of object clitic pronouns in Greek? Note that these elements cannot be hosted within a functional head since they do not have a grammatical function and yet in some ways behave like affixes (see Zwicky 1985).

b) What is the status of the preverbal particles in Greek which express grammatical information similar to that expressed by bound morphemes?

2. The functional categories of the Greek verb

Some significant facts about Greek verbs

i) Greek is a null-subject language with rich person/number subject agreement (e.g. see the present tense of the verb *γrafo* 'I write' in (1):

(1) sing. γ raf-o (1st), γ raf-is (2nd), γ raf-i (3rd),
plur. γ raf-ume (1st), γ raf-ete (2nd), γ raf-un (3rd)

ii) There is no infinitive; the only non-finite forms are the *gerund* (2) and the invariable dependent form, namely the *non-finite* (3), which is only found preceded by the auxiliary ***exo*** 'I have' (4):

(2) γ rafondas 'writing'
(3) -γ rapsi
(4) exo γ rapsi 'I have written'

iii) The imperative, which has its own inflectional endings, divides the Greek linguists into those who interpret it as a finite form (Philippaki-Warburton 1994a, 1996) and those who consider it non-finite (Joseph 1985, Horrocks 1990). The imperative is like the non-finite gerund as far as the order of the object clitic pronouns is concerned: unlike the other verb forms, imperative and gerund precede the clitics.

Clauses may contain a monolectic verb (5) or a periphrastic one preceded by the auxiliary ***exo*** (6):

(5)	ACTIVE		MIDDLE/PASSIVE	
	imperfective	*perfective*	*imperfective*	
perfective				
non-past:	γ rafo	γ rapso	γ rafome	γ rafto
past:	eγ rafa	eγ rapsa	γ rafomuna	γ raftika

(6)	*perfect*	
non-past:	exo γ rapsi	exo γ rafti
past:	ixa γ rapsi	ixa γ rafti

The syntactic structure (the configuration of the functional categories involved in the derivation) of the monolectic verb forms is the one represented in (7) and of the periphrastic ones the one in (8). The derivation involves the operation of head movement of the V head to the relevant functional categories in order to check out the inflectional features which it carries as it enters the syntactic derivation fully inflected from the lexical/morphological component (Chomsky 1995). The order of the functional categories reflects the way the morphological exponenets are arranged. INFL represents the fused agreement and tense features.

(7)	INFL		VOICE	ASPECT	VP[...V...]
(8)	INFL	***exo***	VOICE	ASPECT	VP[...V...]

The auxuliary ***exo*** expresses perfect (Aspect/Tense) and thus in its content it is a verbal functional category. However, this functional role of ***exo*** cannot lead to an analysis which treats the periphrastic perfect constructions as single grammatical words, members of the verb paradigm, which would enter the Syntax from the Lexicon, because the two elements of the periphrasis often appear syntactically separate (9-10):

(9)a. I Maria me exi **poles fores** stenoxorisi
Mary me has many times upset
'Mary has upset me many times'
(10)a. Exi sinantisi o Nikos ton adelfo su?
Has met Nick your brother 'Has Nick met your brother?'
b. Exi **o Nikos** sinantisi ton adelfo su?
Has Nick met your brother

Thus, ***exo*** is a separate lexical entry of the category V and of the subcategory Aux, selecting the *non-finite* main verb form, but as an auxiliary it is not a core representative of the lexical category verb because it does not get the full complement of grammatical properties in that it has no theta-grid. So, the syntactic representation consists of two lexical heads a higher V (the auxiliary) and a dependent lower V (the main uninflected verb) neither of which has the full complement of verb properties.

3. Object clitic pronouns

The non-imperative verb whether monolectic or periphrastic may be preceded by one or two object clitic pronouns (indirect (io) - direct (do)).

(11)a. to eγ rapsa
it I-wrote 'I wrote it'
b. tu eγ rapsa
to-him I-wrote 'I wrote to him'
c. tu to eγ rapsa
to-him it I-wrote 'I wrote it to him'
(12)a. to exo γ rapsi
it I-have written 'I have written it'
b. tu exo γ rapsi
to-him I-have written 'I have written to him'
c. tu to exo γ rapsi
to-him it I-have written 'I have written it to him'

Thus, the order is:

(13) io.cl do.cl INFL ***exo*** VOICE ASPECT VP[...V...]

Such constructions present us with the question whether clitics are affixes (Joseph 1988), licensing a *pro* in the argument position, or independent syntactic units (Philippaki-Warburton 1977, 1987), generated in the object argument position and moving to adjoin somewhere in the syntactic configuration, leaving a trace behind.

In what follows we will advance syntactic, morphological and phonological arguments, both empirical and theoretical, which support the analysis of object clitic pronouns in Greek as syntactically separate units. Furthermore, we propose that this analysis holds for both situations, i.e. when only the clitic pronouns are present in the construction but also when we have both clitic pronouns as well as lexical object DPs (clitic left dislocation and clitic doubling constructions). In these cases the clitic is still viewed as the argument proper, while the corresponding lexical DP is interpreted as an adjunct coindexed with the clitic and providing either a topic (clitic left dislocation) or some sort of apposition (clitic doubling constructions).

3.1. Syntactic evidence

i) Object clitics in Greek are not agreement markers because they are optional elements.

ii) Object clitics when present constitute the object arguments as shown by the fact that lexical object DPs when co-occurring with the clitics are not arguments, since they cannot receive the main stress of the sentence (*14d).

(14)a. Xθes ɐɣ orasa **to kenurjo vivlio tu Chomsky**
Yesterday I bought **the new book by Chomsky**
b. **Xθes** to ɐɣ orasa to kenurjo vivlio tu Chomsky
It was **yesterday** that I bougt the new book by Chomsky
c. To vivlio tu Chomsky to ɐɣ orasa **xθes**
The book by Chomsky I bought it **yesterday**
d. * Xθes to ɐɣ orasa **to kenurjo vivlio tu Chomsky**

iii) Related to the above is the fact that in the presence of its clitic a lexical object DP cannot normally undergo wh-movement in a single clause. Compare (15a) with (15b):

(15)a. Pjo vivlio ɐɣ orases?
Which book did you buy?
b. * Pjo vivlio to ɐɣ orases?

Given that wh-movement leaves a coindexed trace with which the moved wh-item forms a chain the irregularity of (15b) follows from the fact that under our analysis the trace following the verb is derived by the movement of the pronominal clitic, as shown in (16a):

(16)a. to$_j$ ɐɣ orases t$_j$ pjo vivlio
b. * pjo vivlio$_j$ to$_j$ ɐɣ orases t$_j$

In (*16b) the trace forms a chain with the clitic and for this reason it cannot also form a chain with the wh-element. Nor can we have two crossing chains with one included within the other.

We must note however that, although the constructions as in (15a) are the normal and most frequent ones, while (15b) are ungrammatical, the latter may become more acceptable if the sentence is extended with, for example, some adverbial as in (17).

(17)a. ? Pjo vivlio to ɐɣ orases xθes?
b. pjo vilio$_i$ to$_i$ ɐɣ orases t$_i$ xθes

Such evidence may seem to undermine our analysis. However, this evidence is rather weak because constructions as those in (17) are marked and rare and more significantly because the wh-constituent in such constructions is not straightforwardly questioned but has a topic reading. It is possible, therefore, to argue that (17b) is a construction where the wh-DP is a left dislocation construction analogous to:

(18) To Jani ton iða xθes
John him I-saw yesterday 'As for John, I saw him yesterday'

where the dislocated DP requires coindexation with the element [clitic ... t]. For more on this issue see Anagnostopoulou (1994), Androulakis (1997), Theofanopoulou-Kontou (1986-7).

iv) The affix analysis cannot capture the right choise of clitics according to the subcategorization frame of the main verb in the periphrastic constructions in which the clitic appears adjoined to the auxiliary, though constrained by the main verb:

(19)a. eɣ rapsa to ɣ ramma
I-wrote the letter
b. to eɣ rapsa
it I-wrote
c. to exo ɣ rapsi
it I-have written
(20)a. * xamoɣ elasa to Niko
I-smiled Nick
b. * to xamoɣ elasa
it I-smiled
c. * to exo xamoɣ elasi
it I-have smiled
(21)a. eðosa to vivlio tu Niku/sto Niko
I-gave the book to-Nick
b. tu to eðosa
to-him it I-gave
c. tu to exo ðosi
to-him it I-have given

Similarly it cannot explain why in the perfect of the passive voice the clitic cannot appear with ***exo***.

(22)a. * ɣ raftika to ɣ ramma
I-was-written the letter
b. * to ɣ raftika
it I-was written
c. * to exo ɣ rafti
it I-have written

An analysis according to which ***exo*** is inflectionally bound to the main verb (23) may explain these subcategorization restrictions, but it ... be entertained for the reasons presented above (see the discussion of examples 9-1

(23)a. to-exo-ðosi
it I-have given
b. tu-to-exo-ðosi
to-him it I-have given

3.2. Morphological evidence

Clitics inflect for person, number, gender and case, a property of lexical items and not of affixes (recall the complexity test by Zwicky (1985: 288): "Words are frequently morphologically complex...affixal units rarely are"). The feature of case treats them in fact as the arguments proper of the verb. This conclusion is strengthened by the fact that the morphological variation of the clitics follows almost completely the regular morphological pattern also followed by the corresponding stronger pronominal forms. Compare the strong pronominal forms of ***ekinos*** 'that, he' (a), and ***aftos*** 'this, he' (b), and the strong forms of the first and second person pronominals ***ego*** 'I' and ***esi*** 'you' with the corresponding clitic forms (c):

(25)	Masc.Sg.gen.:	a. ekinu	b. aftu	c. tu	
	acc.:	a. ekinon	b. afton	c. ton	
	Masc.Pl.gen.:	a. ekinus	b. aftus	c. tus	
	acc.:	a. ekinon	b. afton	c. tus	(ton)
(26)	1st.Sg.acc.:	emena		c. me	
	2nd.Sg.acc.:	esena		c. se	
	1st.Pl.acc.:	emas		c. mas	
	2nd.Pl.acc.:	esas		c. sas	

3.3. The phonological evidence

i) Stress: In Greek each grammatical word (in the traditional narrow sense) carries one stress only. Furthermore each stress must occur on one of the last three syllables. This constraint is referred to as the *antepenultimate stress rule* or the *trisyllabic rule* (Holton, Mackridge & Philippaki-Warburton 1997). Derivationally and inflectionally (again in the narrow sense) related words may show differences in the position of the stress as shown in the following examples:

(27)a.	vunó		'mountain'	
b.	diminutive:	vunaláki	'little mountain'	
(28)a.	nom.: máθima	b. gen.: maθíma tos		'lesson'
(29)a.	Pres.Act.1st.Sg.:	ðjavázo	'I read'	
b.	Past.Act.1st.Sg.:	ðjávaza	'I was reading'	
c.	Past.Act.1st.Pl.:	ðjavázame	'We were reading'	

In all the examples, both nouns and verbs, the assignment of stress is constrained by the trisyllabic rule but more importantly, from our point of view, the adjustment is achieved by shifting the stress (or reassigning the stress) in ways that will satisfy the constraint. Let us now examine what are the consequences for the position of stress when one or two clitics are attached to a grammatical word (as enclitic) increasing the length of the form and creating units which violate the trisyllabic rule, as is the case of imperatives.

(30)a.	ðjávase	'Read'
b.	ðjávasé to	'Read it'
c.	ðóse	'Give'
d.	ðóse mu	'Give me'
e.	ðóse mú to	'Give it to me'

What we observe here is that there is no shift of the underlying stress (no reassignment of stress). Instead the trisyllabic rule is restored by the development of a secondary stress on the penultimate of the total string. This shows that cliticisation is a different process from inflection proper and that it operates after inflection has been completed. It takes place in a larger domain than a word domain, in the sense of the relevant tests in Zwicky (1985: 288). These phenomena are naturally handled within a theory that recognises that the effect of clitics on the stress takes place within the syntax after combining clitics with their hosts.

ii) Euphonic *-e*: In Greek there is a strong tendency for open syllables in word final position. When a word terminates in the licit final consonant *-n*, there is a tendency for an euphonic *-e* to be added after it:

(31)a.	irθan	→	irθane	'They came'
b.	ton peðjon	→	ton peðjone	'of the children'

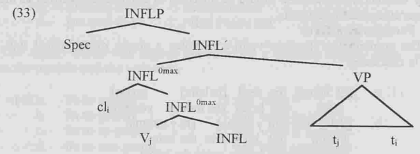
Affixes do not need and nor do they show such a tendency. And yet object clitic pronouns may appear with such final euphonic *-e*:

(32)a.	tone θelume	'We want him'
b.	ðen tine fovate	'He is not afraid of her'

3.4. The analysis of object clitic structures in Greek

The evidence presented above argues strongly that the object clitic pronouns in Greek are not affixes of the verb but constitute separate syntactic units and thus separate syntactic entries. The analysis we assume is the following (see also Kayne 1991, Chomsky 1995):

(33)



4. Particles

4.1. Their forms and their functions

In addition to clitic pronouns a verb may be preceded by one of the two negative morphemes and one of the two mood/tense particles. These are: ***θa*** whose prototypical use is to express futurity; ***na*** which marks the subjunctive mood; the negative particle for the indicative is ***ðe(n)***, while for the subjunctive it is ***mi(n)***. The possible combinations are shown below:

(34)a.	ðen to egrapsa	
	not it I-wrote	'I did not write it'
b.	ðen θa to grapso	
	not will it I-write	'I will not write it'
c.	ðen θa to ixa γrapsi	
	not will it I-have written	'I will not have written'
d.	na min to γrapsis	'You should not write it'
e.	na min to ixes γrapsi	'You should not have written it'

The analysis which will be assumed here is the following (Philippaki-Warburton 1994b, 1996): ***na*** is a subjunctive mood marker generated under a Mood (MD) functional

category. The particles ***ðen*** and ***min*** are generated under a NEG functional category. The particle ***θa*** is generated under a different functional category (lets call it FT, for futurity) from the one of ***na*** (35; Philippaki-Warburton 1996). For a different analysis see Drachman (1994) and Rivero & Terzi (1995)):

(35)	MD	NEG	FT	
a.	Ind. [0]	ðen	θa	to γ rapsis
b.	Sub. [na]	min		to γ rapsis

The full structure with all the possible combinations is the following:

(36)	MD	NEG	FT		INFL	***exo***	VOICE	ASPECT V
a.	Ind. [0]	[ðen]	[θa]	io.cl	do.cl			
b.	Sub.[na]	[min]	---	io.cl	do.cl			

4.2. The intermediate status of particles

Irrespective of the specific details of the analysis of these particles we must now come to the question relevant to the issue of the interface between syntax and morphology, namely whether these particles should be analysed as affixes of the verb or as independent and separate syntactic elements.

4.2.1. Particles as affixes

i) The combination [particle+verb] constitutes a single phonological unit as far as stress is concerned:

(37)a.	[O Nikos]	[θa ferí]	[ta lefta]	[avrio]
	Nick	will bring	the money	yesterday
b.	[θelo]	[na min γ rapsis]	[to γ ramma]	
	I want	not you-write	the letter	
	'I do not want you to write the letter'			

ii) There are certain syntactic phenomena which apply to such strings (i.e. particle+clitic+verb) treating them as single syntactic units, larger than the grammatical word, in the narrow sense of this term, but smaller than the phrase (verb focalisation, which involves movement of the whole verb group to the sentence initial position leaving the object DP behind (38), deletion (39), and co-ordination (40)):

(38)a.	O Nikos θa ðosi ta lefta	
	Nick will give the money	
b.	**[θa ðosi]** o Nikos ta lefta	
	will give Nick the money	
c.	* θa o Nikos ðosi ta lefta	
d.	O Nikos ðe θa ðosi ta lefta	
	Nick not will give the money	'Nick will not give the money'
e.	**ðe θa ðosi** o Nikos ta lefta	
f.	* ðosi o Nikos ta lefta ðe θa	
(39)a.	Ti su ipe na min kanis?	'What did he tell you not to do?'
b.	Na min fiγ o	'Not to leave'
c.	Fiγ o	

(40)a.	* θa erθo avrio ke fiγ o	meθavrio
	will I-come tomorow and I-leave the day after tomorrow	
b.	θa erθo avrio ke θa fiγ o	meθavrio
	will I-come tomorow and will I-leave the day after tomorrow	
	'I will come tommorow and leave the day after tomorrow'	

The evidence presented above shows that the combination [particle+verb] forms a single unit both for phonological but also for syntactic reasons and this can be said to strengthen the proposal that particles are affixes. However this conclusion must be rejected because there is strong evidence against the affixal view, as we will argue below.

4.2.2. Particles as independent syntactic elements

i) If we accept all the arguments offered above that auxiliary ***exo*** is a separate lexical entry from that of main verb, we must reject the view that particles are affixes, because their treatment as affixes will entail that they should appear as affixes both on the monolectic verb forms but also on the auxiliary ***exo*** forms, as in (41). This duplication is both redundant and counterintuitive.

(41)a.	θa fiγ o	'I will leave'
b.	na fiγ o	'... that I leave'
c.	[θa exo] fiγ i	'I will have left'
d.	[na exo] fiγ i	'... that I have left'

ii) The clitic, which was argued before to be a separate lexical entry, intervenes between the particle and the verb form:

(42)a.	θa to grapso	'I will write it'
b.	θa tu grapso	'I will write to him'
c.	θa tu to grapso	'I will write it to him'
d.	θa to exo grapsi	'I will have written it'
e.	θa tu exo grapsi	'I will have written to him'
f.	θa tu to exo grapsi	'I will have written it to him'

iii) If particles are to be treated as affixes, we must also treat as affixes the negative morphemes ***ðen*** and especially ***min***, because the particle ***na*** precedes negative ***min***. Thus if ***na*** is an affix either ***min*** is also an affix or we end up with the same situation as with ***exo*** and the clitics discussed above.

From the above discussion we conclude that particles have also an independent syntactic status. Additional support for this conclusion derives from the fact that particles can be emphatically stressed e.g. ***na MIN to ðosis*** 'you should NOT give it', and also can be nominalised by the use of the definite article, e.g. ***Ta θa ke ta min*** 'The wills and the nos'.

4.2.3. The paradox of the intermediate status of particles

We can now draw the conclusion that particles are separate independent syntactic items which enter the syntax as independent syntactic elements and not as affixes, but somehow they end up united with the verb form which they grammatically modify. Thus, the string [particle+verb], though it is not a unit in the Lexicon, but consists of two independent syntactic elements, nevertheless it can function as a unit in the syntactic component. The challenge is to find a formal account of this phenomenon.

5. The formal account of the intermediate status of clitics and particles

The whole of the verb group consists of a lexical entry for the verb plus a number of reduced grammatical elements which also constitute separate entries at the Morphology/Syntax interface. This conclusion, however, leaves unaccounted for the evidence that these verb groupings behave as single units for the purposes of some phonological and syntactic phenomena.

The solution, which we believe will satisfy both types of properties of these elements is one which formally recognises two different types of word (see also Di Sciullo & Williams 1987). *Primary words*, or *grammatical words* (the *morphological objects* or *syntactic atoms* in Di Sciullo and Williams' terms), are those which enter the syntax as separate entries. These are the units of the Morphology/Syntax interface. These are the inflectionally complete members of the narrowly defined verb paradigm, as well as those words which are either monomorphemic (particles, clitics). Another type of word, which we may call *secondary* or *syntactic word* is formed subsequently after the interface. This unit consists of such reduced lexical elements as particles and clitics in combination with the grammatical word that contains the head of the construction. The questions that are raised now are the following:

i) How do particles and clitics combine into a single syntactic word with the verb?

ii) Where precisely does this union take place?

Some theoretical details are in order here: In the Minimalist Program (Chomsky 1995) functional/grammatical information is projected on the syntactic structure by means of functional heads. These heads consist of certain grammatical features to be satisfied either by verb-movement (Operation Attract/Move) or by merging a functional word (Operation Merge), mostly a particle. In Greek Voice, Aspect, INFL and MD (when imperative) are satisfied by means of verb-movement; their morphophonological exponents are affixed on to the verb stem (verb head) in the Lexicon/Morphology component before syntax. This is what we refer to as grammatical word. On the other hand, NEG, FT and MD (subjunctive) are satisfied by means of merging a particle (the negative ***den*** and ***min***, the future ***θa***, and the subjunctive ***na***). No verb movement is required, and actually it is banned as unmotivated. The theory thus predicts that there is no motivation for the unity of the verbal group. However, we presented evidence showing that the verbal group constitutes a unit for some syntactic operations (focalisation, ellipsis, co-ordination). In order to solve this problem we propose a merging operation, which unites all these elements in the syntactic component, in terms of the syntactic operation of *Move* α. We call this operation *Move-Incorporate*.

It may be argued that our proposal is facing a theoretical problem. According to the restricted theory of the MP, movement is constrained by the economy principle of *Last Resort* and it is thus restricted to take place only in order to satisfy certain functional features on morphologically empty functional heads. Our rule *Move-Incorporate*, however, involves full lexical items and not simply features on lexical heads and this may be undesirable. To overcome this problem we suggest that the grammatical affinity between the particles and the verb form which they modify can be formally captured by assuming that the particles carry the feature [+V] which needs to be satisfied in the syntax. Thus, particles are grammatical words that do not carry a categorial feature but a functional one, like any empty functional head. Given these assumptions we propose that derivation proceeds as follows:

All the functional information coming from the Lexicon is satisfied either by moving the verb all the way up to the functional heads attracted by their abstract features, or by

inserting a particle under the relevant functional head. If the derivation contains a clitic, at some point of the derivation, the clitic will move to adjoin to the INFL-V head creating an $INFL^{0max}$. If a structure contains particles these will be marked by [+V] feature. In fact all projections relevant to the grammatical modification of the verb will be marked by this feature. Thus a verb group structure will be as in (46):

(46)	MD^0	Neg^0	FT^0	$INFL^{0max}$
	+V	+V	+V	+V
	na	min		to γ rapso
		ðen	θa	to γ rapso

A merging operation will now apply moving the unit containing the grammatical word for the verb (the head word) to the next F^0 category until one single word unit is created. Thus $INFL^{0max}$ will be attracted by ***θa*** and it will move to incorporate to it creating the node FT^{0max} (***θa-to-γrapso***). Then the negative particle ***den*** will attract the FT^{0max} which will move to incorporate to the NEG^0 creating a NEG^{0max} consisting of the NEG^0 plus FT^{0max} (***ðen-θa-to-γrapso***) and so on.

We have presented a merging operation *Move-Incorporate*, which acts in a syntactic way, subsumed in fact under the Operation *Move* α. However we must now clarify the differences between *Move-Incorporate*, relevant to the merging of independently existing lexical items, and the standard *Move* α, which operates in order to check functional information represented as features on the heads of affixal functional categories. The differences are as follows:

i) The features which motivate the *Move-Incorporate* are associated with independently existing lexical items and not with morphologically empty functional nodes.

ii) *Move-Incorporate* is relevant to X^{0max} and not to X^0.

iii) *Move-Incorporate* results in right adjunction with the host grammatical word, whereas *Move* α results in left adjunction.

iv) The motivation for *Move-Incorporate* is not to eliminate the functional features of an empty head but, it is the result of the grammatical affinity of the particle to the head as well as its morphophonologically dependent status.

6. Conclusion

In our analysis, which draws a distinction between *grammatical* and *syntactic* word, the debate among various analyses revolving around the lexical vs affixal character of clitics and particles is reconciled. The phonological, morphological and syntactic facts which point to the lexical independence of these items are satisfied by their original lexical status. On the other hand their morphophonological dependence and their forming a single unit with their host, as if they were affixes, is satisfied by the cliticization and *Move-Incorporation* operations triggered by their functional role to grammatically modify their host and the fact that they are morphophonologically reduced. Thus their intermediate status is revealed to be the result of their history within the derivation and the paradox of their conflicting properties is thus resolved and explained.

References

Anagnostopoulou, E. 1994. *Clitic Dependencies in Modern Greek*. Ph.D dissertation: University of Salzburg.

Androulakis, A. 1997. "Wh- and direct object clitics revisited". Ms. (in prep.).

Chomsky, N. 1995: *The Minimalist Program*. Cambridge Mass.: MIT Press.

Di Sciullo, A. M. and E. Williams. 1987. *On the definition of word*. Cambridge Mass.: MIT Press.
Drachman, G. 1994. "Verb movement and minimal clauses". In Philippaki-Warburton, Nikolaidis and Sifianou (eds.).
Holton, D., P. Mackridge and I. Philippaki-Warburton. 1997. *Greek: A Comprehensive grammar of the modern language*. London: Routledge.
Horrocks, G. 1990. "Clitics in Greek: A diachronic view". In Roussou, M. and S. Panteli (eds), *Greek outside Greece II*. Athens: Diaspora books, 35-52.
Joseph, B. 1985. "Complementizers, particles and finiteness in Greek and the Balkans". *Folia Slavica* 7, 390-411.
Joseph, B. 1988. "Pronominal affixes in Modern Greek: The case against clisis". *CLS* 24, 203-15.
Kayne, R. 1991. "Romance clitics, verb movement and PRO". *Linguistic Inquiry* 22, 647-686.
Philippaki-Warburton, I. 1977. "Modern Greek clitic pronouns and the 'Surface Structure Constraints' Hypothesis". *Journal of Linguistics* 13, 259-81.
Philippaki-Warburton, I. 1987. "The theory of empty categories and the pro-drop parameter in Modern Greek". *Journal of Linguistics* 23, 289-318.
Philippaki-Warburton, I. 1994a. "Verb movement and the distribution of clitic pronouns". In Philippaki-Warburton, Nikolaidis and Sifianou (eds), 53-60.
Philippaki-Warburton, I. 1994b. "The subjunctive mood and the syntactic status of the particle *na* in Modern Greek. *Folia Linguistica* 28, 297-328.
Philippaki-Warburton, I. 1996. "Functional Heads and Modern Greek Syntax". Paper presented at the GLOW Workshop on *Current issues in Modern Greek Syntax*. Athens.
Philippaki-Warburton, I., K. Nikolaidis and M. Sifianou (eds) 1994. *Themes in Greek Linguistics: Papers from the First International Conference on Greek Linguistics*. Amsterdam: J. Benjamins.
Rivero, M. L. and A. Terzi 1995. "Imperatives, V-movement and logical mood. *Journal of Linguistics* 31, 135-146.
Theofanopoulou-Kontou, D. 1986-7. "Kenes katigories ke klitika sti NE: I perisptosi tu amesu antikimenu (Empty categories and clitics in MG: The object case)". *Glossologia* 5-6, 41-68.
Zwicky, A. M. 1985. "Clitics and particles". *Language* 61, 283-305.

Susan Steele
University of Connecticut
steele@gulley.vpa.uconn.edu

INFLECTION AND INFORMATION

Abstract
An information-based approach to morphology provides a simple and clean method of distinguishing among morphological operations, by focusing on their informational effects. One value of making distinctions on these grounds is the internal order it brings to morphological processes; another is the possibility it allows for generalizations across the domains of morphology and syntax.

1. Introduction

Two debates inform current morphological theory. One has to do with the status of morphemes. Morpheme-based theories focus on identifying discrete phonological sequences and accounting for the effect of their presence. Pointing to the many morphological phenomena that do not involve the addition of a discrete phonological piece, non-morphemic morphological theories focus rather on accounting for the kinds of relationships that can exist between and among stems and words. A second debate has to do with whether the principles governing morphology are fundamentally distinct from the principles governing syntax. The two issues are logically independent in fact, but in practice they are not. In general, morpheme-based theories also take the position that the principles governing the structure of words and those governing the structure of sentences are essentially identical, while non-morphemic theories argue that the principles governing the two domains are fundamentally distinct.

After establishing the assumptions under which the investigation will proceed, the body of this paper expands on the information-based theory of inflection offered in Steele 1995 and termed 'Articulated Morphology'. Like all information-based theories, Articulated Morphology focuses on the informational relationships among linguistic objects – i.e., on the differences in information between one linguistic object and another. As a processual theory, Articulated Morphology is also fundamentally concerned with the processes that yield these differences. The heart of this paper is an information-based distinction among three kinds of operations – those that add information to the operand (the object they operate on); those that change its information; and those that eliminate information from the operand. Given the primacy of these kinds of informational effects, it is possible to conceptualize more clearly the character of inflection, derivation, and compounding. Further, the classification of morphological operation types corresponds simply and intuitively to a classification of syntactic operation types.

Information-based theories of syntax have achieved wide currency. In contrast, the study of morphology from an informational perspective is in its infancy. This paper redresses this asymmetry and, in the process, offers a new perspective on similarities with the principles of syntax. Although the interest in morphemeless morphological theories has contributed to the development of an information-based view of morphology, being information-based does not in and of itself determine a position on the existence of morphemes. That is, it is logically possible that an informational difference between two linguistic objects would always be associ-

ated with the presence of discrete phonological sequences. The logical possibility, however, is not an article of faith; rather, it is an empirical question. Articulated Morphology does not require the existence of morphemes and easily accommodates the phenomena driving the development of non-morphemic theories. In regard to the issue of the principles of syntax and morphology, then, the information-based approach argued for here affords generalizations across the two domains, a conclusion more commonly associated with morpheme-based theories.

2. Background

I begin with three assumptions. First, drawing on information-based syntactic theories (like HPSG or LFG) and consistent with the morphological work of Aronoff 1994, I take linguistic objects, whether they involve stems, words, phrases or sentences, to be signs involving a phonological part, a semantic part and, crucially for this paper, a syntactic part.

1. [phonology]
   [syntax]
   [semantics]

Second, I assume that the syntax of a sign is an articulated attribute/value structure. That is, each of the three parts of the sign in 1 involves a set of features and associated values. Because I will take no position for the purposes of this paper on the internal structure of the phonological and semantic parts, 2 expands accordingly the syntax of a sign only. Each of the superscripted *F*s stands for an attribute; the lower case letters represent values.

2. [phonology]
   $\begin{bmatrix} F^1\text{: a} \\ F^2\text{: b} \\ F^3\text{: c} \end{bmatrix}$
   [semantics]

Third, morphological and syntactic operations take signs and manipulate their properties. Three logical possibilities present themselves: An operation may (a) modify, (b) add to, or (c) subtract from the information represented in the attribute/value structure of the operand. Although, as we will see, these three logical possibilities are not entirely mutually exclusive, 3 presents a schematic representation of each as an independent option.

3. Informationally Additive Operation:
   [Property X] → [Property X & Property Z]
   Informationally Modificational Operation:
   [Property X] → [Property Y]
   Informationally Subtractive Operation:
   [Property X & Property Z] → [Property X]

An example from Potawatomi (an Algonquian language) illustrates the first type and also gives a general sense of the approach to be adopted.[1] Intransitive verb stems in Potawatomi may include information about the animacy and person of their subjects. The intransitive stem *nis:e*

[1] The analysis on which this example is based is found in Steele 1995. The Potawatomi data are drawn from the published work of Hockett. The Luiseño data used later in this paper are from my work with the late Villiana Hyde and are written in the orthography introduced in Hyde 1971.

'fall down' requires an animate third person subject. (Consistent with the focus on the syntax of a sign, the phonological and semantic properties are maximally simplified. The phonology is the orthographic representation. The semantic properties appear as a simplified predicate calculus.)

4. Phon: *nis:e*
   Syntax $\begin{bmatrix} \text{ANIMATE: +} \\ \text{PERSON: 3} \end{bmatrix}$
   Seman: FALL DOWN(x)

A word includes not only information about the animacy and person of the subject, but also information about its number. So, the word *nis:e-k* 'they fall down' requires an animate third person **plural** subject.

5. Phon: *nis:e-k*
   Syntax $\begin{bmatrix} \text{ANIMATE: +} \\ \text{PERSON: 3} \\ \text{NUMBER: pl} \end{bmatrix}$
   Seman: FALL DOWN(x)

The attribute/value pair having to do with number differentiates a word (which has it) from a stem (which doesn't). The mapping between 4 and 5 is a morphological operation that takes an animate third person stem and yields an animate third person plural word, by both changing the phonology and adding 'NUMBER: pl'. (The operation in 6 adds both the attribute NUMBER and its value, in order to emphasize the fact that the domain lacks number and the codomain includes number. It could equally be the case that the domain includes the attribute and the operation adds only its value.)

6. pl:
   Phon: Z → Phon: Z*k*
   Syntax $\begin{bmatrix} \text{ANIMATE: +} \\ \text{PERSON: 3} \end{bmatrix}$ → Syntax $\begin{bmatrix} \text{ANIMATE: +} \\ \text{PERSON: 3} \\ \text{NUMBER: pl} \end{bmatrix}$
   Seman: V(x) → Seman: V(x)

The two other types of operations in 3 will be exemplified below. But, given this example, we can consider and exemplify the two fundamental principles of Articulated Morphology:

7. **Principle 1**: Morphological objects can be differentiated according to the kind of information they present.
   Principle 2: Morphological operations can be differentiated according to the kind of object they are performed on and the kind of information they manipulate.

We will also assume the following special case of Principle 1, as establishing initial parameters.

8. **Assumption A**: Stems are informationally reduced, relative to words.

Ex 4 is a Potawatomi stem; 5 is a Potawatomi word. The fact that there is a difference is consistent with Principle 1 and the particular contrast is consistent with Assumption A. Moreover, the distinction between Potawatomi (verb) stems and the words that contain them is consistent with both Principle 1 and Assumption A. As in 4, the syntactic information associated with a Potawatomi (verb) stem indicates how many arguments the verb requires (here one) and the animacy of at least one (the only argument in the intransitive case and the second argument in a transitive case); the syntactic information may also include the person of one of the arguments (here third person). The syntactic information associated with a Potawatomi word includes all the in-

formation associated with the stem and more, as Assumption A requires. In particular, each argument is associated with both person and number. Schematically:

9. Stem: # of arguments = X
   # of instances of person ≤ X
   # of instances of number = 0
   Word: # of arguments = X
   # of instances of person = X
   # of instances of number = X

Principle 2 refers to the character of operations. The operation in 6 that adds number information takes a stem and yields a word. However, it is not the case that all morphologically simple stems include a particular person value. Morphologically simple transitive animate stems, for example, are subject to operations that determine the person of their arguments. The animate transitive stem *wapm* 'x see anim', for example, can occur in the following complex forms: *wapma* 'x see 2anim', *wapmUkO* '3 see anim', *wapmn* 'x see 2anim', and *wapmy* '2 see anim'. The operations in 10, each of which has a distinct informational effect on the attribute PERSON, yield this array.

10. a. Phon: *Z* — Syntax: [P:] [ANIM: +, P:] → Phon: *Za* — Syntax: [P:] [ANIM: +, P: 3]

b. Phon: *Z* — Syntax: [P:] [ANIM: +, P:] → Phon: *ZukO* — Syntax: [P: 3] [ANIM: +, P:]

c. Phon: *Z* — Syntax: [P:] [ANIM: +, P:] → Phon: *Zn* — Syntax: [P:] [ANIM: +, P: 2]

d. Phon: *Z* — Syntax: [P:] [ANIM: +, P:] → Phon: *Zy* — Syntax: [P: 2] [ANIM: +, P:]

The contrast between the operation in 6 and the operations in 10 illustrates the distinction between operation types allowed by Principle 2. The Potawatomi operations that add number take stems and yield words. The operations in 10 that add person take stems and yield stems. These operations illustrate one more point that is essential to our discussion. The schematic statements about addition, subtraction and modification in 3 do not differentiate between attributes and values. Our focus here is on values. Addition always involves the addition of a value; modification always changes a value; and subtraction always removes a value. It is logically possible, as noted above, that some operations might involve both an attribute and a value and others only a value. Because we will presume that stems and words include the same attribute sets, only the latter is an option.

11. a. Addition: [F:] → [F: x]
    b. Modification: [F: x] → [F: y]
    c. Subtraction: [F: x] → [F:]

Refining the differences among these three operational types is possible through an analysis of Luiseño, a Southern California Uto-Aztecan language of Northern San Diego County.

3. Addition and Modification.

The Potawatomi case illustrates that addition involves adding a type of information previously absent. The Luiseño case contrasts this straightforward example of addition with an equally straightforward example of modification – replacing an instance of a particular kind of information by the same kind of information.

The syntax of Luiseño stems involves the four features in 12.

12. Syntax: CAT(EGORY):
    N(UMBER):
    ASP(ECT):
    SUBCAT(EGORIZATION):

Every stem has a value for the features CAT and SUBCAT; a stem may, but need not, have in addition a value for the features ASP and N. The values available to the last three are reasonably familiar; only the first requires any introductory comments. Although we could employ any of a variety of terms to distinguish among category types, I continue from previous discussions (e.g. Steele 1988) the mnemonic labels in 13, where +/-poss indicates the possibility (or the impossibility, respectively) that a stem may combine with a possessive and +/-abs indicates the possibility (or the impossibility, respectively) that a stem may combine with an absolutive. (**Absolutive** is the term used in Uto-Aztecan for a set of morphs – *-t, -ta, -sh, -chaˌ, -la, -l* – that appear on the citation (or absolute) form.)

13.

[CAT: -poss, +abs]	[CAT: +poss, +abs]
hunwu 'bear'	too 'rock'
qeengi 'squirrel'	huu 'arrow'
tuupa 'sky'	kutapi 'bow'
[CAT: -poss, -abs]	[CAT: +poss, -abs]
heela 'sing'	peew 'spouse'
'aamo 'hunt'	yo' 'mother'
pelee'i 'lick'	kaamay 'son'

The informational effect accompanying the addition of the morph *-ki* changes the categorial assignment of a stem. As illustrated in 14, stems including this morph are based on members of the first category, but they behave morphologically like members of the fourth.

14. a. hunwut 'bear' — nohunwuki 'my (toy) bear'
    *nohunwu — *hunwuki-sh
    b. tuupash 'sky' — notuupaki 'my (piece of) sky'
    *notuupa — *tuupaki-sh

These data can be captured with a morphological operation that modifies the categorial information of a stem – replacing the value '-poss, +abs' **with another value of the same type** '+poss, -abs'.

15. [+poss, -abs]:

Phon: *X* → Phon: *Xki*

Syntax [CAT: -poss, +abs] → Syntax [CAT: +poss, -abs]

Another example is useful to underline the character of modification as replacing one value with another of the same type. In addition to the attribute CAT(egory), the syntax of Luiseño stems includes information about their subcategorization possibilities. (See Steele 1990 for an extensive discussion.) For example, *pelee'i* 'lick' requires two arguments, a subject, which need not have lexical instantiation, and an object-marked object; *heela* 'sing', in contrast, requires only a subject, which similarly need not be lexically instantiated. Ex. 16 provides some simple examples and 17 illustrates how the subcategorization properties are to be represented. The lower case 'obj-mrk' identifies the formal property of the obligatorily present argument – it must be marked for object; the capital 'SUBJ' indicates the necessity of a subject but does not carry requirement as to its formal character.

16. a. pommay pum **pelee'i**wun 'They are licking their hands.'
their.hands.obj aux are.licking

b. **heela**q up 'S/he is singing.'
is.singing aux

17. a. Phon: pelee'i

Syntax: [CAT: -poss, -abs; SUBCAT: [obj-mrk; SUBJ]]

Seman: LICK(x y)

b. Phon: *heela*

Syntax: [CAT: -poss, -abs; SUBCAT: [SUBJ]]

Seman: SING(x)

Luiseño has a morphological causative, the effect of which is to change the subcategorization properties of a stem, as the contrast between 16a and 18 will make clear.

18. pommay pum pomoomi **pelee'ini**wun
their.hands aux them.obj is.making.lick
'They are making them lick their hands.'

The syntactic information of the stem *pelee'ini* is, thus, distinct from the syntactic information associated with the stem *pelee'i*.

19. Phon: pelee'ini

Syntax: [CAT: -poss, -abs; SUBCAT: [obj-mrk; obj-mrk; SUBJ]]

Seman: LICK(x y z)

The difference between 17a and 19 might appear to be characterizable as addition – 19 has, after all, an 'additional' argument in SUBCAT. However, the value of an attribute is not internally decomposable. Just as the operation in 15 replaces a complex value with another complex value, the causative operation replaces one non-decomposable argument structure type with another.

20. causative:

Phon: X → Phon: X*ni*

Syntax: [CAT: -poss, -abs; ASP:; SUBCAT: [obj-mrk; SUBJ]] → Syntax: [CAT: -poss, -abs; ASP:; SUBCAT: [obj-mrk; obj-mrk; SUBJ]]

Seman: S(x y) → Seman: S(x y z)

Although the argument types – e.g. 'obj-mrk' or 'SUBJ' – comprise a small set, their combinations are also drawn from a small set, one that does not exhaust the logically possible combinations of argument types. (Argument types include, in addition to 'obj-mrk' and 'SUBJ', those marked for number and those with postpositions, among other possibilities.) For example, no argument structure is composed of more than four arguments and only those that have the formal property 'obj-mrk' can occur more than once. Were the attributes for the feature SUBCAT decomposable, operations that create non-existent combinations of arguments would require adhoc proscription. Nothing would preclude three, four or more causatives, because each would simply add 'obj-mark'. Thus, the value for the attribute SUBCAT (or any other attribute) can be internally complex, but change in their values must be viewed as replacing in the syntax of the stem one non-decomposable value with another.

4. Inflection and Derivation.

Against this background, 21 represents the conceptual difference between the addition of information and the modification of information.

21. a. Addition [F :] → [F : x]

b. Modification [F : y^i] → [F : y^j] (where y^i and y^j are values of the same type)

In Steele 1995 I proposed that inflection be defined as operations involving the addition of information. If this is reasonable and if our morphological primitives are inflection, derivation and compounding, it is also reasonable to define derivation as operations that modify information. The schema in 21a would, thus, represent inflection and that in 21b, derivation. However, not all operations are purely one or the other. The question, then, is how to treat such hybrid operations.

Essentially all Luiseño stems which are [CAT: -poss, -abs] may undergo any of a rich set of aspectual operations. Ex. 22 introduces the result of three such operations on the stem *'aamo* 'hunt', bolding the phonological effect.

22. 'aamo**lo** 'hunt (generically)'
'aamo**qala** 'hunt (changing over time)'
'aamo**mokwi** 'hunt (past)'

Not only do the aspectual operations add to the syntax of a stem aspectual information like that represented somewhat clumsily by the glosses in 22, they also replace the category of the stem with a different category of the same type. That is, they combine an informationally additive effect with an informationally modificational effect.

The stem *'aamo* 'hunt' has the attribute/value structure in 23.

23. Phon: *'aamo*

Syntax: [CAT: -poss, -abs; N:; ASP:; SUBCAT: [obj-mrk; SUBJ]]

As the value for CAT indicates, *'aamo* is subject to operations introducing neither the absolutive nor the possessive. Rather, stems of this type may undergo any one of a set of tense operations.

24. **'aamoq** 'is hunting'
'aamon 'will hunt'
'aamoqu 'was hunting'

The forms in 22, in contrast, are not subject to tense operations, but they do exhibit different possibilities relative to the absolutive and the possessive.

25. possessive only [CAT: +poss, -abs]
po'aamolo 'for 3sg to hunt'
possessive and absolutive [CAT: +poss, +abs]
po'aamoqala '3sg hunting'
'aamoqal ('aamoqala-l) 'hunting'
absolutive only [CAT: -poss, +abs]
'aamomokwi**sh** 'hunted'

These facts can be accommodated by a set of aspect operations. Each takes a stem where the value for ASP is not specified and adds a value for this attribute; each, in addition, takes a stem specified [CAT: -poss, -abs] and replaces the value for this attribute with another of the same type. The three operations in 26 exemplify, therefore, an operation type that simultaneously adds and modifies information

26. a. generic:

Phon: *X*				Phon: *Xlo*		
Syntax	CAT:	-poss -abs	→	Syntax	CAT:	+poss -abs
	ASP:				ASP:	generic

b. changing:

Phon: *X*				Phon: *Xqala*		
Syntax	CAT:	-poss -abs	→	Syntax	CAT:	+poss +abs
	ASP:				ASP:	changing

c. unchanging past:

Phon: *X*				Phon: *Xmokwi*		
Syntax	CAT:	-poss -abs	→	Syntax	CAT:	-poss +abs
	ASP:				ASP:	unchanging past

The aspectual operations provide information about aspect not found in the stems to which they apply, but they also change the category of the stem – i.e. in regard to aspect they are informationally additive, but in regard to category they are information changing. If inflection and derivation exhaust our choices, the options are clear:

27. a. informationally additive = inflection
informationally modificational only = derivation
b. informationally additive only = inflection
informationally modificational = derivation

In fact, the option in 27b has been proposed in Lieber 1992. If her account of such complex operations is correct, we would identify aspect as derivation and expect it to pattern with examples of simpler derivation types like *-ki* in 15.

Lieber assigns all morphemes but inflectional affixes a 'categorial signature', a bundle of attributes which, for each morpheme, may but need not be accompanied by values. Inflectional affixes are accompanied by a reduced set of attributes, and each attribute present must be specified with a value. The categorial signature of a word is constructed from the information contributed by the morphemes it contains, according to a system of 'percolation'. First, given a binary branching tree, one branch leads to a 'head' and the other to a non-head. Second, the information from whichever morpheme is the head is percolated to the dominating node; if the head lacks information found in the nonhead, this additional information is carried up as well. Consistent with Lieber's idea that an inflectional affix cannot be a head, its partial informational structure can only be added to the categorial signature of the dominating node. 'In derivational word formation the value for a feature of a head morpheme will supercede or override that of an inner morpheme. Features from inflectional morphemes can never override features from their bases, but can only fill in values unspecified in the categorial signatures of their bases. Inflectional word formation is therefore **additive** in a way that derivational word formation...[is] not.' (p.112) Ex 28 reformulates the operation in 15 in conformity with Lieber's proposals.

28. a. *tuupa* 'sky'

CAT:	-poss +abs
ASP:	
SUBCAT:	null

b. *-ki*

CAT:	+poss -abs
ASP:	
SUBCAT:	null

c.

CAT:	-poss +abs
ASP:	
SUBCAT:	null

Head Percolation

CAT:	+poss -abs
ASP:	
SUBCAT:	null

tuupa 'sky'

CAT:	-poss +abs
ASP:	
SUBCAT:	null

ki

Aside from the general difficulties associated with positing morphemes (richly detailed by Anderson 1992, among others), this analysis is reasonably consistent with the comparable analysis of *-ki* above.

The flaws in Lieber's notion of inflection are obvious, however, when we consider an example, like Luiseño aspect, which simultaneously adds and modifies information. Because Luiseño as-

pect overwrites the value for CAT, on Lieber's account it must be of the same morphological type as Luiseño *-ki* – that is, derivational. The morpheme *-mokwi*, for example, would have the categorial signature in 29.

29. *-mokwi* [CAT: -poss, +abs; ASP: unchanging past; SUBCAT:]

As the head of the structure in 30, presumably, it would override the stem's value for CAT by 'head percolation', supplying as well a value for ASP. Because *mokwi* lacks a value for SUBCAT, the subcategorization value for *'aamo* would fill in by 'backup percolation'.

30.

[CAT: -poss, +abs; ASP: unchanging past; SUBCAT: [obj-mrk; SUBJ]]

Backup Percolation

[CAT: -poss, +abs; ASP: unchanging past; SUBCAT:] Head Percolation

[CAT: -poss, -abs; ASP: ; SUBCAT: [obj-mrk; SUBJ]] [CAT: -poss, +abs; ASP: unchanging past; SUBCAT:]

'aamo 'hunt' *mokwi*

In terms of the behavior of the resulting stems, the model in 27b as interpreted in 29 and 30 seems an incorrect result. First, no aspectual element replaces the subcategorization properties of the stem with which it combines, a fact that is without explanation on Lieber's analysis. It is simply an accident in this analysis that heads that are aspectual elements lack a value for SUBCAT. Second, any operation that is purely informationally modificational, like the causative, precedes the aspectual operations, and no operation with any modificational effect is possible after the aspectual operations. On Lieber's analysis, the causative and other modificational operations could just as well follow the aspectual operations, since all are derivational.

Of the two models in 27 that in 27a seems the best characterization of the mapping from informational effects to the distinction between inflection and derivation. That is, we can identify operations that are informationally modificational only with derivation and we can identify with inflection operations that involve informational additivity, whatever other informational effects they might have. This model can be presented somewhat more formally as follows:

31. Derivation: $[F: y^i] \rightarrow [F\ y^j]$

Inflection: $\begin{bmatrix} (F^1: y^i) \\ F^2: \end{bmatrix} \rightarrow \begin{bmatrix} (F^1: y^j) \\ F^2: x \end{bmatrix}$

On this model, inflection has a broader domain than does derivation within the informational parameters at issue. It follows, as well, that the morphological types involved will also have a wider range. The examples offered above demonstrate this latter point. The Luiseño examples of derivation and inflection both involve mapping from stem to stem, but the Potawatomi examples of inflection include an operation mapping from a stem to a stem (as in 10) and an operation mapping from a stem to a word (as in 6).

5. Compounding

To complete the picture we must consider, if briefly, the informational domain of compounding. Given the options offered at the outset, the informational 'core' for compounding must be informational subtraction.

32. Compounding: [F: x] → [F:]

An example is found again in Luiseño. The operation adding the morph *-vichu* 'want to' requires stems that are specified [ASP: unchanging], but the resulting stems lack a value for ASP. Both facts are illustrated in 33. The requirement of an aspectual value in the operand is represented by the morph *-x* on the stems *heela* 'sing' and *pella* 'dance'; this is a morph like the three presented in 22 above. The fact that the combination with *-vichu* is [ASP:] is indicated by the presence of the tense/aspect morph *-q* 'present'; this morph is mutually exclusive with aspect, as demonstrated in Section 4.

33. heela**x**vichu**q** 'wants to sing'
pella**x**vichu**q** 'wants to dance'

The operation that adds the morph *-vichu*, thus, must destroy the aspectual information associated with the stem, consistent with 32.

34. *-vichu:*

Phon: *X* → Phon: *Xvichu*

Syntax [CAT: +abs, +poss; ASP: unchanging; SUBCAT: […]] → Syntax [CAT: -abs, -poss; ASP: ; SUBCAT: […]]

The operation in 34 involves more than subtraction; it also modifies the value of CAT – [+abs, +poss] to [-abs, -poss]. Assuming that this is a reasonably representative example of compounding, it suggests that, like inflection and unlike derivation, compounding need not be informationally simple but can have multiple informational effects. Furthermore, the informational complexity for both compounding and inflectional operations involves the possibility that information can be modified as well as subtracted or added respectively.

35. Derivation: $[F: y^i] \rightarrow [F: y^j]$

Inflection: $\begin{bmatrix} (F^1: y^i) \\ F^2: \end{bmatrix} \rightarrow \begin{bmatrix} (F^1: y^j) \\ F^2: x \end{bmatrix}$

Compounds: $\begin{bmatrix} (F^1: y^i) \\ F^2: x \end{bmatrix} \rightarrow \begin{bmatrix} (F^1: y^j) \\ F^2: \end{bmatrix}$

The model in 35 makes a final prediction about the mapping between the three informational

options and the three morphological domains: No operation is simultaneously additive and subtractive.

6. Conclusion

Ex 35 maps the informational effects of addition, modification and subtraction onto traditional morphological divisions. But the primitives are the informational effects themselves. That is, the three-way informational division in 35 is a fundamental fact about morphological operations, whatever label might be applied to any of the three types. Thus, having established the primitive informational effects for morphological operations, we can consider a parallel to syntax. Although the standard view of syntax is not processual, the informational effects are arguably not exclusive to the morphological domain.

A prima facie case exists for the syntactic subtraction of information. This possibility is represented explicitly in the categorial grammar operation of functional application. For example, a transitive verb has the category VP/NP; application of this category to an NP yields a VP – VP/NP NP → VP. The combination of elements to yield many standard phrasal categories, in fact, involves the elimination of information associated with the 'non-head'. Informationally additive syntactic operations appear to be much more limited. Required is something that maintains (at least some of) the information in a syntactic domain while adding new information. The one reasonably good example might be the addition of clitics. For example, the Luiseño second position clitic complex supplies the speaker's assessment of the situation described in its complement. The contrast between the two sentences in 36 is illustrative.

36. a. noo n takwayaq 'I'm sick.'
I clitic.complex is.sick
b. noo kunun takwayaq 'I'm sick, I gather.'
I clitic.complex is.sick

Think of the complement to the clitic complex (e.g. *noo takwayaq*) as something with a temporal value. The clitic complex doesn't change or eliminate this value; rather, it adds to it a judgement. We might represent this as follows:

37. Clitic: [TNS: X] → [TNS: X; JUDGE: Z]

Finally, the requirements for informational modification limit the syntactic application. The one syntactic operation that arguably involves the replacement of one value with another of the same type is agreement across the members of a constituent. The most telling example is a case like Hopi where the dual is the result of a plural subject and a singular verb.

Although these syntactic examples demand further scrutiny, the potential parallels with morphology are intriguing. Intuitively, morphological compounds most closely resemble simple functor/argument relationships and cliticization is the most morphological part of syntax.

References

Anderson, S.M. 1992. A-Morphous Morphology. Cambridge: Cambridge University Press.

Aronoff, M. 1994. Morphology by itself. Cambridge, MA: MIT Press.

Hyde, V. 1971. An introduction to the Luiseño language. Banning, CA: Malki Museum Press.

Lieber, R. 1992. Deconstructing morphology: Word formation in syntactic theory. Chicago: University of Chicago Press.

Steele, S. 1988. Lexical categories and the Luiseño absolutive: Another perspective on the universality of noun and verb. IJAL 54:1-27.

Steele, S. 1990. Agreement and anti-agreement: A syntax of Luiseño. Dordrecht: Kluwer.

Steele, S. 1995. Towards a theory of morphological information. Language 71: 260-309.