Carnets de Grammaire

Rapports internes de CLLE-ERSS Rapport n°22 – décembre 2014

Actes des Décembrettes

8^e édition du colloque international de morphologie

Proceedings of the Décembrettes 8th International conference on morphology

6 et 7 décembre 2012 / December 6-7, 2012

Edités par / Edited by Sandra Augendre, Graziella Couasnon-Torlois, Déborah Lebon, Clément Michard, Gilles Boyé, Fabio Montermini

> Equipe de Recherche en Syntaxe et Sémantique Cognition, Langues, Langage, Ergonomie UMR 5263 CNRS & Université Toulouse – Jean Jaurès

Carnets de Grammaire est le nom d'une série de rapports internes édités par CLLE-ERSS. Cette série de rapports prédiffuse des travaux que leur degré d'aboutissement, leur nature ou leur longueur ne permettent pas de publier rapidement par les canaux habituels.

Comité de rédaction Michel Aurnague, Cécile Fabre, Claudine Garcia-Debanc, Hélène Giraudo, Frédéric Lambert, Fabio Montermini, Ludovic Tanguy

ISSN: 1965-0019

CLLE-ERSS – Maison de la Recherche – Université Toulouse - Jean Jaurès 5, allées Antonio Machado 31058 Toulouse Cedex 9

Actes des Décembrettes 8^e édition du colloque international de morphologie

Proceedings of the Décembrettes 8th International conference on morphology

6 et 7 décembre 2012 / December 6-7, 2012 Bordeaux

Le laboratoire CLLE-ERSS (UMR5263), en collaboration avec l'université Bordeaux-Montaigne et le LABRI (UMR5800), a organisé les 6 et 7 décembre 2012, la huitième édition des Décembrettes, une conférence internationale de morphologie qui se tient tous les deux ans à la fin de la première semaine de décembre. Ce colloque faisait suite aux sept éditions précédentes alternant entre Toulouse et Bordeaux.

Cet ouvrage rassemble certaines des communications faites lors de cette conférence. Les articles issus de ces communications ont été retravaillés par leurs auteurs et relus par un comité scientifique.

Pour cette huitième édition des Décembrettes, nous avions reçu comme invités Olivier Bonami (Paris4-Sorbonne) et Mark Aronoff (Stony Brook), et c'est par l'article de ce dernier que ce recueil débutera. S'en suivront les articles des conférenciers. Ces articles étant ordonnés alphabétiquement selon les noms des auteurs.

The research group CLLE-ERSS (UMR5263), in collaboration with Bordeaux-Montaigne University and LABRI (UMR5800), organized in Bordeaux, on December 6-7, 2012, the 8th Décembrettes, an international conference on morphology which takes place every two years at the end of the first week of December. This conference followed the seven previous editions alternating between Toulouse and Bordeaux.

This volume gathers some of the papers presented during this conference. These articles have been reworked by their authors and reviewed by a scientific committee. For this eighteenth edition of the Décembrettes, we received as invited speakers Olivier Bonami (Paris4-Sorbonne) and Mark Aronoff (Stony Brook), and that is by the article to the latter that this collection starts. This is followed by the articles sorted alphabetically by name of the authors.



Comités / Committees

Le colloque / The conference

Comité de sélection / Selection committee

Dany Amiot (Lille 3) Matthew Baerman (Surrey) Gilles Boyé (Bordeaux 3) Berthold Crysmann (LLF, Paris) Alain Kihm (LLF, Paris) Laurence Labrune (Bordeaux 3) Fabio Montermini (CLLE-ERSS, Toulouse) Anna Thornton (L'Aquila) Jesse Tseng (CLLE-ERSS, Toulouse)

Comité scientifique / Scientific committee

Dany Amiot (Université Lille 3, STL) Denis Apothéloz (Université de Nancy 2, ATILF) Giorgio Francesco Arcodia (Università di Milano-Bicocca) Mark Aronoff (Stony Brook University) Matthew Baerman (Surrey Morphology Group) Christian Bassac (Université de Lyon 2, CRTT) Laurie Bauer (Victoria University of Wellington) Ricardo Bermudez-Otero (University of Manchester) Delphine Bernhard (LiLPa, Université de Strasbourg) Olivier Bonami (Université Paris-Sorbonne, LLF) Gilles Boyé (Université de Bordeaux, CLLE-ERSS) Dunstan Brown (University of Surrey) Patricia Cabredo Hofherr (CNRS UMR7023 SFL) Basilio Calderone (CNRS UMR5263 CLLE-ERSS) Berthold Crysmann (CNRS UMR 7110 LLF) Georgette Dal (Université Lille 3, STL) Bernard Fradin (CNRS UMR 7110 LLF) Livio Gaeta (Università di Napoli "Federico II") Sascha Gaglia (Georg-August-Universität Göttingen) Nuria Gala (CNRS UMR7279 LIF) Hélène Giraudo (CNRS UMR5263 CLLE-ERSS) Natalia Grabar (CNRS UMR 8163 STL) Nicola Grandi (University of Bologna) Nabil Hathout (CNRS UMR5263 CLLE-ERSS) Gerd Jendraschek (University of Regensburg) Françoise Kerleroux (Université de Paris Ouest, MODYCO) Alain Kihm (CNRS UMR 7110 LLF) Laurence Labrune (Université de Bordeaux, CLLE-ERSS) Stéphanie Lignon (Nancy Université, ATILF)

Maria-Rosa Lloret (Universitat de Barcelona) Judith Meinschaefer (Freien Universität Berlin) Fabio Montermini (CNRS UMR5263 CLLE-ERSS) Fiammetta Namer (Nancy Université, ATILF) Vito Pirrelli (ILC CNR Pisa) Claudia Pons (Universitat Autònoma de Barcelona) Renate Raffelsiefen (IDS Mannheim) Pollet Samvelian (Université de Paris 3) Christoph Schwarze (Universität Konstanz) Andrew Spencer (University of Essex) Pavel Stichauer (Charles University, Prague) Gregory Stump (University of Kentucky) Ludovic Tanguy (Université Toulouse 2, CLLE-ERSS) Anna Maria Thornton (Università dell'Aquila) Jochen Trommer (University of Leipzig) Jesse Tseng (CNRS UMR5263 CLLE-ERSS) Kristel Van Goethem (Katholieke Universiteit Leuven) Florence Villoing (Université Paris 8, SFL) Madeleine Voga (Université Montpellier 3) Pierre Zweigenbaum (CNRS UPR3251 LIMSI)

Comité d'organisation / Organizing committee

Sandra Augendre (Université Bordeaux 3, CLLE-ERSS) Sébastien Bouchet (Université Bordeaux 3) Gilles Boyé (Université Bordeaux 3, CLLE-ERSS) Lionel Clément (Université Bordeaux 1, LABRI) Graziella Couasnon-Torlois (Université Bordeaux 3, CLLE-ERSS, Littorale & Scilab) Jean Cruchet (Université Bordeaux 3) Nabil Hathout (CLLE-ERSS, CNRS, Université de Toulouse 2 Le Mirail) Laurence Labrune (Université Bordeaux 3) Déborah Lebon (Université Bordeaux 3) Claire Marquebielle (Université Bordeaux 3) Clément Michard (Université Bordeaux 3) Fabrice Pellet (Université Bordeaux 3, Littorale & Scilab) Rachel Puyol (Université Bordeaux 3, Littorale & Scilab) Gauvain Schalchli (Université Bordeaux 3)

Les Actes / The Proceedings

Comité scientifique / Scientific committee

Dany Amiot (Université Lille 3, STL) Mark Aronoff (Stony Brook University) Laurie Bauer (Victoria University of Wellington) Olivier Bonami (Université Paris-Sorbonne, LLF) Gilles Boyé (Université de Bordeaux, CLLE-ERSS) Bernard Fradin (CNRS UMR 7110 LLF) Livio Gaeta (Università di Napoli "Federico II") Sascha Gaglia (Georg-August-Universität Göttingen) Natalia Grabar (CNRS UMR 8163 STL) Gerd Jendraschek (University of Regensburg) Stéphanie Lignon (Nancy Université, ATILF) Fabio Montermini (CNRS UMR5263 CLLE-ERSS) Fiammetta Namer (Nancy Université, ATILF) Christoph Schwarze (Universitä Konstanz) Madeleine Voga (Université Montpellier 3)

Comité d'édition / Editing committee

Sandra Augendre (Université Bordeaux 3, CLLE-ERSS) Graziella Couasnon-Torlois (Université Bordeaux 3, CLLE-ERSS, Littorale & Scilab) Déborah Lebon (Université Bordeaux 3) Clément Michard (Université Bordeaux 3) Gilles Boyé (Université Bordeaux 3, CLLE-ERSS) Fabio Montermini (CNRS UMR5263 CLLE-ERSS)

Les éditeurs souhaiteraient remercier le comité de rédaction des *Carnets de Grammaire* pour nous avoir permis de publier les actes des Décembrettes 8 dans leur revue.

The editors would like to thank the editorial committee of *Carnets de Grammaire* for having given us the opportunity to publish the proceedings of the Décembrettes 8 in their journal.

Contact / Information :

Les Décembrettes Département Sciences du Langage UFR Langues et Civilisations Université Bordeaux Montaigne Domaine Universitaire 33607 Pessac Cedex France Tel. +33 557-126-168 Fax +33 557-124-601 Web: http://w3.erss.univ-tlse2.fr/decembrettes2012/ E-mail: decembrettes@gmail.com

Sommaire / Table of contents

PARTIAL ORGANIZATION IN LANGUAGES: LA LANGUE EST UN SYSTÈME OÙ LA PLUPART SE TIENT1
Mark Aronoff & Mark Lindsay
VERBAL REDUPLICATION IN SINITIC
GIORGIO FRANCESCO ARCODIA, BIANCA BASCIANO & CHIARA MELLONI
THE ORGANIZATION OF CHINANTEC TONE PARADIGMS 46
MATTHEW BAERMAN & ENRIQUE L. PALANCAR
THE NOMINAL MORPHOLOGY OF LOVARI FROM AN ANALOGICAL PERSPECTIVE
MARTON ANDRAS BALO
A CROSS-LINGUISTIC INSIGHT ON AGENTIVE NOUN FORMATION IN ITALIAN AND FRENCH
VERB STEM ASPECT IN MIAN
MEASURING MORPHOLOGY: THE TIP OF THE ICEBERG? A RETROSPECTIVE ON 10 YEARS OF MORPHOLOGICAL PROCESSING 136 Helene Giraudo & Madeleine Voga
INTERNAL LOCALISATION NN_{ADV} REDUPLICATION IN SICILIAN
A CONSTRUCTIONIST ACCOUNT OF THE EMERGENCE OF A NEW LEXEME-FORMATION PATTERN: ITALIAN RIGHT-HEADED VERBAL COMPOUNDS

CLAUDIO IACOBINI

NIKOS KOUTSOUKOS

DANIELA MARZO

ISABEL PUJOL PAYET

MARION SCHULTE

PAVEL ŠTICHAUER

JANA STRNADOVA

CHRISTIAN SURCOUF

FRANCESCO-ALESSIO URSINI

PARTIAL ORGANIZATION IN LANGUAGES: LA LANGUE EST UN SYSTÈME OÙ LA PLUPART SE TIENT¹

Mark Aronoff & Mark Lindsay Stony Brook University

Abstract

At least some of the systematicity of language can be rooted in imperfection and flux. We assume that languages are the product of undirected (cultural) evolution with neither plan nor purpose and explore the role of competition in the organization of linguistic systems, with a narrow focus on inflectional morphology. In actual languages, there is often more than one way to express the same notion and these must compete in a Darwinian fashion. The competition may not be resolved quickly and instead persist for a long time. We will explore in detail one such example, the English comparative construction. Conventional grammatical wisdom is that the two ways of forming the comparative of adjectives (suffixal -er and periphrastic more) are in complementary distribution. We review the recent corpus-based literature on the English comparative and add finding of our own, based on the Google Books N-gram Corpus. We show that the two strategies have competed for millennia, with no resolution on the horizon. A case like this, though rare, is important because it makes sense only in a framework based on competition. On the morning of 28 November, 2012, while traveling by train between Manchester and York, one of us noticed the following sentence in the lead article on the front page of that day's edition of The International Herald Tribune. The new allies of Hamas want a more quiet region. Curious about this use of more quiet rather than the usual quieter, he looked at the original version of the article, published the previous day in The New York Times, where he found the following sentence: Egypt, Qatar and Turkey all want a more quiet, stable Middle East. He typed both sentences into Microsoft Word's grammar checker, which flagged the expression more quiet with the heading Comparative Use in both instances and suggested quieter instead. In the remainder of this article, we will discuss a framework in which it is not unreasonable for the synonymous expressions quieter and more quiet to coexist.

¹ We are grateful to the organizers of Décembrettes 8 for their kind invitation to present this work at the meeting in Bordeaux and for the depth of their hospitality.

The reader should know that this article is almost entirely phenomenological in nature. It does not provide much if any analysis of the phenomena presented and has little theoretical content, *stricto sensu*. Instead, we defend in it a particular metatheoretical or ideological point of view that is alien to many linguists, though not to Martians, one hopes.

Most grammarians and linguists assume that languages are fully organized systems où tout se tient, where everything holds together. One notable exception is Edward Sapir: "Were a language ever completely 'grammatical' it would be a perfect engine of conceptual expression. Unfortunately, or luckily, no language is tyrannically consistent. All grammars leak." (Sapir 1921: 38). Sapir understood that a perfect language, a grammar that didn't leak, would never change. Not being far removed in time from the neogrammarians, he also understood well that one of the most interesting properties of languages was the systematic ways in which they do change, which is made possible by their imperfection. Saussure, too, understood that the stop-action shot that was the object of synchronic linguistics was as much an analytical convenience as a reality. Here we will show how at least some of the systematicity of language can be rooted in imperfection and flux. We will couch the discussion in evolutionary terms, but we will have nothing to say about the evolution of the language faculty. Our sole interest is the evolution of individual languages (Hurford 2011).

There are three basic schools of thought concerning why languages are organized systems. The first and most widely discussed is that there is an initial cause: Languages are the product of an already organized language generating machine, a language organ. The second is that of a final cause or purpose: Languages are tools for communication, driven by a unique need to share information and states of mind that Fitch (2010) calls Mitteilungsbedürfnis. The last, which we advocate here, is that there is no cause: Languages are the product of undirected (cultural) evolution with neither plan nor purpose. The languages that we witness are survivors. These languages survive because they run in the sense that a Tinguely machine runs, not because they do anything but because a machine cannot run unless it is organized. We do not deny the value of the other two ideas. There must be some innate component to human language, since all human groups have a language, and languages clearly have communicative value. But the ludic and accidental nature of much human behavior and their importance for language have been neglected in the literature.

Only by accepting all three factors can we understand that languages are the product of innate human-specific capacities that combine with an overwhelming human need to share and the unique human capacity for cultural evolution, allowing groups of humans to construct languages, many of whose properties are historically contingent byproducts of this interaction.

The rest of this article will be concerned with the role of competition in the organization of linguistic systems. Most similarities between languages likely result from innate mechanisms and the exigencies of communication. The differences between languages most likely result from cultural evolution (Richerson & Boyd 2005). Cultural evolution works in terms of cultural selection, which in turn works by competition.

In this article, the discussion of competition in language will have a narrow focus on inflectional morphology. There has been a good deal of excellent detailed work in the last decade on competition between rival derivational affixes, much of it on French (e.g., Dal 2003; Grabar et al. 2006; Fradin et al. 2008). We discussed some of the complexities of competition between rival derivational affixes in our last presentation at Décembrettes (Lindsay & Aronoff) and we plan to return to this topic in future research.

It is important to emphasize in any discussion of the topic that most competition is not goal-driven or directed. One area of morphology where a competition model has been popular is that between words and rules in lexical access (Pinker 1999). Here is a simple example. Why do speakers of English say *went* and not *goed*? The assumption is that two mechanisms compete in producing the past tense of the lexeme GO. First, there is a general rule that operates on all English verbs to produce a past tense form by suffixing *-ed*. This rule produces *goed*. Also, for this lexeme only, there is a stored form *went*, which is retrieved at the same time as the rule is invoked. Since *went* is a very frequent word, it will be accessed very quickly and so win out over the slower rule, which needs time to compute. That is why people do not normally say *goed*. A language learner might, but that is because they do not know the form *went*.

The relation between the rule and the stored form is sometimes referred to as a horse race, the basic idea being that the fastest to the finish line, which in this case means the fastest to form a word, wins. The term *horse race*, though, is misleading, since it carries with it the implication that the two mechanisms are racing against each other towards a finish line, in the way individual people might race. In fact, there is no sense in which the mechanisms know that they are competing, let alone that there is a finishing line. The criterion for victory is completely external to the competitors. The winner is the survivor, not a triumphant victor. The same is true for competition in biology and it is this sense of competition that we are interested in here.

In a perfect language, there would be only one way to realize a meaning. In actual languages, there is often more than one way to express the same notion, as we see in this simple example. These must compete in a Darwinian fashion. The expressions are eventually sorted systematically. The sorting among alternative expressions is accomplished by competition over

time and one of several outcomes eventually results. In the simplest case, one (or more) of the alternatives dies out and one emerges victorious. It is also possible that one of the alternatives retreats to a niche and the other emerges as the default expression. This is what happened with the various ways of forming the past tense in English. Another possibility is that each alternative retreats to a distinct niche and there is no default. We will not explore this case. Finally, the competition may not be resolved quickly and instead persists for a long time. One example of this, which we will explore in detail, is the English comparative construction. Such cases are especially important, because it is only in a framework based on competition that they make sense. In most cases, the competition will eventually be sorted out. Because most frameworks are discrete, however, there is no discussion of how the nondiscrete sorting out process takes place. These theories assume that the competition itself is uninteresting and that it always leads to a discretely defined distribution of the competing expressions in which one expression is the default. We hope to show otherwise here.

Competition involves systematic distribution of elements within a system. One of the first and still most important discoveries in modern linguistics was the distinction between phonemes and allophones that underlay the phonemic principle. Saussure (1916) and others discovered the phonemic principle of contrast early on. It was only later, however, that Sapir (1933) and others showed that the subphonemic level at which elements do not contrast was equally interesting. The basic discovery was that each distinct phoneme was not always phonetically uniform. Instead, the phonetic realization could be described in terms of a set of non-contrasting allophones that were in complementary distribution, with each allophone or positional variant occurring in a specific subset of the larger environment in which the phoneme occurs.

The general theoretical distinction between contrasting emic distribution and complementary etic distribution was quickly extended to other areas of language and even culture (Pike 1967). It remains important in anthropology, where it has been much broadened (Kottak 2008). In linguistics, its most prominent extension was in morphology, where morphemes could be seen to have positional variants, allomorphs, that were distributed in a way that appeared to be analogous to allophones: in complementary distribution.

Although it is not traditional to think of complementary distribution in terms of competition, the connection is clear. Each allophone of a phoneme and each allomorph of a morpheme can be thought of as competing with all the others. The distribution that we see at any stage of the language is a resolution of this competition, with each allophone and allomorph settling into a particular environment or niche. Most of the time, one allophone or allomorph will emerge as dominant, while the others will be more specialized. The dominant variant is the default. Of course, no distribution is completely stable, since no language is perfect.

Morphology differs from phonology in the characterization of the environments in which the rival expressions occur. Most notably, the environment of one or more of the expressions may be lexical, consisting of one or more specific lexemes, as we saw above with *went*, the past tense form of go. Still, in inflection, just as in phonology, one of the competing expressions can usually be characterized as the elsewhere or default variant. The default variant is not normally assigned lexically. In the case of the English past tense, the default variant is -ed, which is assigned by rule.

Brown and Hippisley (2012) provide a computationally implementable general account of complementary distribution in inflectional morphology. The most important mechanism in this account is default inheritance within a network. Default inheritance encodes the system of defaults very elegantly within network morphology. More specific variants or lexical specifications override the default, which emerges where it is not overridden.

What about non-contrastive, non-complementary distribution? Within structuralist phonology, this was included in the category of free variation. Allophones whose distribution could not be predicted were said to be in free variation. The study of systematic non-complementary distribution emerged in the 1960's with William Labov's work on inherent sociolinguistic variation and the idea that non-complementary distribution and statistical variation are general characteristic of languages has gained importance in the study of dialects and sound change. Not so much in morphology. But if we view the distribution of morphological variants as a form of competition, we expect to encounter variable distribution throughout language. This expectation is borne out even in standard English, as we will show. We did not see the variability because we were not looking for it.

There has been very little work on variable distribution in 'standard' languages. This may be because of a mistaken prejudice that variation should be more characteristic of nonstandard varieties. The work of Anna Thornton (2011), who has discussed the concepts of 'overabundance' and 'cell-mates' in Modern Standard Italian, is a notable exception.

Because of the history of the field, most syntacticians and morphologists, when they do encounter variation, have either dismissed it as performance rather than competence or attempted to reduce it to a discrete system. Good examples of this method can be found in Adger's (2006) discussion of a Scottish dialect and Adger and Trousdale's (2007) treatment of variable agreement in a British dialect. Hudson (2007) criticizes Adgers's method and concludes that no account of the Scottish data that fails to incorporate inherent variability at its heart can be sastisfactory. The real question is whether languages encode phenomena statistically at their heart.

We turn now to Standard English and to the two forms of the comparative with which we began, which we will call *periphrastic* and *suffixal*. Conventional grammatical wisdom for the last century has been that they are in complementary distribution. We will show, quite to the contrary, that the two competing strategies have been at work side by side for millennia, with no resolution on the horizon. From the perspective that we defend here, such unresolved competition is exactly what we expect, at least in some instances. Languages are not fully organized systems où tout se tient.

The comparison of adjectives (degree) in English is famously expressible by two means, the suffixes -er, -est and the adverbs *more*, *most*.² As the quotations from *The International Herald Tribune* and *The New York Times* show, despite the best efforts of prescriptive grammarians, the two are rivals in non-complementary distribution. Some of the earliest English-language documents show that the rivalry has gone on for at least 1200 years and each of the two strategies has roots in Indo-European.

This sort of competition is common in derivational morphology, where rival suffixes rise and fall with great frequency, as we showed at the last Décembrettes meeting (Lindsay & Aronoff 2013). What is remarkable about adjective comparison in English is that it bears all the hallmarks of being inflectional rather than derivational and examples of systematic rival inflectional forms are vanishingly rare.

Degree morphology is usually considered to be syntactic rather than lexemic and hence inflectional (Zwicky 1989). The adverbial expressions of degree is accordingly termed *periphrastic* morphology (Chumakina & Corbett (2012). Periphrasis is usually thought of as a syntactic method that fills cells in a lexemic paradigm alongside those filled by morphology. Periphrasis is thus lexical in that it furnishes members of a lexeme's paradigm but also nonmorphological, in that it is not a part of morphology in the narrow sense: it does not form words. The most commonly cited periphrastic example is the Latin perfect passive. Most of the forms in a Latin verb's paradigm consist of single words: *laudō* 'I praise'; *laudābitur* 'she will be praised'. The exception is the perfect passive, which is a two-word form consisting of the present tense of the verb *esse* 'be' preceded by a passive participle: *laudātus est* 'he was praised'.³ Sadler and Spencer (2000) provide an analysis of this construction in which "the syntax actually fills cells in the morphological verbal paradigm" (*ibid*: 73). The same seems to hold for the English periphrastic comparative: it is part of the paradigm of adjectives.

 $^{^2}$ In what follows, we will use the blanket term *comparative* to include both the comparative and superlative forms.

³ This same periphrastic form also functions as the active perfect of deponent verbs (Flobert 1975).

What evidence do we have that the English comparative is inflectional? First, like all good inflection, degree morphology does not change the category of its base, which remains an adjective. Also, the expression of degree is always available for a gradable adjective; a degree form never seems novel in the way a newly derived word may. Individual degree forms do not drift semantically and degree forms have no special connotations. For all these reasons, Zwicky (1989) concludes that we are dealing here with inflection.

These two means of forming the comparatives appear at first glance to be in complementary distribution, like other competing inflectional realizations. Words of one syllable generally take the suffixed forms. Two-syllable words ending in -y and -le are said to take the suffix but other two-syllable words do not: *sillier, livelier, nobler,* but **foolisher, *rampanter.* Adjectives that can appear in the predicate only must take the periphrastic form: **awarer, *afraider, *contenter.* Elsewhere, only periphrastic forms occur, notably with adjectives of more than two syllables. But there are many exceptions and uncertainties. Some one-syllable words avoid suffixation: **loucher.*

Most uncertainties occur among two-syllable words. Some two-syllable words ending in unstressed syllables other than -y also prefer suffixation: *narrow, clever.* But some two-syllable words prefer periphrasis: *vapid.* Zwicky, noting the variability, quotes Evans and Evans (1957): "But this is a description of what usually happens, not of what must happen. Mark Twain must have agreed, for he wrote: the confoundedest, brazenest, ingeniousest piece of fraud." According to Jespersen (1949: 347) "a good deal is left to the taste of the individual speaker or writer" and that the "rules given in ordinary grammars are often too dogmatic." Individual linguists differ in their intuitive judgments about individual words. Zwicky notes that disyllables with tense vowels in their final syllable take the suffix: *profounder, politer, sincerer, obscurer, shallower.* My own intuition is that disyllables with tense vowels in their final syllable accept both suffixed and periphrastic forms.

Those who have looked at real data note that "Disyllabic words have always been subject to more variation." (Kytö & Romaine 2000: 180) Frequency also plays an important role among two-syllable words (Graziano-King 1999). A number of authors claim that there are stylistic differences between the alternatives, with the periphrastic form more common in written registers, but this has not been established empirically.

The one class of words for which there is certainty is participles, which categorically allow only the periphrastic form. The restriction holds even for adjectival participles: *a more rousing/*rousinger cheer*; *a more fitted/*fitteder bodice*. The historical origins of this restriction are well known. All present participles are at least two syllables long, because the

affix *-ing* appears exceptionlessly with present participles. Most past participles are also more than one syllable long.⁴

The two strategies are very old, with the comparative and superlative suffixes dating back to Indo-European (Sihler 1995). Latin used the periphrastic expressions *magis* and *plus* for participles and other non-adjectival forms and the suffixes for adjectives: *longus* 'long'; *longior* 'longer'; *longissimus* 'longest'. Vulgar Latin lost the suffixed forms early on and they are absent from all modern Romance languages (Herman 1967). By contrast, Germanic languages other than English including Modern German, have only the suffixed forms, except for participles. Thus, most of English's geographically close relatives have resolved the conflict by picking one strategy or the other.

Gonzales-Diaz (2008) cites a number of Old English examples of the periphrastic construction with the adverbs *ma*, *bet*, and *swidor*: *Qaet hi syn sylfe ma gode donne odre men* 'that they themselves are more good than other men'.⁵ There are even examples of double periphrastics in Old English of the sort that persist today in a few expressions like *more better*. Old English examples of the adverbs in question with (usually past) participles in predicate position are attested. The periphrastic use of more increased in Middle English, with support from French and Medieval Latin. According to Kytö and Romaine (2000) the modern distribution developed gradually over a period of centuries.

What is the actual distribution of the two constructions and are they in complementary distribution? We will review the two recent major historical and synchronic studies of the question, both of which conclude that, though one predominates in certain environments, the distribution is not discrete and has never been. Gonzalez-Diaz is based mostly on the British National Corpus (henceforth BNC). Table 1, drawn from this work, contains all comparative forms of two-syllable adjectives ending in -ly, -y, and -le in the BNC. These sets are supposed to prefer the inflectional form, according to the standard descriptions, but the actual data show that they are split almost down the middle and that there is no difference between the three subclasses.

⁴ This distribution may have contributed to the general distribution of the inflected and periphrastic forms. Since participles occur only in the periphrastic form and participles are generally more than one syllable long, it is likely that a word more than one syllable long will show the periphrastic form.

⁵ This particular example is notable for the periphrastic comparative form of *gode* 'good' rather than the suppletive *beter* 'better'.

Туре	Adjective	Inflectional	Periphrastic
-ly class		119 (51%)	115 (49%)
	Lonely	11	4
	Lively	55	37
	Lowly	6	7
	Friendly	47	67
-y class		190 (53%)	175 (47%)
	Shaky	7	1
	Weighty	19	6
	Clumsy	12	4
	Glossy	7	3
	Empty	11	5
	Xosy	20	14
	Scary	8	6
	Angry	38	29
	Risky	39	41
	Sleepy	3	6
	Ready	23	52
	Cloudy	3	8
Syllabic /l/ class		31 (63%)	18 (37%)
	Noble	20	11
	Feeble	11	7
TOTAL		340 (53%)	308 (47%)

Table 1: Actual occurrences of con	mparative forms of adjectives ending	g in
<i>—ly, -y,</i> and <i>—le</i> in the BNC	(data from Gonzalez-Diaz 2008)	

It has often been suggested that the periphrastic option is more likely to be used predicative position, while the inflectional form is more common in attributive position (immediately before a noun). The postpositive position following the noun is fairly rare, occurring only about 3% of the time. Again, Gonzalez-Diaz's findings from the BNC are instructive, as shown in Table 2.

Position	Inflectional	Periphrastic	Total
Attributive	152 (45%)	72 (23%)	224 (35%)
Predicative	179 (52%)	224 (73%)	403 (62%)
Postpositive	9 (3%)	12 (4%)	21 (3%)
Overall	340 (100%)	308 (100%)	648 (100%)

 Table 2: Inflectional and periphrastic forms according to syntactic position (data from Gonzalez-Diaz's analysis of the BNC)

We see that periphrastic forms are more that three times more likely to occur in predicative position than in attributive position. Inflectional forms, however, show no preference for the attributive position, appearing in the predicative position slightly more frequently, though this distribution is likely due to the fact that overall the comparative is used in the predicative position almost twice as often as in the attributive.

One more variable that Gonzalez-Diaz explores is the presence or absence of a *than-phrase*. Overall, only 116 of the 648 examples in the last table are followed by a *than-phrase*. All but 13 of these appear in predicative position and 58% of those are inflected rather than periphrastic. We find the opposite ratio when there is no *than-phrase* in the same predicative position. So overall, the presence of the *than-phrase* appears to favor the use of the inflected form.

How do we make sense of this system? Whatever is going on, it does not resemble complementary distribution. If that were so, we would find the inflectional forms overwhelmingly in attributive position and the periphrastic forms equally overwhelmingly in predicative position. If, instead, we conceive of the inflectional and periphrastic forms as entirely independent entities that happen to compete for the same resource, adjectives, in order to accomplish the same goal, realizing English comparative morphosyntax, then there is no expectation that the two will be in complementary distribution. They are competitors. In some arenas, such as one-syllable words, one will be dominant. In other arenas, such as words of three or more syllables or participles, the other will be dominant. But, as Gonzalez-Diaz has shown in beautiful detail, the competition between the two entities still rages when the food source is two-syllable words. What is remarkable is only that this battle has gone on for so long, at least a millennium, and shows no sign of abating. We see similar prolonged contests between suffixes in English derivational morphology (Lindsay & Aronoff 2013), but not in inflection, where there is usually one overwhelming choice or, if there are several, they sort themselves out into inflectional classes, accompanied by a few lexical exceptions. We have no explanation for why the English comparative remains such a hotlycontested battleground, but the incontestable fact that it does provides striking support for a framework in which competition plays a central role.

In the same year as Gonzalez-Diaz, Martin Hilpert (2008) did an independent logistic regression analysis of comparative forms in the BNC and found a number of additional significant factors, including individual word frequency and the ratio of the comparative to the positive form of any given word, both of which favored the inflected form. Hilpert confined himself to 247 "alternating" adjectives, those that appeared with both inflected and periphrastic forms. He found a number of significant variables of several kinds that had been identified in previous studies as candidates, most of them

phonological and a few syntactic. Almost all the effects he found were gradient.

Hilpert concludes with the following remark:

The results of the present study suggest that the comparative alternation is governed by functionally motivated factors as well as by formal phonological factors that do not necessarily reflect such a motivation. This may seem a provocative statement to the functional linguistic community, and indeed it is intended to be a strong hypothesis which I hope others will attempt to falsify. *(ibid.:* 413)

Unfortunately, Hilpert does not explain what he means by functionally motivated factors or functional linguistics. He makes it clear, though, that his goal is "to predict the distribution of the two variants with a high degree of accuracy" (p. 412). If he means predicting which of the two will occur in a given specific environment, then it is precisely here that we part company with Hilpert and, we suspect, most linguists. The key term here is *variant*. For us, these are not variants but rather competitors and as such there is no reason to believe that their distribution should be complementary in the way that linguists have come to expect of variants. There may be systematic pressure for the distribution to become complementary but these two have staunchly resisted that pressure for a long time.

Once we frame the discussion in terms of competition, we can shift our focus to the opposite cases from those that interest Hilpert and everyone else, *balanced pairs*, in which the inflected and periphrastic forms of a given adjective are close to equal in their numbers of occurrences in a corpus. Balanced pairs have not received attention in the literature on comparatives because previous researchers have all looked for complementary rather than identical distribution, but they provide clues to the environments in which the two rival strategies are most competitive, which is precisely what our new point of view directs our attention to.

A single balanced pair provides only anecdotal evidence and so we need a way to find many of them. A natural source was the Google Books N-gram Corpus, which contains 500 billion part-of-speech tagged words in over 5 million books published since 1500. We restricted our search to books published in English between 1900 and 2012, in order to control for possible diachronic (as well as orthographic) effects. We found 3551 adjectives used in comparative constructions. In each pair, we compare the number of tokens for each form; the form with the most tokens is the 'winner' for that pair. Looking at the total number of winners for each form, a pattern emerges. We have only begun to explore the data and it is a little dirty, in spite of the part-of-speech tagging. False positives include *phase, skid, text, report,* and *bargain.* We consider here only the 972 pairs in which the token count of each member of the pair meets a threshold of 500, in order to limit the

number of false positives, though this is not completely foolproof. Still, we find comfort in large numbers. Indeed, we find that the ratio of -er winners to *more* winners stays relatively constant no matter the threshold: between 1.78 and 2.05 when considering a threshold from 0 to 1000. The difference in tokens between the winner and loser of each pair differed by at least an order of magnitude in 41.2% of cases.

The greatest imbalance tends to appear with monosyllables. They strongly favor -er, just as Hilbert found. In our case, -er was favored by a ratio of 3.5 in monosyllables; that ratio becomes 1.25 with polysyllabic adjectives. Polysyllables ending in -y favor -er 220 to 97 (a 2.27 ratio), while all other polysyllabic pairs strongly favor *more*: 105 to 53 (a 1.98 ratio). It is in the latter category where *more* seems to find its niche.

Disyllables not ending in -y rarely make the 500 hit threshold but go both ways. The forms that prefer suffixation are mostly trochaic (e.g. *shallow*, *narrow*) but not all trochees prefer suffixation. Even some monosyllables favor periphrasis, e.g. *more prone* > *proner* (difference log 2.667). But the word *prone* appears almost exclusively in predicate position, supporting the importance of this factor. Disyllables ending in -y tend to have the most balanced distribution. *Blocky, leaky, lonely, scaly, starry,* and *haughty* are among the 10 most balanced. Many monosyllables are also among the most balanced (contradicting most previous assertions): *sour, terse, lewd, sly, ripe, odd, cute, stark, mute, frank*.

In any competition-based account, there are no a priori conditions on what will constitute a viable environment. A viable environment can only be defined a posteriori as one is which something has been found to thrive. Because true competition is never head-to-head, we should expect to find environments where two or more rivals thrive, as revealed here through balanced pairs. Disyllables ending in -y constitute an example of such an environment. This is the first linguistic example known to us of two competitors both thriving in a specific environment. More should be out there but we linguists will to change our research methods in order to find them.

Overall, we have identified a number of environments in which one or the other competitor is stronger, and also environments in which the two are close competitors. Inflection is more likely with monosyllables although not with very infrequent or phonologically marked lexemes. Inflection is more likely in attributive position and less likely in predicative position. Both options are found with disyllables ending in -y but with clear lexical effects.

Most generally, there is no complementary distribution. This, we believe, is the clue to the mystery of the English comparatives. In language, when two realizations compete for the same resources, one usually emerges as the default and the other either disappears, changes its meaning, or retreats to a special niche (Brown & Hippisley 2012; Aronoff 2013). That is one major aspect of the organization of languages. Here, for whatever reason, none of these developments has occurred. The result is a standoff in which each party to the battle has a substantial territory, with occasional incursions into the other's feeding ground. We see no resolution on the horizon.

References

- Adger D., G. Trousdale 2007. "Variation in English Syntax: Theoretical Implications." *English Language and Linguistics* 11: 261-278.
- Adger D. 2006 "Combinatorial Variability." Journal of Linguistics 42: 503-530.
- Brown D., A. Hippisley 2012. *Network Morphology*. Cambridge: Cambridge University Press.
- Chumakina M., G. Corbett, eds. 2012. *Periphrasis: the Role of Syntax and Morphology in Paradigms*. Oxford: Oxford University Press.
- Dal G 2003. Productivité morphologique: definitions et notions connexes. Langue Française 140: 3-23.
- Evans B., C. Evans 1957. A Dictionary of Contemporary American Usage. New York: Random House.
- Fitch T 2010. *The Evolution Of Language*. Cambridge: Cambridge University Press.
- Flobert P 1975. Les Verbes Déponents Latins Des Origines à Charlemagne. Paris: Sorbonne.
- Fradin B., G. Dal, N. Grabar, S. Lignon, F. Namer, D. Tribout, P. Zweigenbaum 2008. *Langages* 171: 34-59.
- Gonzalez-Diaz V 2008. English Adjective Comparison: A Historical Perspective. Amsterdam: John Benjamins.
- Grabar N., G. Dal, B. Fradin, N. Hathout, S. Lignon, F. Namer, C. Plancq, D. Tribout, F. Yvon, P. Zweigenbaum 2006. Productivité quantitative des suffixation par -ité et -Able dans un corpus journalistique du français. In Actes de la 13e Conférence Annuelle sur le Traitement Automatique des Langues Naturelles (TALN-2006). Louvain: Presses Universitaires de Louvain, 167-175.
- Graziano-King J 1999. Acquisition Of Comparative Forms In English. Ph.D. Dissertation, CUNY.
- Herman J 1967. Le Latin Vulgaire. Paris: Presses Universitaires de France.
- Hilpert M 2008. "The English Comparative Language Structure and Language Use." *English Language and Linguistics* 12: 395–417.
- Hudson R 2007. "Inherent Variability and Minimalism: Comments on Adger's 'Combinatorial Variability'." *Journal of Linguistics* 43: 683-694.

- Hurford J 2011. The Origins Of Grammar: Language In The Light Of Evolution. Oxford: Oxford University Press.
- Jespersen O 1949. A Modern English Grammar on Historical Principles. Part VII: Syntax. London: Allen & Unwin.
- Kottak C. 2008 *Mirror for humanity: a concise introduction to cultural anthropology* (6th edn.). New York: McGraw-Hill.
- Kyto M., S. Romaine 2000. "Adjective comparison and standardization processes in American and British English from 1620 to the present" in L. Wright (ed.), *The Development of Standard English*, 1300-1800. Theories, Descriptions, Conflicts. Cambridge: Cambridge University Press, 171-194.
- Lindsay M., M. Aronoff 2013. "Self-Organizing Morphological Systems." In F. Montermini, G. Boyé, J. Tseng (eds.), Selected Proceedings of the 7th Décembrettes. München: Lincom Europa, 133-153.
- Pike K. L 1967. Language In Relation To A Unified Theory Of The Structure Of Human Behavior. The Hague: Mouton.
- Pinker S 1999. Words And Rules. New York: Basic Books.
- Richerson P., R. Boyd 2005. Not by Genes Alone. Chicago: University of Chicago Press.
- Sadler L., A. Spencer 2000. "Syntax as an exponent of morphological features". *Yearbook of Morphology* 2000: 71-96.
- Sapir E 1921. Language. New York: Harcourt and Brace.
- 1933. "La réalité psychologique des phonemes." Journal de Psychologie Normale et Pathologique 30: 247-265.
- Saussure F. de 1960. Cours de Linguistique Générale. Publié par Charles Bally et Albert Sechehaye avec la Collaboration d'Albert Reidlinger (5th edn.). Paris: Payot.
- Sihler A 1995. *New Comparative Grammar Of Greek And Latin*. Oxford: Oxford University Press.
- Zwicky A 1989. Quicker, more quickly, *quicklier. *Yearbook of Morphology* 2: 139-173.

VERBAL REDUPLICATION IN SINITIC^{*}

Giorgio Francesco Arcodia(1), Bianca Basciano(2) & Chiara Melloni(3) (1) Università di Milano-Bicocca (2) Università Ca' Foscari Venezia (3) Università di Verona

Abstract

The main aim of this paper is to underpin the connection between the semantic relationship binding the constituents of verbs and the formal and semantic properties of their reduplication in Sinitic. We will first discuss in detail verbal and adjectival reduplication in Standard Mandarin, the best described Chinese language; we also collected data on adjectives, in order to compare them to verbs. Then, we will discuss data from a convenience sample of twelve Chinese 'dialects', representing the eight major groups of Sinitic, comparing them to Mandarin. We will show that whereas the ABAB reduplication pattern often has a (counter-iconic) diminishing meaning and appears as close(r) to syntax, being also sensitive to the aspectual properties of the base, the AABB pattern always has an increasing function, regardless of the word class of the base, and it is a phenomenon conditioned by morphological factors, being sensitive to the relation holding between the constituents of the base verb.

1 Introduction

The topic of reduplication in Chinese has been investigated in depth in the literature (see *e.g.* Li & Thompson 1981, Tang 1988, Zhu 2003, Tsao 2004, Wang & Xie 2009, Xu 2012, *inter alios*). Many word classes, including nouns, classifiers, verbs and adjectives undergo full reduplication in Chinese

^{*} Traditional characters have been used as a default for Chinese. The romanisation system used for (Standard) Mandarin Chinese is *Hanyu Pinyin*, whereas for other Chinese varieties the transcriptions are given as provided by the sources. When no transcription is provided, we will use toneless smallcaps *Pinyin* following Mandarin pronunciation. The glosses follow the general guidelines of the Leipzig Glossing Rules. For academic purposes, Giorgio F. Arcodia is responsible for sections 3 and 4, Bianca Basciano is responsible for sections 1, 2, 2.1 and 2.2, Chiara Melloni is responsible for sections 2.3 and 2.4. Authors' names are alphabetically listed.

languages, with both iconic and non-iconic meanings, as *e.g.* Mandarin 紅紅 *hóng~hóng* 'red~red, very/quite red', and 看看 kàn~kan 'look~look, have a look'. Special attention has been accorded to adjectival and verbal reduplication, not only in Standard Mandarin, but also in the so-called 'Chinese dialects', *i.e.* Sinitic languages other than the national standard. In this paper, we will focus on phenomena of *full* reduplication (at the segmental level, *i.e.* disregarding tone change), excluding partial reduplication (*e.g.* Mandarin 冷冰冰 *lěng-bīng~bīng* 'cold-ice-ice, ice-cold'; *cf.* 冰冷 *bīng-lěng* 'ice-cold') and reduplication involving the addition of other segmental material (as *e.g.* Cantonese 肥 *féi* 'fat' > 肥肥哋 *fèih~féi-déi* 'rather fat, chubby'; Matthews & Yip 2011: 186).

The main aim of this paper is to underpin the connection between the semantic relationship binding the constituents of verbs and the formal and semantic properties of their reduplication in Sinitic. To this end, we carried out a detailed survey of patterns of verbal and adjectival reduplication in (Standard) Mandarin Chinese, the best-described Sinitic language, and we then looked for analogous data in a convenience sample of twelve Chinese dialects, with at least one representative for each of the eight major groups of Sinitic, comparing them to Mandarin. We collected data also on adjectives in order to compare the features of adjectival reduplication with those of verbs and to highlight the connection between form and meaning characterising full reduplication; moreover, the classes of verbs and adjectives are not always well distinguished in isolating languages, including Sinitic varieties (see Dixon 2004).

Our main claim is that there is a very strong correlation between form and meaning/function in reduplication which applies fairly consistently throughout Sinitic. Thus, whereas the ABAB reduplication pattern often has a (counter-iconic) diminishing meaning and appears as close(r) to syntax, being also sensitive to the aspectual properties of the base, the AABB pattern always has an increasing function, regardless of the word class of the base, and the input is conditioned by morphological factors, being sensitive to the relation holding between the constituents of the base verb, but not to its aspectual features. However, there is also considerable variation both within and across individual languages, which shows up chiefly in patterns of monosyllabic reduplication.

This paper is organised as follows. In section 2, we will provide an overview of reduplication in Mandarin, discussing the correspondence between form and function of the attested patterns, the constraints on the input and output of processes of reduplication, and we will propose a syntactic analysis for diminishing reduplication. In section 3, we will discuss data from our sample of Chinese dialects, highlighting the commonalities and the differences

among them and comparing them to Mandarin, showing that many of the generalisations we may draw on the latter apply also to the former. In the last section of this paper, we will summarise our main conclusions and provide some hints for further research.

2 Mandarin data and background

As mentioned in the introduction, reduplication in Mandarin has both iconic and counter-iconic uses. Typically, the *diminishing* (counter-iconic) function is associated with verbs (1), whereas the *increasing* (iconic) function is associated with adjectives (2):

(1)	教	\rightarrow	教教
	jiāo		jiāo~jiao
	teach		teach~teach
	'teach'		'teach a little'
(2)	小	\rightarrow	小小
	xiăo		xiăo~xiāo
	small		small~small
	'small'		'very/really small'

Diminishing reduplication marks the so-called 'tentative' or 'delimitative' aspect (Chao 1968, Li & Thompson 1981, Tsao 2004), meaning to do something "a little bit/for a while" (Li & Thompson 1981:29), to do something quickly, lightly, casually or just for a try; it has the pragmatic function of marking a relaxed tone, casualness (Ding 2010), and thus reduplicated verbs are also used as mild imperatives (see Xiao & McEnery 2004). Increasing reduplication for adjectives indicates a higher degree of liveliness or intensity (see Tang 1988, among others). However, as a matter of fact, increasing reduplication is possible also for verbs, but only if the base is bimorphemic and its constituents are in a relation of coordination. See the example in (3), where the reduplication shows two interrelated actions which are performed alternately, repeatedly.

in and out'
i

This kind of reduplicated verbs, besides expressing pluriactionality or action in progress (see Hu 2006, Ding 2010), can also express vividness (4), or other kinds of more abstract meanings (5), depending on the linguistic context (on the meaning of AABB verbal reduplication, see Hu 2006).

(4)	跑跳 <i>pǎo-tiào</i> run-jump 'run and jump'	\rightarrow	跑跑跳跳 pǎo~pǎo-tiào~tiào run~run-jump~jump 'skip, run about, run and jump in a vivacious way'
(5)	偷摸 <i>tōu-mō</i> steal-touch 'pilfer'	\rightarrow	偷偷摸摸 <i>tōu~tōu-mō~mō</i> steal~steal-touch~touch 'furtively, do a thing covertly'

The distinction between diminishing and increasing reduplication, thus, crosscuts lexical categories, rather than being firmly associated with a word class.¹ Rather, it appears that the two (contradicting) functions of reduplication are associated with a set of formal and selectional properties. This will be the topic of the following two subsections.

2.1 Correspondence between form and function

One of the most striking features of the Modern Chinese lexicon is the prevalence of polysyllabic words, most often disyllabic (see Shi 2002); given that the overwhelming majority of syllables correspond to morphemes in this language, we may say that Chinese words are mostly complex, typically composed of two syllables/morphemes:

(6)	a.	逼供	b.	酸辣
		bī-gòng		suān-là
		force-confess		sour-hot
		'extort a confession'		'hot and sour

Nevertheless, a considerable number of words (especially, very common ones) are monosyllabic/monomorphemic, as those in examples (1) and (2). This distinction is very relevant for Mandarin, because reduplication works in a significantly different way for monosyllabic and disyllabic words. From the formal point of view, the difference between increasing and diminishing

¹ Reduplication of coordinate nouns is also attested, but it is not productive.

reduplication is visible only at the suprasegmental level, in that the reduplicated verb is toneless, whereas the reduplicated adjective always bears the first tone (Tang 1988: 282, Paul 2010: 120; but *cf*. Li & Thompson 1981: 33). However, for disyllabic bases (AB), the difference arises at the segmental level. In the diminishing function, the base is reduplicated as a whole (ABAB):

(7) 休息	\rightarrow	休息休息
xiūxi		xiūxi~xiūxi
'rest'		rest~rest
		'rest a little, for a while'

In the increasing function, each morpheme is reduplicated by itself (AABB), as seen above for coordinated verbs (3-5). This is true for adjectives as well:

(8)	乾淨	\rightarrow	乾乾淨淨
	gān-jìng		gān~gān-jìng~jìng
	dry-clean		dry~dry-clean~clean
	'clean'		'very/totally clean'

Thus, it appears that there is a strong correlation between the function and the form of reduplication. This is very interesting especially because many (if not most) languages do not exhibit such a clear correspondence between patterns and functions in reduplication (Mattes 2007). Moreover, the difference between these two patterns is not only semantic, but also concerns the restrictions on the input and on the output, as we will show in what follows.

2.2 Input and output constraints

As seen above, whereas increasing reduplication involves (a subclass of) adjectives and verbs, diminishing reduplication only allows verbs as input, either monosyllabic or polysyllabic.² Moreover, not all verbs may enter the diminishing reduplication construction. The base verb must be a dynamic and volitional verb (Li & Thompson 1981), *i.e.* it should possess the features [+controlled], [+dynamic], [+durative]; all inherently telic verbs are excluded:

² An adjective such as 高興 *gāoxìng* 'happy' may also reduplicate as 高興高興 *gāoxìng~gāoxìng,* with the diminishing meaning 'have some fun'; this is restricted to those adjectives which may be used as dynamic predicates in Mandarin (basically, stage-level adjectives); see Sybesma (1997), Liu (2010).

- (9) *贏贏那場比賽 (Xiao & McEnery 2004: 155; characters added)
 * yíng~ying nà chăng bĭsài win~win that CLF match 'win that match a bit'
- (10) *喝醉喝醉
 *hē-zuì~hē-zuì
 drink-drunk~drink-drunk
 'get drunk (a bit)'

The diminishing (AA) reduplication of monosyllabic verbs like $\frac{\pi}{i}$ 'come' or $\frac{\pi}{i}$ 'in 'enter' is thus ruled out by aspectual constraints; generally speaking, the delimitative aspectual semantics of the diminishing pattern is incompatible with the *Aktionsart* of accomplishments and achievements. Moreover, stative verbs generally cannot reduplicate (see Tsao 2004).³

As to the output, delimitative aspect turns an unbounded dynamic event into a holistic / temporally bounded event (see Xiao & McEnery 2004). This is apparent if we consider that, differently from the base verb, reduplicated (non-coordinate) verbs are incompatible with the progressive aspect marker $\mathbb{E}\pm zh\dot{e}ngz\dot{a}i$, but are perfectly compatible with the perfective aspect marker $-\vec{J}$ *-le*, which signals completion or termination of an action (Xiao & McEnery 2004, Ding 2010):

 (11) 學習了學習 xuéxí-le xuéxí study-PFV study
 'studied a bit'

Differently from diminishing reduplication, increasing reduplication requires that its base adjectives and verbs have specific structural properties. As for adjectives, increasing reduplication applies both to monosyllabic and to disyllabic bases; however, the AABB pattern requires a disyllabic *and* bimorphemic base, whereas disyllabic monomorphemic words cannot be reduplicated (Paul 2010: 137):

³ However, some stative verbs expressing states of mind which can have a dynamic interpretation, as *e.g.* $\exists \mu i a o j i e$ 'understand', may actually reduplicate (Ding 2010: 283).

(12)	窈窕	\rightarrow	*窈窈窕窕
	yăotiăo		*yǎo~yǎo-tiǎo~tiǎo
	'graceful, gentle'		

It thus appears that here units are handled on a morphemic base, rather than on a prosodic base. Moreover, the possible bases for AABB reduplication are either lexicalized, non-transparent bases (13a), adjectives formed by two morphemes with a similar meaning (13b) or in logical coordination (13c)

(13) a	a.	馬虎	\rightarrow	馬馬虎虎
		mă-hu		mă~ma-hū~hū
		horse-tiger		horse~horse-tiger~tiger
		'careless, casual'		'careless, casual (stronger)'
ł	э.	快樂	\rightarrow	快快樂樂
		kuài-lè		kuài~kuài-lè~lè
		pleased-happy		pleased~pleased-happy~happy
		'happy'		'very/really happy'
c	с.	高大	\rightarrow	高高大大
		gāo-dà		gāo~gāo-dà~dà
		tall-big		tall~tall-big~big
		'tall and big'		'(very) tall and big'

As to verbs, increasing reduplication has no aspectual requirements on the base verb, since all kind of verbs, including inherently telic verbs like $\pi l \dot{a} i$ 'come', $\pm j in$ 'enter' or $\pm ch \bar{u}$ 'exit' are allowed (see *e.g.* 3), but requires bases with specific structural properties. As a matter of fact, AABB increasing reduplication is possible only for coordinated complex verbs, the constituents of which may be either in a relation of logical coordination (14a), synonyms (14b) or antonyms (see above, ex.3):

ì.	說笑	\rightarrow	說說笑笑
	shuō-xiào		shuō~shuō-xiào~xiào
	talk-laugh		talk~talk-laugh~laugh
	'talk and laugh'		'talk and laugh continuously'
).	叫嚷	\rightarrow	叫叫嚷嚷
	jiào-răng		jiào~jiào-răng~răng
	call-shout		call~call-shout~shout
	'shout, howl'		'shout repeatedly'
	l.	 說笑 shuō-xiào talk-laugh 'talk and laugh' 叫嚷 jiào-răng call-shout 'shout, howl' 	a. 說笑 → $shu\bar{o}$ -xiào talk-laugh 'talk and laugh' b. 叫嚷 → jiào-răng call-shout 'shout, howl'

Note that in (14a-b) the bases of reduplication are existing verbs, but this is not necessarily always the case, as e.g. 走走停停 zǒu~zǒu-tíng~tíng 'walk

and stop' (there is no corresponding base verb 走停 $z \delta u - t i ng)^4$. Also, a coordinate compound made of synonymous constituents as 討論 $t \delta o - l u n$ 'discuss-discuss = discuss' reduplicates as 討論討論 $t \delta o - l u n \sim t \delta o - l u n$, meaning 'discuss a little' (i.e. diminishing, rather than increasing). Arguably, this happens because such highly lexicalised word forms are unanalysable for the average speaker, and hence are treated as non-coordinate (on lexicalisation, see Packard 2000). We will get back to this in §3.1.

Thus, the constituents of increasing AABB reduplication must be either coordinate (and non-lexicalised, in the case of verbs) or lacking a semantic/structural head. Notably, disyllabic adjectives with a modifier-head structure, such as 雪白 xuě-bái 'snow-white', reduplicate as ABAB (雪白雪白 xuě-bái~xuě-bái), with an increasing meaning. This is actually the only exception to the form-function identity between ABAB reduplication and diminishing meaning in Mandarin. Moreover, adjectival reduplication normally requires as input a [+gradable] base (either monosyllabic or disyllabic), thus a non-gradable adjective such as 方 fāng 'square' cannot reduplicate (*方方 *fāng~fāng; Paul 2010: 139, fn. 19); modifier-head adjectives are the only non-gradable adjectives which may reduplicate. It is also worth remarking that 'rhotacisation', a morphophonological phenomenon consisting in the addition of a retroflex approximant at the end of a word, occurs after the reduplicated adjective in AABB reduplication (高高興興兒 gāo~gao~xìng~xìng-r 'really happy'), but after each AB in the case of modifier-head compound adjectives (雪白兒雪白兒 xuě-bái-r~xuěbái-r; see Lee 2012). These facts suggest that adjectival AA/AABB and ABAB reduplication are two distinct phenomena, albeit both morphologically conditioned (*i.e.* they have specific structural requirements on the base).

To sum up, it appears that increasing reduplication is sensitive to the morphological makeup of its input, rather than to any semantic feature. Also, we showed that there appears to be an exception to the strong correspondence between form and function in Mandarin reduplication, which involves a very peculiar subclass of adjectives. In the next subsection we will focus on verbs, outlining an analysis of the data discussed here.

⁴ One could argue then that verbal AABB reduplication is the result of the coordination of two reduplicated verbs, [A~A] [B~B]. However, note that reduplication of monosyllabic verbs expresses a delimitative meaning, so the coordination of two monosyllabic reduplicated verbs should result in delimiting semantics. Moreover, note that telic verbs like 進 jin 'enter', as said above, cannot reduplicate by themselves, * 進進 jin~jin~(cf. ex. 3).

2.3 Analysis: diminishing reduplication

In the previous section, we showed that despite superficial similarities the diminishing and increasing patterns of Mandarin reduplication are characterized by different properties that make a unified analysis of the two phenomena untenable. Quite to the contrary, we purport the view that, whereas increasing reduplication is sensitive to *morphological* constraints and its building blocks allegedly are chunks of structure below the X° level, diminishing reduplication is a *syntactic* phenomenon, which combines larger structures within the *v*P domain.

This rationale is motivated primarily by the separability of the verbal complexes obtained via the diminishing pattern, which challenges the alleged syntactic atomicity or lexical integrity of words (see Lapointe 1979, *inter alios*). Specifically, in (11), we remarked that the aspect marker $-\overline{J}$ *-le*, usually occurring at the rightmost side of verbs, is 'interfixed' between the base and the reduplicant, and other elements can in fact occur between them (see Basciano & Melloni 2013). Furthermore, under the acknowledged view that aspectual properties are *syntactically* encoded,⁵ the range of aspectual constraints described in the previous section for the input verbs is unexpected if one treats this pattern as a strictly morphological phenomenon. Besides this, and differently from increasing reduplication, there is a lack of purely morphological constraints that impose specific requirements on the structural makeup of input verbs.

We thus propose a syntactic analysis of diminishing reduplication in the constructionist framework put forth by Ramchand (2008), which is based on a syntactic decomposition of the event structure ('first phase syntax'). In this system, the event structure can be decomposed into a maximum of three subevents, each represented with its own projection, ordered in a hierarchical causal embedding relation: the causative subevent (*initP*), which introduces the causation event and the verb external argument hosted in its specifier (*i.e.* the subject of cause or *initiator* in Ramchand's theory); the process subevent (*procP*), which specifies the nature of the change or process and introduces the entity undergoing the change or process (*i.e.* the subject of process or *undergoer*);⁶ the result subevent (*resP*), which provides the *telos* or *result state* and hosts the subject of result or *resultee*.

⁵ Since the early 1990s, a number of studies have advanced the hypothesis that thematic and aspectual requirements of events are directly encoded in the syntax: see among others, Travis (2000, 2010); Borer (1994, 2005); McClure (1995); Ramchand (1997, 2008).

⁶ The *proc*P is the heart of the dynamic predicate, since it represents change through time and it is present in every dynamic verb.



In this framework, lexical items specify the syntactically-relevant information by means of a category label or 'tag', which permits their insertion in the eventive structure, and may have multiple category features. Telicity in this framework can arise in two ways: either it is lexically encoded (in Ramchand's terms, the lexical item is marked by [res] feature) or it is compositionally obtained in *proc*P by means of a spatial bounded path (usually acknowledged as 'incremental theme') in the complement position.⁷

The present analysis rests upon the main hypothesis that diminishing reduplication spells out two copies of the same element within the vP domain. Let us now see the details of our proposal. First, it should be premised that the group of Chinese verbs that can undergo diminishing reduplication are easy to delimit in Ramchand's framework since – being activities and accomplishments taking a non quantized object – they are lexically marked by the tags [init, proc]. All verbs tagged with [res] (*i.e.* achievements and resultatives) are excluded. Relevant literature (see Xiao & McEnery 2004) advances a purely semantic explanation for the incompatibility between inherently telic verbs and diminishing reduplication. We contend instead that this fact straightforwardly follows from the inner structure of reduplicated verbs.

It has been noticed that the main semantic function of diminishing reduplication is to delimit the temporal duration of an otherwise unbounded event. We thus claim that the reduplicant adds a [+bounded] temporal path to the [-bounded] situation codified by the base verb. Being a Path of process verbs, we claim that the reduplicant (the verb lower copy) occupies a

⁷ "The complement position of a process head is associated with the semantic relation of structural homomorphism, regardless of the category of that complement." (Ramchand 2008: 47). We refer the reader to Ramchand's (2008) book for further details.

dedicated syntactic position in the complex structure of vP, *i.e.*, it is the complement of the Process head in Ramchand's (2008) framework.

This analysis implies a *structural* incompatibility between the reduplicant and the resP, which sits in the complement of procP; hence it syntactically accounts for the aspectual restrictions exclusively ascribed to the semantic level in previous analyses. As mentioned in section 2, limited temporal duration is not the only semantic value conveyed by diminishing reduplication: besides this, a number of related semantic effects such as casualness, tentativeness, etc. are also listed as possible meanings of reduplication in reference grammars of Mandarin. As a matter of fact, the aspectual constraints on input verbs hold in all the instances of diminishing reduplication, independently from the overall semantics of the output. We argue, however, that the aspectual restrictions on input verbs are hardly justified in an account that derives them from the lexical-semantic incompatibility between the inner temporal constitution of the base and the varied (hardly predictable) semantics of the reduplication template. On the other hand, the derived semantic nuances of diminishing reduplication are structurally justified in the present analysis provided that they are analysed as shifted semantic correlates of the core meaning of the procP - Path template, *i.e.* temporal boundedness.

2.3.1 Reduplicants as objects

In order to grasp the technical details of our analysis, let us start from the 'simplest' case of an intransitive verb, $\neq z \delta u$ 'walk':



The structure in (16c) shows that the reduplicant, occupying the verb complement, turns a basically unergative verb into a transitive one, a solution

which is reminiscent of Hale & Keyser's (1993) understanding of unergative verbs. Thematically, the object is not a Patient, but a temporal Path which provides a temporal boundary for the event.

If all reduplicated verbs undergo a kind of 'transitivization', the obvious issue to address concerns the position of syntactic objects of inherently transitive bases, provided that the reduplicant should cause the unavailability of the complement of *proc*P. In Ramchand's framework, however, 'objects' of the verb can originate in different places in the *v*P. Within the *proc*P, they can be either Undergoers (*i.e.* the entity undergoing the change or process), which originate in the specifier of *proc*P, or Paths in the complement position of *proc*P (see above). Let us consider the case of verbs with Undergoers first. It is worth noting that Undergoers cannot measure out the event, since they are not incremental themes, but do undergo the change described by the event. Therefore, no incompatibility arises in cases such as 試試 *shi~shi* 'try~try, try on (shortly, for a while)' because the syntactic object and the reduplicant occupy different structural positions:

(17) a. 我试试衣服
wǒ shì~shi yīfu
lSG try~try dress
'l'll try the dress on'
b. *init*P (causing projection)
我 wǒ T
试 shi 'try' procP (process projection)
衣服 yīfu 'dress'
试 <shi 'try'> 试 shi

Ramchand (2008) observes that some transitive verbs are characterized by having an object which is not the Undergoer of the Process but a Path (more typically acknowledged as incremental theme). With respect to the traditional Vendlerian taxonomy, verbs taking a Path as object are accomplishments.

According to Ramchand, when the verb takes a Path object, the property mapped onto the process is inherent to the DP and does not change; the homomorphism with the process of the event is established via the scalar structure of the inherent property, and the process is defined by its progress through the scale provided by the Path object. This class includes creation/consumption (or ingestive) verbs, like *eat*, *drink*, *read*, *write*, etc. Ramchand assumes that in these cases the specifier position of *proc*P is not

filled by the direct object of the verb, which is a Path, and that it is the Initiator itself which fills the Undergoer position too, given its status as a continuous experiencer of the process. See the example below:



As for diminishing reduplication, accomplishment verbs do not behave in the same way. Typically, they cannot undergo reduplication when they are combined with a quantized object; however, when taking a non-quantized object, accomplishments can undergo reduplication too.

 (19) 喝喝茶
 hē~he chá drink~drink tea
 'have some tea'

Assuming that the object is a Path in the complement position of *proc*P, we should exclude the possibility that the reduplicant is a Path itself, since the complement position is already occupied by the object (see ex. 18 above). We thus advance a tentative hypothesis which might be able to capture their structure and semantics. Interestingly, a cross-linguistic look shows that reduplicated verbs combined with incremental themes are close, at the semantic level, to light verb constructions (henceforth, LVCs) such as the following (attested in many Romance and Germanic languages):

(20) Italian

a. *fare* (**bere*) *una bevuta di tè* do-INF (*drink-INF) a-SG.F of tea
English
b. *take a drink of tea*

In these cases, a semantically light verb (such as do, make, take, give, etc.) takes as its object a complex DP that, beyond codifying the core event semantics, is able to delimit the event temporally; in particular, in (20) una bevuta (di...) / a drink (of...) is a DP headed by an event noun which acts as a measure phrase, able to turn the mass noun 'tea' into a quantized nominal expression. At the vP level, this DP also provides a "boundary" to the unbounded process encoded by the verb bere/drink.

Provided that LVCs are formed by a process verb combined with a DP complement, which in turns embeds a complement (*di tè* is the internal argument of the event noun *bevuta*, from *bere* 'to drink'), we argue that the semantic parallel between LVC and diminishing reduplication can be translated into a syntactic one. Structurally, both *bevuta* / *drink* and the reduplicant \mathbb{B} *hē* 'drink' are Paths able to delimit the event; further, just as *di tè* / *of tea* acts as the complement of *bevuta* / *drink*, in Chinese $\frac{\pi}{K}$ *chá* 'tea' would be the complement of \mathbb{B} *hē* 'drink'.





In this picture, many issues still deserve further understanding; above all, we still lack an assessment of the categorial nature of the reduplicant, which as a verb should not sit in the complement of *proc*P, but as a noun should not be able to license its nominal complement ($\stackrel{\scriptstyle \leftarrow}{\times} ch a$).

We believe however that this line of analysis, whose details are omitted here due to space limitations, offers novel insights on a phenomenon generally ascribed to the lexical/morphological domain. First, it structurally accounts for the counter-iconic semantics of the diminishing/delimiting pattern. Furthermore, it can justify the lack of lexical integrity of the complex since, as a phenomenon affecting the 'first phase syntax' of the verb, diminishing reduplication is not expected to created syntactic atoms. Also, it offers a structural explanation for the incompatibility between Result State and diminishing reduplication: under the present analysis, the result state and the
reduplicant cannot be base-generated in the same structural position. Finally, it predicts the semantics of direct objects of reduplicated verbs, which are never Paths/Incremental Themes; they can be either Undergoers (originated in the specifier position of *proc*P) or complements of the reduplicant itself.

2.3.2 Reduplicants as Cognate Objects

Other evidence in support of the object analysis of reduplicated verbs comes from the heterogeneous class of cognate object constructions (henceforth: COC). It has been remarked (see Chao 1968 and Hong 1999) that Mandarin V-yi-V reduplicating construction can be understood as a kind of COC, sharing many properties of Indoeuropean COCs. Consider the following English standard case of COC:

(22) laugh a (scornful) laugh

A cognate object such as *a* (...) *laugh* possesses the following three characteristics: from the point of view of its morphological form, *laugh* bears the same form as the verb *laugh*; from the point of view of its syntactic function, *a laugh* is the syntactic object of the verb *laugh* (at least according to Massam 1990, Macfarland 1992, and Pham 1999); as to its semantic function, *a laugh* is delimitative, since it temporally bounds the process codified by the verb *laugh* (see Hong 1999: 263). Chao (1968) and Hong (1999) argue that delimitative reduplication is in fact a type of COC, but their claim is limited to those cases where — *yi* (*yī*) 'one' precedes the reduplicant.

(23) 看一看 kàn-yi-kàn look-one-look 'have a look/look for a while'

On the other hand, this analysis does not take into account two interesting facts. First and foremost, there seems to be no difference in semantics between reduplicated forms with and without the numeral -yi ($y\bar{i}$):

(24) 看看 *kàn~kan*look~look
'have a look/look for a while' (*cf.* ex. 23)

Furthermore, $-y\bar{i}$, as a numeral taking a classifier, is often omitted in speech.

(25) 我想买(一)本书
wǒ xiǎng mǎi (yī) běn shū
1SG want buy (one) CLF book
'I want to buy a book'

Therefore, here we put forth the tentative hypothesis that not only monosyllabic verbs reduplicated with — $yi(y\bar{i})$ are instances of COC, but that the COC analysis applies to all instances of diminishing reduplication, which would contain — $yi(y\bar{i})$ covertly or overtly. Under this analysis, the double parallelism arising between DR and COCs is easy to capture; that is to say, both reduplicants and cognate objects provide a (temporal) boundary to the event; syntactically, they can be analysed as complements of the verb, specifically as delimiting Paths of process heads.

2.4 Further remarks

In the previous sections, we outlined a syntactic analysis of diminishing verb reduplication, mainly on the grounds of the fact that reduplication modifies the aspectual structure of the base verb; its base is indeed aspectually constrained, yet not conditioned by morphology. This picture does not take into account reduplication of coordinate verbs though. As we have shown in $\S2.2$, this kind of reduplication is akin to adjectival reduplication in that it expresses an increasing meaning and its input, differently from diminishing verbal reduplication, is conditioned by morphological factors⁸.

Semantically, both diminishing reduplication and increasing adjectival reduplication affect boundedness (intended as gradability for adjectives, see Alexiadou 2010): in particular, reduplication turns a [-bounded] (atelic) event into a [+bounded] one and a [+gradable] adjective into a [-gradable] one. An open question concerns increasing verbal AABB reduplication: can this kind of reduplication too be accounted for in term of boundedness? This hypothesis is intuitively appealing: the typical meanings of this pattern, as shown before, include pluractionality and action in progress, *i.e.* essentially unbounded aspectual profiles (we will get back to this in §3.1). However, at present we do not have an analysis able to account for the semantics expressed by this kind of reduplication. We leave this for further research.

⁸ An analysis of adjectival reduplication is beyond the scope of this paper, but we believe that, though constrained by morphological factors, it should be understood as a phenomenon pertaining to the syntax, rather than to the morphological or lexical module of grammar. We leave this for further research.

3 Beyond Mandarin: reduplication in other Sinitic languages

Sinitic is the largest branch of the Sino-Tibetan family in terms of number of speakers, with a number of dialect groups varying from 7 to 10, according to different classifications (see Kurpaska 2010); most of the variation within Sinitic is found in Central and Southern China, whereas the North of the country is dominated by Mandarin dialects (from which Standard Mandarin Chinese originated). Chinese 'dialects', thus, are not varieties of a unitary language but, rather, varieties *related* to Standard Mandarin, just as Dutch and Swedish are related to English, and should be viewed as distinct objects for comparison (Norman, 2003); the difference is that whereas English, Dutch and Swedish all have a long written history and recognised standard varieties taught in schools and used in media discourse (as well as a number of regional dialects/varieties), within Sinitic only Mandarin and, in a sense, Cantonese are standardised language varieties.

As stated in the introduction, for the purposes of our study, we looked for data on reduplication in twelve Chinese dialects, with at least one representative for each of the eight major groups. In table 1 we provide a list of the varieties considered, together with their affiliation.

Language	Group	Source			
Chengdu	Mandarin	Yang (2005)			
Huojia	Jin	He (1989)			
Xiangtan	Xiang	Zeng (2001)			
Taiwanese	Min	Tsao (2004), Chuang			
Southern Min		(2007)			
Zhangzhou	Min	Ma (1995), Li (2013)			
Gutian	Min	Li (2006), Li (2013)			
Hong Kong	Yue	Matthews & Yip (2011)			
Cantonese					
Taiwanese	Hakka	Lai (2006)			
Hakka ⁹					
Shanghai	Wu	Zhu (2003)			
Wenzhou	Wu	Chi & Wang (2004), Wang			
		F. (2011)			
Suzhou	Wu	Wang P. (2011), Fu & Hu			
		(2012)			
Yanshan	Gan	Lin & Hu (2008)			

 Table 1: Our sample of Chinese dialects

Unfortunately, we do not have data of the same quality as for Mandarin for any of these dialects, since the descriptions are not nearly as detailed, also with considerable variation from dialect to dialect; nevertheless, we will show that some clear tendencies are visible even in the (incomplete) data we could gather.

3.1 Verbal reduplication

Verbal reduplication is found in all the languages of our sample except Xiangtan, in which, according to the description we consulted, only adjectives reduplicate; the (near) absence of verbal reduplication appears to be a common feature of the Xiang group (Wu 2005: 11-12). Generally speaking, in the dialects of our sample the reduplication of monosyllabic and (non-coordinating, non-lexicalised) disyllabic verbs has the same function as in Mandarin, *i.e.* indicating short duration, 'tentativeness' (see §2), and, in the latter case, it follows the ABAB pattern, as in the following example:

⁹ 'Taiwanese Hakka' is used here loosely as a cover term for the Hakka dialects spoken in Taiwan. The transcriptions of the examples represent the Siyen (四縣, Mandarin *Sìxiàn*) variety.

(26) Shanghai (Zhu 2003: 86)

 帮助
 →
 pòngzu

 'help'

幫助幫助 *pòngzu~pòngzu* help~help 'help out a bit'

However, progressive/iterative semantics is also attested for reduplication of monosyllabic verbs in several Chinese dialects; see Fu & Hu (2012) for examples from Min, Wu and Yue dialects, and Wang (2005) for examples from a Mandarin dialect (Taonan). In Wenzhou, reduplication of monosyllabic verbs may mean 'repetition/continuation over a short period of time';¹⁰ in the following example (adapted from Chi & Wang 2004: 150), the actions of 'reading' and 'writing' are performed alternately and repeatedly over a quite long period, but each individual action is performed for a limited time:

(27) 渠東见東见 写写 gi^2 $ts^h j^5 \sim ts^h j^5$ XIE~XIE 3SG.M read~read write~write 'He is reading and writing'

Shi (2007) proposes that progressive/iterative verbal reduplication is a feature distinguishing Southern China from Northern China, and that it reflects the Middle Chinese pattern of verbal reduplication, whereas the diminishing pattern is an innovative feature (see also Fu & Hu 2012). However, all the examples quoted in Shi (2007) and Fu & Hu (2012), as well as those from our sample, involve monosyllabic verbs; hence, although ABAB reduplication *might* in principle have increasing semantics, we could not find any instance of this, and in all the varieties considered, if ABAB reduplication of verbs is possible, it has a diminishing function, as in the following examples (and ex. 26 above):

(28) Zhangzhou (Ma 1995: 127)

a. 修理	\rightarrow	修理修理
siu ⁴⁴ li ⁵³		siu ²² li ⁴⁴ ~siu ²² li ⁵³
'repair'		repair~repair
		'try to fix, repair a bit'
Gutian (Li 2006: 7	1)	

¹⁰ The 'true' delimitative and the iterative/progressive patterns of reduplication in Wenzhou are distinguished by suprasegmental means (*i.e.* different tone patterns; see Chi & Wang 2004, Wang F. 2011).

b.	研究	\rightarrow	研究研究
	yiey ⁴² kiu ²¹		nien ³⁵ kiu ⁵³ ~nien ³⁵ kiu ⁵³
	'study, research	,	research~research
	-		'study a bit, do some research'

We will get back to progressive/iterative reduplication of monosyllabic verbs below.

Just as for Mandarin, AABB reduplication in 'our' dialects typically conveys vividness, iteration and alternation of actions, as in the following examples:

(29) Hong Kong Cantonese (Matthews & Yip 2011: 40)

a.	上落 →	上上落落
	séuhng-lohk	séuhng~séuhng-lohk~lohk
	rise-fall	rise~rise-fall~fall
	'rise and fall'	'go up and down'
С	hengdu (Yang 2005: 85)	
b	商量 →	商商量量
	SHANG-LIANG	SHANG~SHANG-LIANG~LIANG
	discuss-consider	discuss~discuss-consider~consider
	'discuss, consult'	'discuss repeatedly/for a while'

The patterns exemplified here, however, have different degrees of generality; for instance, according to Matthews & Yip (2011), AABB reduplication is found with directional verbs, and they provide no data on other types of verbs; also, according to Yang (2005), the AABB pattern in Chengdu is available only for a small set of verbs.

Yang also claims that the verbs reduplicating as AABB in Chengdu, nearly all made of coordinate (often synonymous) constituents, correspond to ABAB reduplicates in Mandarin, *i.e.* to highly lexicalised compound verbs (see above, §2.2). Nevertheless, with a cursory Google search, we actually found that *e.g.* both 商量商量 *shāng-liang~shāng-liang* (delimitative) and 商商量量 *shāng~shāng-liang~liang* (iterative) are commonly found in written Chinese¹¹, attesting not only to the differences in the perception of the

 $^{^{11}}$ 882,000 hits for the ABAB version and 609,000 hits for the AABB version (11/11/2013).

structure of this word by different speakers¹², but also to the strong connection between the AABB pattern and increasing semantics, on the one hand, and the ABAB pattern and diminishing semantics, on the other hand. One last aspect of verbal reduplication in the dialects of our sample which is worth mentioning is its interaction with the resultative verb construction. In Mandarin, resultative verb compounds, being inherently telic, cannot be reduplicated, as shown by the ungrammaticality of (10) above. In §2.3.1 we provided our analysis of this incompatibility: in diminishing reduplication, the reduplicant and the result state cannot be base-generated in the same structural position. However, reduplication of (monosyllablic) verbs with resultative elements is not uncommon in our dialect sample:

(30) Wenzhou (Wang F. 2011: 60)

a.	逮魚洗洗光生
	DAI YU XI~XI-GUANGSHENG
	OBJ fish wash~wash-clean
	'wash the fish clean'
Та	iwanasa Southarn Min (Chuang

Taiwanese Southern Min (Chuang 2007: 6; characters added)

拍死 →	拍拍死
phah ⁴ -si ²	phah ⁴ ~phah ⁴ -si ²
hit-die	hit~hit-die
'beat to death, kill'	'beat savagely, to death'
	拍死 → $phah^4$ - si^2 hit-die 'beat to death, kill'

According both to Chi & Wang (2004) and to Wang F. (2011), in Wenzhou reduplication with a resultative element is typically found in imperative sentences, as it softens the tone of the request; Chi & Wang also remark that in this construction a reduplicated verb indicates an action which has not yet occurred (irrealis?). According to Fu & Hu (2012), in these sentences the focus is on the result state, whereas the reduplicated verb indicates that the action leading to the result state is carried on (or repeated) for some time. Basing on an extensive cross-dialectal survey, Fu & Hu (2012) suggest that all monosyllabic patterns of verbal reduplication with a progressive/iterative meaning are found in background sentences, which are necessarily followed by another clause (see 31 below), by a resultative complement (30a-b), or by a directional or a quantifier. Their function is to indicate the manner, reason or circumstances of the occurrence of the following predicate or result state, state, state, for the following predicate or result state, whereas the reduplication with a progressive/iterative meaning are found in background sentences, which are necessarily followed by another clause (see 31 below), by a resultative complement (30a-b), or by a directional or a quantifier. Their function is to indicate the manner, reason or circumstances of the occurrence of the following predicate or result state, see the following predicate or result state, the occurrence of the following predicate or result state, the occurrence of the following predicate or result state, the occurrence of the progressive/iterative or the progre

¹² Compare Mandarin 來往 *lǎi-wǎng* 'come and go', which reduplicates as 來來往往 *lǎi~lǎi-wǎng~wǎng* 'go back and forth, come and go in great numbers', and 來往 *lái-wang* 'have contacts with', fully lexicalised (note the neutral tone of the second constituent), which reduplicates as 來往來往 *lái-wang~lái-wang* 'have some contacts with'.

whereas, on the other hand, the function of the result state is to provide a boundary to the continuation of the action.

Hence, it appears that progressive/iterative verb reduplication differs from diminishing reduplication not only because of its meaning, but also because of its aspectual properties: in the former pattern, reduplication apparently *detracts* from the boundedness of the verb, rather than adding a boundary. The analysis we proposed above for Mandarin, thus, cannot apply as such for these cases.

As to Taiwanese Southern Min, whereas Tsao (2004) believes that the reduplication of the verb in the resultative construction indicates 'rapid completion', being thus somehow consistent with a delimitative interpretation (short duration > rapid completion), Chuang (2007) proposes that the actual meaning conveyed by verbal reduplication, here, is 'intensity'. Thus, in (30b), the reduplication of $\frac{1}{10} phah^4$ 'hit' somehow adds intensity to the predicate, indicating "intensification on the action causing a change of state" (Chuang 2007: 84). Interestingly, in Suzhou, a Wu dialect just as Wenzhou, reduplication of monosyllabic verbs appears to work similarly to the latter, indicating continuation of an action in the background (Fu & Hu 2012: 145):

(31) 我打打球,小王来喊啧
WO DA~DA QIU XIAO-WANG LAI HAN ZE
1SG play~play ball young-wang come call PERF
'I was playing [a ball game], when Young Wang came to call me'

However, when the (monosyllabic) verb in a resultative construction is reduplicated, it is said to indicate that the action has been already completed (Wang P. 2011: 332):

(32) 烧烧熟 SHAO~SHAO SHU cook~cook cooked 'Cooked'

Just as in Wenzhou, the reduplicated resultative construction is found chiefly in imperative sentences (judging from the examples provided in Wang P. 2011). In yet another Wu dialect, Yongkang (not included in our sample due to the lack of adequate data), verb reduplication is one of the devices used to express, again, perfective aspect/completion of an action (Huang 1996: 175): (33) 信寄寄就来
 XIN JI~JI JIU LAI
 letter send~send then come
 '(I, she, etc.) will come after sending the letter / (please) come after sending the letter'

Since no context is provided, it is unclear whether (33) is to be understood as a declarative or as an imperative sentence. Note that in Wenzhou the reduplication of a monosyllabic verb, if followed by an aspectual(/modal) particle as $\gtrsim hus^0$, indicating perfect aspect, may mean 'sudden change' (Chi & Wang 2004: 151):

(34) 鸡都死死爻

JI DOU SI~SI huo^0 chicken already die~die PERF 'The chicken has already died'

The reader may have noticed the use of a verb like 'die', which is not allowed in Mandarin delimitative reduplication because of its inherent telicity. Wang F. suggests that $\gtrsim huo^0$ is added only to those reduplicated verbs whose base form indicates non-volitional, instantaneous actions, and the construction indicates "suddenness, broad scope, gravity of the consequences, etc." (2011: 87; our translation). Moreover, according to her analysis, the reduplication of \overline{M} SI contains a "subjective evaluation on the part of the speaker, expressing 'surprise', 'disappointment'" (2011: 71).

3.2 Adjectival reduplication

Adjectival reduplication seems to be even more common than verbal reduplication in our sample: it is attested in each of the dialects considered, and, apparently, it is less restricted. Given that our main concern here are verbs, we shall provide but a few remarks on adjectives, focussing on the comparison with verb reduplication, just as we did for Mandarin above.

One first remark is that, perhaps surprisingly, reduplication of monosyllabic adjectives, which is the structurally simplest pattern, is not available in all the dialects considered: it is apparently not attested at all in Shanghai, whereas in Xiangtan reduplicated monosyllabic adjectives are part of a pattern including other morphemes (*e.g.* 好高不高 HAO-GAO-BU-GAO 'very-tall-not-tall = very tall'; Zeng 2001: 52). Moreover, reduplicated monosyllabic adjectives do not always convey increasing semantics. In Taiwanese Hakka and Southern Min, reduplication of adjectives has a diminishing meaning, whereas triplication has an increasing meaning, as hinted at above:

(35) Taiwanese Hakka (Lai 2006: 490; characters added)

a.	紅	\rightarrow	紅紅	\rightarrow	紅紅
	fung ¹¹		fung ¹¹ ~fung ¹¹		fung ¹¹ ~fung ¹¹ ~fung ¹¹
	red		red~red		red~red~red
	'red'		'kind of red'		'very red'
Та	iwanese	Southern	Min (Chuang 20	07: 2; ch	aracters added)
b.	紅	\rightarrow	紅紅	\rightarrow	紅紅
	ang^5		ang ⁵ ~ang ⁵		ang ⁵ ~ang ⁵ ~ang ⁵
	red		red~red		red~red~red
	'red'		'reddish'		'very red'

Thus, notwithstanding the differences between these two varieties and Standard Mandarin, we still do have a clear correspondence between pattern and function. An analogous distinction is found again in the reduplication of disyllabic adjectives :

(36) Taiwanese Hakka (Lai 2006: 491, fn. 8)

a.	風神	\rightarrow	風神風神
	fung ²⁴ -sen ¹¹		fung ²⁴ -sen ¹¹ ~fung ²⁴ -sen ¹¹
	style-smart		style-smart~style-smart
	'awe-inspiring'		'quite awe-inspiring'
b.	淨利	\rightarrow	淨淨利利
	qiang ⁵⁵ -li ⁵⁵		qiang ⁵⁵ ~qiang ⁵⁵ -li ⁵⁵ ~li ⁵⁵
	clean-sharp		clean~clean-sharp~sharp
	'clean'		'very clean'

The same situation is found in Taiwanese Southern Min (Tsao 2004). Interestingly, verbal reduplication for disyllabic verbs works just as Mandarin both in Taiwanese Hakka and in Southern Min, *i.e.* the ABAB pattern is associated with diminishing semantics, and the AABB pattern with increasing semantics; hence, there appears to be a perfect correspondence between the ABAB pattern and diminishing semantics, and between the AABB pattern and increasing semantics, which equally applies both to adjectives and to verbs. Tsao (2004: 306) suggests that diminishing reduplication for disyllabic adjectives possibly is a Taiwanese innovation; we may speculate that the basis for this was an extension of the ABAB verbal pattern to adjectives may reduplicate as AABB, and he suggests that these cases may be interpreted as the result of Mandarin influence on Taiwanese Southern Min. In the other Min dialects of our sample, *i.e.* Gutian and Zhangzhou, both spoken in mainland China, AABB reduplication has increasing semantics, whereas

ABAB reduplication is apparently found (in Zhangzhou) only for modifierhead adjectives, just as seen above for Mandarin.

Thus, in short, it appears that the strong correspondence between form and function in disyllabic reduplication is consistent across word classes and across dialects. As to monosyllabic reduplication, we discussed some very significant differences in semantic and aspectual features among different dialects, again both for verbs and for adjectives.

3.3 Summary

The picture sketched above for verb reduplication in the dialects of our sample is very complex, if compared both to the situation of adjectives and to what we saw earlier for Mandarin. Moreover, whereas the behaviour of disyllabic verbs is consistent across dialects, monosyllabic verbs are found in several kinds of constructions, apparently expressing incompatible meanings. In the adjectival domain, again, we find much more consistency for disyllabic items than for monosyllabic ones. Thus, there seems to be a general tendency for variation to occur in constructions involving the reduplication of monosyllabic, rather than disyllabic items. The most striking fact, however, is the use of verb reduplication to indicate background open-ended events, contrary to Mandarin, where reduplication typically conveys temporal delimitation /boundedness of the event expressed by the base verb.

As a (tentative) conclusion, we may propose that there are two core semantic values for reduplication, both involving the notion of 'iteration' (incidentally, iconically coded in the construction): iteration over a long/undefined period of time and iteration over a short/defined period of time. The former should reflect the older use of reduplication, and the latter should reflect the 'innovative' uses, as e.g. those of Mandarin. Iteration over a long/undefined period of time may easily be reanalysed as expressing progressive/unbounded semantics, as in Suzhou, whereas iteration over a short period of time may be reanalysed as indicating perfective-like meanings, as rapid completion and suddenness, and, also, tentativeness. These processes of reanalysis, needless to say, are construction-specific, *i.e.* they depend on the interaction between verb semantics and the other items, such as resultatives or aspect markers, if present. Thus, reduplication may add a temporal boundary, as in Mandarin, but may also act to the contrary; in Suzhou, for instance, both effects of reduplication are attested (compare ex. 31 and 32). However, more data taken from a broader variety of contexts is needed to provide a proper assessment of these phenomena.

4. Concluding remarks

In this paper we provided an illustration of the patterns of verbal and adjectival reduplication in Mandarin and in a convenience sample of twelve Chinese dialects, showing some interesting correlations between form, structure and meaning in reduplication which crosscut lexical classes. One of the most striking aspects of reduplication in Sinitic is that there appears to be a very significant difference between monosyllabic/monomorphemic and disyllabic/bimorphemic items; it seems that word structure constrains meaning somehow.

Monosyllabic/monomorphemic verbs and adjectives exhibit a wide range of behaviours in the languages considered; reduplicated monosyllabic verbs, in particular, may express meanings as different as delimitative aspect, tentativeness, rapid completion, suddenness, greater intensity, etc. We proposed that these functions are all somehow connected to two semantic (macro-)values, namely repetition over an unbounded time span, which appears to be the earlier use for verb reduplication in the history of Chinese, and repetition over a bounded time span, the innovative usage; these were extended to include the disparate functions and values which reduplicate construction possess in modern Sinitic varieties. As to disyllabic/bimorphemic reduplication, we showed that the association between functions and patterns is much more stable and consistent, both across word classes and across dialects. For instance, we did not found a single instance of a disyllabic (non-coordinate) verb reduplicating as ABAB and expressing increasing, rather than diminishing semantics, and in the dialects which allow ABAB reduplication of adjectives, as Taiwanese Southern Min, this has diminishing semantics.

As to the deeper significance of the distributional and selectional properties of reduplicative constructions, we outlined a syntactic analysis of the Mandarin data, though limited to the diminishing pattern. We argued that diminishing verbal reduplication is subject to aspectual constraints only, and appears to modify the eventive structure of the base verb (providing a temporal boundary to the event described). On the contrary, increasing verbal and adjectival reduplication is subject to *structural* morphological constraints on the input; we leave for future research whether increasing reduplication too, along the lines of the analysis put forward here for diminishing reduplication, may be accounted for in syntactic terms. Unfortunately, we could not provide a formal analysis for all the patterns exemplified due to lack of adequate data for varieties other than Standard Mandarin; we hope that further research, based on a large number of actual occurrences of reduplication in different contexts, rather than on individual examples, will make it possible to provide a unified analysis of verbal (and adjectival) reduplication in Sinitic based on the framework introduced here.

References

- Alexiadou A. 2010. "Reduplication and doubling contrasted: implications for the structure of the DP and the AP". *Linguística – Revista de Estudos Linguísticos de Universidade do Porto* 5: 9-25.
- Basciano, B., C. Melloni 2013. "Verb reduplication in Mandarin Chinese".Paper presented at the 39th Incontro di Grammatica Generativa (IGG), Modena-Reggio Emilia, 21-23 February.
- Borer H. 1994. "The projection of arguments" in E. Benedicto (ed), University of Massachusetts Occasional Papers in Linguistics 17. Amherst, MA, 19–47.
- 2005. In Name Only. Structuring Sense (Volume I). Oxford: Oxford University Press.
- Chao Y. 1968. *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.
- Chi C., C. Wang 2004. "温州话动词重叠式分析". Zhejiang Daxue Xuebao 34 (5): 149-157.
- Chuang H. 2007. *Verbal Reduplication in Taiwan Southern Min.* MA Dissertation. National Chung Cheng University, Minxiong.
- Ding Y. 2010. A Complete Guide to Teaching Chinese as a Second Language - 对外汉语教学参考. Beijing: Beijing Language and Culture University Press.
- Dixon R.M.W. 2004. "Word: a Typological Framework" in R.M.W. Dixon, A.Y. Aikhenvald (eds), Word: a Cross-linguistic Typology. Cambridge: Cambridge University Press, 1-41.
- Fu X., H. Hu 2012. "汉语方言单音节动词重叠式比较研究". Nanchang Daxue Xuebao 43 (5): 143-150.

- Hale K., S.J. Keyser. 1993. "On argument structure and the lexical expression of syntactic relations" in K. Hale, S.J. Keyser (eds), The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger. Cambridge, MA: MIT Press, 51–109.
- He W. 1989. 获嘉方言研究. Beijing: Shangwu Yinshuguan.
- Hong Z. 1999. "Cognate objects in Chinese". Toronto Working Papers in Linguistics 17: 263-284.
- Hu S., Z. Lin 2008. *铅山方言研究*. Beijing: Wenhua Yishu Chubanshe & Zhongguo Shehui Kexue Chubanshe.
- Hu X. 2006. "动词重叠AABB式的语法化". Hanyu Xuexi 8 (4): 18-25.
- Kurpaska M. 2010. *Chinese Language(s)*. A Look through the Prism of the *Great Dictionary of Modern Chinese Dialects*. Berlin-New York: Mouton de Gruyter.
- Lai H. 2006. "Iconic coding of conceptualization: Hakka reduplicative constructions". *Language and Linguistics* 7 (2): 483-499.
- Lapointe S. 1979. "A Lexical Analysis of the English Auxiliary Verb System". *GLOT* 2: 215-254.
- Lee S. 2012. "Syntax-based phonological asymmetries: the case of adjective reduplication in Mandarin Chinese". Paper presented at the 43rd annual meeting of the North East Linguistic Society, New York, 19-21 October.
- Li B. 2006. "闽东古田方言动词的重叠式". Fujian Jiaoyu Xueyuan Xuebao 7 (7): 67-72.
- 2013. "闽东古田话形容词的重叠式". Longyan Xueyuan Xuebao 31 (1): 44-51.
- Li C.N., S.A.Thompson 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Liu C.L. 2010. "The positive morpheme in Chinese and the adjectival structure". *Lingua* 120 (4): 1010-1056.

Ma Q. 1995. "漳州方言重叠式动词研究". Yuyan Yanjiu 1: 124-131.

- Macfarland T. 1992. "Event structure and argument structure of cognate objects". Paper presented at Console I, Utrecht, 16 December.
- Massam D. 1990. "Cognate objects as thematic objects". *The Canadian Journal of Linguistics* 35 (2): 161-190.
- Mattes V. 2007. Reduplication in Bikol. PhD dissertation. Graz University.
- Matthews S., V. Yip 2011. *Cantonese*. A *Comprehensive Grammar* (2nd edition). London: Routledge.
- McClure W.T. 1995. *Syntactic Projections of the Semantics of Aspect*. PhD dissertation. Cornell University, Ithaca (NY).
- Norman J. 2003. "Chinese dialects: phonology" *in* G. Thurgood, R.J. LaPolla (eds), *The Sino-Tibetan Languages*. London: Routledge, 72-83.
- Packard J. 2000. *The morphology of Chinese*. Cambridge: Cambridge University Press.
- Paul W. 2010. "Adjectival modification in Mandarin Chinese and related issues". *Linguistics* 43 (4): 759-793.
- Pham H. 1999. "Cognate objects in Vietnamese transitive verbs". *Toronto Working Papers in Linguistics* 17: 227-246.
- Ramchand G. 1997. Aspect and Predication: The Semantics of Argument Structure. Oxford: Clarendon Press.
- —. 2008. Verb meaning and the lexicon. Cambridge: Cambridge University Press.
- Shi Y. 2002. The Establishment of Modern Chinese Grammar. The Formation of the Resultative Construction and its Effects. Amsterdam-Philadelphia: John Benjamins.
- 2007. "汉语方言中动词重叠的语法意义和功能的差别". Hanyu Xuebao 4 (20): 59-63.

- Sybesma R. 1997. "Why Chinese verb-*le* is a resultative predicate". *Journal* of East Asian Linguistics 6 (3): 215-261.
- Tang
 T. 1988.
 "國語形容詞的重疊規律"
 in
 T. Tang
 (ed),

 漢語詞法句法論集.
 Taipei: Student Book, 29-57.
- 1997. "On reduplication of adjectives in Chinese: a Comparative study of Mandarin and Southern Min" in C. Cheng et al. (eds), Proceedings of the Eighth North American Conference on Chinese Linguistics. Los Angeles: GSIL, University of Southern California, 314-331.
- Travis L. 2000. "Event structure in syntax" in C. Tenny, J. Pustejovsky (eds), Events as Grammatical Objects: The Converging Perspectives of Lexical Semantics and Syntax. Stanford, CA: CSLI Publications, 145–185.
- 2010. Inner aspect: the articulation of VP. Dordrecht: Springer.
- Tsao F. 2004. "Semantics and Syntax of Verbal and Adjectival Reduplication in Mandarin and Taiwanese Southern Min" in H. Chappell (ed.), Sinitic grammar: Synchronic and diachronic perspectives. New York: Oxford University Press, 285–308.
- Wang F. 2011. A Study of Reduplication of Verbs and Adjectives in Wenzhou Dialect. MA Dissertation. National University of Singapore, Singapore & Beijing University, Beijing.
- Wang G., X. Xie (eds) 2009. 汉语重叠问题. Wuhan: Huazhong Shifan Daxue Chubanshe.
- Wang H. 2005. "吉林洮南方言的持续体范畴". Xueshu Jiaoliu 131 (2): 148-151.
- Wang P. 2011. 苏州方言研究. Beijing: Zhonghua Shuju.
- Wu Y. 2005. A Synchronic and Diachronic Study of the Grammar of the Chinese Xiang Dialects. Berlin-New York: Mouton de Gruyter.
- Xiao R., T. McEnery 2004. Aspect in Mandarin Chinese: A corpus-based study. Amsterdam-Philadelphia: John Benjamins.

- Xu D. (ed) 2012. *Plurality and Classifiers across Languages in China*. Berlin-New York: Mouton de Gruyter.
- Yang L. 2005. "成都话动词重叠格式的句法和语义特征". *Chengdu Daxue Xuebao* 2: 84-86.

Yang X. 2003. 汉语语素论. Nanjing: Nanjing Daxue Chubanshe.

Zeng Y. 2001. 湘潭方言语法研究. Changsha: Hunan Daxue Chubanshe.

Zhu X. 2003. A Grammar of Shanghai Wu. Munich: Lincom Europa.

THE ORGANIZATION OF CHINANTEC TONE PARADIGMS

Surrey Morphology Group, University of Surrey

Enrique L. Palancar SeDyL-CELIA (UMR8202), CNRS Surrey Morphology Group, University of Surrey

Abstract

The tonal paradigm of verbs in Tlatepuzco Chinantec counts as one of the most complex and opaque ever described, with five tone values distributed over twelve cells (distinguishing person/number and aspect) to yield c. 70 distinct paradigm types, with no consistent mapping between morphological form and morphosyntactic function. We suggest that useful generalizations will emerge only when we consider units of analysis larger than the individual inflected form, which we dub *inflectional series*. For Tlatepuzco Chinantee this means concatenating the three aspectual forms for each person/number value. The resulting units allow us to see structural relationships between the elements of the paradigm which were previously inaccessible.

1 Introduction¹

Chinantec (or Chinantecan) is a language family spoken in Mexico which is a branch of Oto-Manguean, which in turn is one of the most diverse linguistic phyla in the Americas. Chinantec verb paradigms have served as the poster child for complex inflectional systems (e.g. Finkel & Stump 2009), mainly because there is no clear form-function mapping and little predictability of inflection classes. In this paper, we attempt a step in the right direction, looking at the person marking of the verbs of Tlatepuzco Chinantec

¹ This paper was written under the auspices of the ESRC/AHRC grant RES-062-23-3126 "Endangered Complexity: Inflectional classes in Oto-Manguean languages" and the ERC grant ERC-2008-AdG-230268 MORPHOLOGY. We thank Olivier Bonami for his invaluable feedback and an anonymous reviewer for his/her very insightful comments.

(henceforth TlCh), a Central Chinantec language. Our analysis is based on the study of the data and the grammatical description in the large dictionary by Merrifield & Anderson (2007).

1.1 Some preliminaries about Tlatepuzco Chinantec.

Like the other Chinantec languages, TlCh uses tone as an exponent of inflection. Tonal inflection consists of three level tones (1, 2, 3 = high, mid, low) and two contour tones (31, 32 = high rising, mid rising). Verbs mark subject person/number and aspect. While the pronominal system distinguishes six person/number values, verbs distinguish just four, namely first person singular, first person plural, second person and third person. This can be seen in (1).

(1)	'I bent it'?	ka ³ -hú ? ³	hní ²
	'we (INCL) bent it'	ka ³ -hú ? ³¹	hniaŋ ¹
	'we (EXCL) bent it'	$ka^3-hú?^{31}$	hnie? ¹
	'you (SG) bent it'	ka ³ -hú ? ³	?niŋ²
	'you (PL) bent it'	ka ³ -hú ? ³	?nia? ³²
	'she, he, they bent it'	ka ³ -hú ? ²	dsa ²

(

There are three aspectual stems (PRS, FUT, CPL),² from which additional TA values can be derived through simple prefixation, as shown in (1). For example, the PRS stem on its own expresses incompletive and habitual aspectual values, while the addition of the prefixes mi^{1} - and ma^{2} - yield the so-called "Imperfect tense" and the "Perfect tense". For the purposes of this paper, it is sufficient to focus just on the bare stems in (3).

1)	TA	1sg	1 pl	2	3	Stems
	Present	hú? ³²	hú? ³²	hú? ³²	hú? ²	PRS
	Imperfect	mi ¹ -hú? ³²	mi ¹ -hú? ³²	mi¹ -hú?³²	mi¹- hú?²	
	Perfect	ma ² -hú? ³²	ma ² -hú? ³²	ma ² -hú? ³²	ma ² -hú? ²	
	Future	hú? ³¹	hú? ³¹	hú? ³¹	hú?²	FUT
	Potential	mi¹ -hú?³¹	mi ¹ - hú?³¹	mi¹ -hú?³¹	mi¹- hú?²	
	Past	ka ³ - hú?³	ka ³ -hú? ³¹	ka³ -hú?³	ka ³ -hú? ²	CPL
	Hodiernal	na² - hú?³	na² -hú?³¹	na²-hú?³	na ² - hú?²	

 $^{^2}$ Following the practical system employed by Merrifield and Anderson (2007) we employ tense labels 'present' and 'future', for the first two stems, when in reality they could be alternatively characterized as 'incompletive' and 'irrealis'.

(3)		1SG	1pl	2	3
	PRS	hú? ³²	hú? ³²	hú? ³²	hú? ²
	FUT	hú? ³¹	hú? ³¹	hú? ³¹	hú? ²
	CPL	hú? ³	hú? ³¹	hú? ³	hú? ²

1.2 The nature of the problem.

CPL

In a maximally transparent inflectional system, we would expect a one-to-one mapping between form and function. That TlCh is not like this is already patently obvious from the repertoire of inflectional formatives: given that there are five tonal formatives distributed across twelve functions (four person/number values x three aspects), a one-to-many mapping (syncretism) is inevitable. And in fact, it works in the other direction as well, with the multiple tonal formatives mapping onto the same function (allomorphy), depending on the lexeme. For a concise illustration, compare the tonal paradigms of $2nai^{32}$ 'sell something' and $kieg^2$ 'lie down' in (4a) and (4b): both employ the same set of tones, but in an almost completely different distribution.

(4) a. [*Inai*³² 'sell something', Class 13b] b. [*kieg*² 'lie down', Class 51f] 2 1pl 2 3 1sg 1pl 3 1sg 2 32 32 2 32 2 2 2 PRS PRS 3 31 31 3 31 FUT 1 1 FUT 1 3 1 1 3 CPL 2 31 3 2

While such many-to-many mappings are a familiar feature of inflectional systems, TlCh - and indeed all Chinantec languages -takes this to an extreme. An impression of the full extent can be gleaned from Table 1, which charts the tone formatives found in the inflected forms of 613 verbs of TlCh for each person/number value and each TA stem.³ There is an almost complete lack of form-function mapping, e.g. all tone units are used for all possible values.

³ Verbs from Merrifield & Anderson's (2007) dictionary that have complete paradigms, *i.e.* omitting intransitive inanimates that lack first and second person subject forms, as well as a few items with incomplete or contradictory information.

		1SG			1pl			2			3	
Tone	PRS	FUT	CPL									
1	149	198	149	146	265	264	151	289	225	10	7	10
2	96	21	92	172	20	43	184	20	85	242	200	240
3	11	67	287	11	68	56	11	67	176	31	335	300
31	70	315	71	74	248	237	70	226	86	12	22	12
32	287	12	14	210	12	13	197	11	41	318	49	51
Total	613	613	613	613	613	613	613	613	613	613	613	613

Table 1: Tone formatives in the inflected forms of 613 verbs of TICh

At first glance one might believe this is an almost random distribution of forms, but the fact that they fall into 'only' around 70 distinct paradigm types suggests that there is some kind of order to them.⁴ But what? As a first step, consider the entropy measures in Table 2, generated using Raphael Finkel and Gregory Stump's *Analyzing Principal Parts* software.⁵ The table illustrates the predictability of the forms shown in columns, given the forms shown in the rows (measured in bits x 100), factoring in type frequency. The higher the number, the less predictable the form; thus the correct form of the 3rd person present for a given verb can relatively easily be predicted from the 3rd person completive (value of 19), while the correct 2nd person completive is very difficult to predict if one knows just the 1PL future.

	1sgP	1sgF	1sgC	1plP	1plF	1plC	2P	2F	2C	3P	3F	3C
1sgP		72	70	121	139	140	40	98	95	53	103	65
1sgF	106		34	136	127	142	133	38	131	135	70	134
1sgC	95	25		128	140	131	123	63	110	123	94	113
1plP	177	158	158		75	64	177	156	190	160	153	170
1plF	186	140	161	66		54	188	139	194	178	139	173
1plC	186	155	152	54	54		186	153	185	174	152	165
2P	40	- 99	97	121	142	140		66	90	43	98	56
2F	134	40	73	135	128	142	101		126	131	58	129
2C	87	89	76	126	139	130	81	83		90	119	84
3P	65	113	109	116	144	139	55	107	110		63	21
3F	139	72	105	133	128	142	134	59	163	87		84
3C	75	110	98	125	137	129	66	104	103	19	58	

Table 2: Conditional entropy of the tone formatives

What is striking about these figures is that, overall, the entropy measures divide the paradigm into cross-cutting distinctions of person and aspect, as broken down in Table 3.

⁴ As pointed out by an anonymous referee, five allomorphs over twelve cells yields

^{248,832} possible paradigm types, so we would expect closer to 613 types.

⁵ http://www.cs.uky.edu/~raphael/linguistics/analyze.html

3 rd person:	55
1pl:	61
1sg:	67
2 nd person:	91
Present:	94
Future:	108
Completive:	123
Differing in both person and aspect:	129

Table 3: Average entropy within restricted domains of the paradigm

That is, within a given value for person, it is easier to predict the different aspectual forms for a lexeme than it is to predict the different person forms within a given aspect, and both are easier than predicting between forms that differ in both person and aspect. In this way person and, to a lesser extent, aspect, are organizing parameters of the paradigm, in that they define domains of relative interpredictability. This suggests that, in spite of the apparent amorphousness of the tonal paradigm, it is in some sense built on conventional featural distinctions, with person playing a dominant role. We propose here to pursue this idea further by providing an account of the TICh tonal paradigm as a person-marking system, in which the properties of and relationships between morphological exponents can be described in relatively simple terms. This allows us to demystify, at least somewhat, the otherwise puzzling complexity of the system.

The crucial point to be gleaned from comparing the figures in Tables 1 and 3 is that the discernible role of person in the organization of the paradigm only begins to emerge when we look at its exponence across multiple aspect values. We propose therefore to treat as our basic unit of form not the individual exponents of person and aspect, but rather the combination of all aspect values for a given person. But before we attempt to characterize the paradigm in such terms, we need to dispense with other possible explanations for the tone patterns and the resulting wealth of inflection classes.

2 Rejected explanations

We need to discard first a set of possible alternative explanations for the variation we observe.

2.1 Semantic explanations

Verbs fall into a few different syntactic/semantic classes based on transitivity and animacy. These are relevant for valence increasing mechanisms, for the syntax of agreement and for the syntax of voice (inverse, passive), but have no observable relationship to the tonal formatives. For example, the verbal lexicon is largely split according to the animacy of the ABS argument (object of transitive or subject of intransitive). The members of animate/inanimate pairs differ along various morphological parameters, including tone, but there are no dedicated animate or inanimate tone patterns. Consider for this purpose the tone of the CPL stem of three gender pairs in Table 4.

CPL		inani	mate		animate					
	1sg	1pl	2	3	1sg	1pl	2	3		
'buy'	lá ²	lá ³¹	lá ³	lá ³	lán ²	lan ³¹	lán ³	lan ²		
'grab'	gúu? ²	guu ? ³¹	guu? ¹	guu? ³	guén? ²	guen? ³¹	guén? ³	guen? ³		
'hit'	bá ²	bá ³¹	bá ³	bá ²	bán ²	bán ³¹	bán ³	bán ²		

Table 4: Completive tonal paradigm of three animate/inanimate pairs

The inanimate verb 'buy (something inanimate)' has the same tonal inflection as the animate verb 'grab (somebody)', but its animate counterpart 'buy (something animate)' has the same tone pattern as the pair of verbs meaning 'hit', which do not mark animacy through tone. This lack of correlation between tone patterns (or indeed any individual tone value) and animacy is typical of the verbal lexicon as a whole.⁶

2.2 Resegmentation

Besides tone, there is an additional prosodic feature, involving the contrast between the so-called 'ballistic stress' ("characterized by an initial surge and rapid decay of intensity, with a resultant fortis articulation of the consonantal onset" (Foris 2000: 16) and 'controlled stress' (characterized by a longer duration and medium intensity) (see Foris 2000 or Pace 1990; for an alternative phonological oriented account as a glottis phonation feature, see Silverman 1994). If we combine this with tone, it doubles the number of morphological exponents, but it does not bring any noticeable clarity.

2.3 Stem phonology

Tonal alternations do not obviously correlate with any other phonological properties of stems, e.g. nasalization, nucleus quality, etc.

2.4 Morphological correlations

Around a fifth of the verbs (137/613) display stem alternations. While these show some interesting interactions with prosodic alternations there is again no clear predictability in either direction (cf. Baerman, forthcoming).

⁶ The relation between animacy and inflection class is a conceivable one which is for example observed in the distribution of the subject suffixes of another Chinantec language such as Lealao Chinantec (for more details, see Palancar, submitted).

3 The proposal: redefinition of the morphological unit

Following Palancar (forthcoming), and the reasoning outlined in \$1.2, we suggest dividing the paradigm up not into twelve cells marking person and aspect, as in (5a), but into four cells marking person, as in (5b). Palancar calls the resulting units 'inflectional series'.



This is the concept of morphological unit we will adopt in this paper, and it will have the shape [# # #], consisting of *present-future-completive tone*. In the following sections, we will show that this new segmentation of the cell reveals a certain degree of regularity both in inflection class assignment and in form-function mapping.

This notion of inflectional series should be distinguished from other types of subparadigmatic unit. For example, Finkel and Stump (2007) employ the concept of *segregated inflection classes*. As an example they take the Sanskrit verb paradigm, which can be divided into the present system and the aorist system. Each of these constitutes a self-contained system, with its own stem and its own set of inflection classes. By contrast, the mapping between the two systems is arbitrary and unsystematic. The Chinantec inflectional series is in effect the mirror image of this. That is, while the individual inflectional series do not themselves constitute morphologically coherent units, once this grouping is made, there is a certain degree of predictability between the resulting units. By analogy, we can think of the inflectional series as phonemes, and the individual tone values as constitutive, distinctive features.

3.1 Inflection class assignment

Applying the concept of an inflectional series to inflectional allomorphy makes it possible to obtain two interesting generalizations. We treat inflectional series as principal parts, employing a maximum of two: one for the 3rd and another for the 1sG (the 1PL and the 2nd would serve equally well, with very minor changes in the numbers):

- As shown in (6), if the 1SG form is the same as the 3^{rd} person, then the forms for the 1PL and 2^{nd} person are the same too. This accounts for 123 lexemes, with only one exception.



- As shown in (7), if the 1SG form consists of three identical tones, the 1PL and 2^{nd} person are identical to it. This accounts for 198 lexemes, with five only exceptions.



Overall these two generalizations cover 321/613 = 52% of the lexicon, mapping onto what Merrifield & Anderson (2007) classify as types B & C (there are five exceptional lexemes). Everything else belongs to their class A. This is shown in (8).



The lines in (8) represent the maximal permitted distinctions, but often there will be conflation within that, e.g. the following represents what would count as a class B paradigm (treated as B1f in Merrifield & Anderson, 2007):

⁷ Future/Completive syncretism in the 1PL is almost exceptionless.

⁸ Merrifield & Anderson (2007) characterize class B simply in terms of person syncretism, but since 97% (198/205) of the lexemes also display aspectual syncretism, we treat both as characteristics of this (revised) class.



The remainder of the lexicon (class A) still needs another account. One option is to construe one of the inflectional series as the default exponent for each person value. We have chosen the most frequent ones, and they account for 55% (129/285) of the remaining data-points.

(10)	1sg	1pl	2	3
PRS	32	32	32	32
FUT	31	31	31	3
CPL	3	31	3	3
	233	156	105	135
	(82%)	(54%)	(37%)	(48%)

Note that the figures in (9) represent the frequency for each form in isolation, not a default paradigm as such; only 16 lexemes (class A11a) actually show all four at once.

3.2 Form-function mapping

3.2.1 The marking of person

The individual inflectional series can be seen as exponents of person, at least to a certain extent. This is shown in Table 5 below (in the table, everything with a token count under 6 (1%) is not highlighted). If we set aside the forms which do not distinguish person at all (C paradigms), the following two generalizations hold:

- (i) The 3rd person form is distinct from 1st (SG/PL) and 2nd person.
 (ii) While 1SG is always distinct from 1PL, 2nd person is usually syncretic with 1SG or 1PL.9

Considering now the C paradigms, we can make a third observation:

(iii) C paradigm forms typically coincide with the distinct 3rd person forms found in the A and B paradigms (88/117).

⁹ 2nd person is [2 1 1] or [32 31 31] only if 1PL is too (65/67 and 15/15); 2nd person is [32 31 3] only if 1SG is too (102/105).

The broad generalization that emerges is that the system potentially distinguishes 1SG, 1PL, 2 and 3, but that (i) 2^{nd} person often takes the form of either 1SG or 1PL, (ii) all 1^{st} and 2^{nd} person values may conflated in a dedicated non- 3^{rd} person form (B paradigms), and (iii) forms that do not distinguish person at all often coincide with 3^{rd} person form.

3.2.2 The marking of aspect

The inflectional series are made up of tonal alternations for aspect. By looking now at the system from the perspective of aspectual marking, we gain some insight into how the series are constituted. Consider again the figures from Table 1 above. If we now divide them by paradigm type (A, B, C), we see there are some striking regularities in the distribution of the tonal components.

		А	В			С		
Series	1sg	2	1pl	3	1/2	3		1/2/3
[1 1 1]	4	6	1	1	137			6
[1 3 1]								3
[2 1 1]		67 (118					
[2 1 2]	52	46	4					
[2 1 3]		20						
[2 1 32]		10						
[2 2 2]	1			172		3		14
[2 2 3]				1				4
[2 3 2]				16		1		17
[2 3 3]				1		2		7
[2 31 3]								1
[2 31 31]			7					
[3 3 3]					1	21		10
[31 31 31]	1		4	1	60	1		10
[32 2 3]				1				2
[32 3 3]			1			76		
[32 3 32]								2
[32 31 1]		7						
[32 31 3]	233	→105	→ 2					9
[32 31 31]	1 ←	15 ←	155					
[32 3 2]				16				1
[32 3 3]				211				26
[32 31 32]		16						
[32 32 32]				15		24		11

Table 5: Inflectional series as exponents of person

Take first the A paradigm verbs illustrated in Table 6 (in the tables, figures under 12 lexemes (2%) are not highlighted).

The patterns emerge most clearly if we group the individual tonal exponents into three tone types:

- HIGH, made up of high tone (1) and ascending to high (31)

- MID, made up of mid tone (2) and ascending to mid (32)

- MID/LOW, made up of mid (2) and low (3)

	PRS				FUT				CPL			
Tone	1sg	1pl	2	3	1sg	1pl	2	3	1sg	1pl	2	3
1	3	0	5	1	53	120	144	1	3	118	79	1
2	51	127	139	120	1	0	0	104	51	2	44	131
3	0	0	0	0	0	1	0	168	234	3	123	141
31	0	4	0	1	235	168	146	1	1	167	16	1
32	236	159	146	163	1	1	0	11	1	0	28	11

Table 6: Tones and aspect in A paradigm verbs

The tone/aspect mappings that result are the following:

- (i) The PRS stem is MID tone.
- (ii) The FUT of non-3rd person is HIGH tone and MID/LOW for 3rd, keeping the marking of 1/2 vs. 3rd distinct.
- (iii) The marking of the CPL stem is complex: 1PL is syncretic with FUT; 2nd person can be realized by almost anything; and 1sG and 3rd person have MID/LOW tone. In the 3rd person, this means that the same tone profile as in the FUT, but the forms are not necessarily syncretic.

All this can give us a perspective on the syncretism of person values seen above in Table 5. For example, the distinct 3^{rd} person FUT tone profile ensures that the 3^{rd} person form will never be syncretic with the 1^{st} and 2^{nd} person forms. Other person values have distinct tone profiles only in the CPL. The fact that 2^{nd} person CPL can have any tone means that sometimes it will be distinct from the others, sometimes coincide with them, accounting for the variable syncretic behaviour of the 2^{nd} person seen in Table 5. Now let us see what happens in B and C paradigms.

		В					С				
	1/2		3			(TA	(TA inflecting)			(uninflecting)	
Tone		PRS	FUT	CPL		PRS	FUT	CPL			
1	137	0	0	0		3	0	3		6	
2	0	76	73	74		29	6	18		14	
3	1	21	100	99		0	56	49		10	
31	60	1	1	1		0	10	0		10	
32	0	100	24	24		40	0	2		11	
	/			1		• n	10	1			

 Table 7: Tones and aspect in B and C paradigm verbs

The aspectless 1/2 form of the B paradigms has the same tone profile (MID) as the 1/2 FUT of the A paradigms. Whereas C paradigm verbs that inflect for TA behave more or less like 3rd person forms, the tone of uninflecting verbs does not show any particular tendency, suggesting that tone assignment here is given in the lexical entry.

4 Conclusion

In this paper, we have argued that the notoriously opaque tonal inflection of TlCh can be at least partially tamed by a judicious choice of analytic units. First, we showed that person can be taken as the dominant organizing principle, in that person values define the most reliable domains of interpredictability within the paradigm. From this observation we extracted a morphological unit that we have labelled inflectional series, made up of the three aspectual forms for each person value. By thus construing the tonal paradigm as a four-cell paradigm marking person, we were able to describe the distribution of forms in terms of a few patterns of syncretism, and concisely characterize the inflection classes in terms of implicational relationships between these forms. We then examined the nature of the inflectional series more closely by looking at aspectual marking in isolation, consolidating tonal marking into three major types. This showed that, given the person/number value and aspect, tonal inflection is highly constrained, the complexity of the system emerging at certain discrete points (e.g. the completive of A class verbs).

For obvious reasons, our analysis in terms of inflectional series cross-cuts an analysis in terms of optimal principal parts. Because of the relative interpredictability of individual tones within each inflectional series, optimal principal parts will tend to be spread across the series. For example, as Olivier Bonami (personal communication) has shown, if we limit ourselves to the four cells that are the best predictors of the entire inflectional paradigm, they will be drawn from each of the four inflectional series, as shown in (11). (These reduce the average conditional entropy of the remaining cells to 0.138 bits, or 138 in terms of the scale used in Table 2.)¹⁰ Significantly, these optimal principle parts are drawn from each of the three aspects as well; recall from Table 3 that aspect also defines a zone of interpredictability (though less reliable than person), so this is what we would expect. Thus the selection of principle parts is complementary to the parameters of paradigmatic organization that we have claimed operate in TlCh, namely person and aspect.



References

- Baerman, M. Forthcoming. "Inflectional class interactions in Oto-Manguean", in J-L. Léonard and A. Kihm (eds.), Issues in Meso-American morphology. Paris: Michel Houdiard.
- Finkel, R and G Stump. 2009. "Principal parts and degrees of paradigmatic transparency", in J. P. Blevins and J. Blevins (eds), Analogy in Grammar: Form and Acquisition. Oxford: Oxford University Press, 13-53.
- Finkel, R and G. Stump. 2007. "Principal parts and morphological typology". *Morphology* 17: 39-75.
- Foris, D. P. 2000. A grammar of Sochiapam Chinantec (Studies in Chinantec Languages 6). Dallas, TX: SIL International and the University of Texas at Arlington.

¹⁰ Though note that this measure does not factor in type frequency; if it did, the figure would be lower.

- Merrifield, W. R. and A. E. Anderson. 2007. Diccionario Chinanteco de la diáspora del pueblo antiguo de San Pedro Tlatepuzco, Oaxaca. [2nd Edition] (Serie de vocabularios y diccionarios indígenas "Mariano Silva y Aceves" 39). Mexico DF: Summer Linguistic Institute.
- Pace, W. J. 1990. "Comaltepec Chinantee verb inflection", in W. R. Merrifield and C. R. Rensch (eds.), Syllables, tones and verb paradigms (Studies in Chinantee Languages 4) Arlington, Texas: The Summer Linguistic Institute and the University of Texas at Arlington, 21–62.
- Palancar, E. L. Forthcoming "Revisiting the complexity of the Chinantec verb conjugation classes", in J-L. Léonard and A. Kihm (eds.), *Issues in Meso-American morphology*. Paris: Michel Houdiard.
- Palancar, E. L. Submitted "When two agreement systems compete: The suffix classes of Lealao Chinantec" MS submitted to *Morphology*.
- Silverman, D. 1994. "A Case Study in Acoustic Transparency: [spread glottis] and Tone in Chinantec". *Proceedings of NELS* (Amherst: University of Massachusetts) 24, 559-572.

THE NOMINAL MORPHOLOGY OF LOVARI FROM AN ANALOGICAL PERSPECTIVE

Márton András Baló

Research Institute for Linguistics, Hungarian Academy of Sciences

Abstract

This paper attempts to provide evidence that analogy-based approaches make language change, as well as unstable and variegated forms and word classes easier to understand and grasp than they would be in a traditional synchronic framework or through a rule-based diachronic analysis. The example presented is the declension of the Lovari dialect of Romani, a dialectally most diverse Indo-European language that is often exposed to contact-related influences. Although analogical effects are mentioned in the relevant literature, the overall approach tends to rely on traditional rules, and the fact that analogy can actually be regarded as the cementing force of a language is often not taken into account. The available information regarding the inflectional nominal paradigms of the Lovari dialect of Romani spoken in Hungary is contradictory, especially concerning the oblique stem. However, the seemingly high number of nominal paradigms can be reduced to just two if we reconsider the defining criteria. A unique feature of the Romani language, the strict split between the morphology of inherited and borrowed vocabulary is also seen in a new light if we examine the possible processes behind the apparent erosion of this system.

1 Introduction

The Romani language, due to the circumstances in which it is used, frequently borrows lexical items. Borrowing mostly happens from the local language, that is, the one spoken by the majority society surrounding the given Roma community and has been particularly intensive since the appearance of the Roma in Byzantium and their dispersal in Europe. Romani is originally divided into dialects on a geographical basis (*cf. e. g.* Matras 2005; Bakker & Matras 1997; Miklosich 1872-80).



Figure 1: The estimated areas where Lovari and the Central dialects are spoken (*cf.* the project "The Linguistic Atlas of Central Romani" and Matras 2005)

The dialects established in that manner are split into further varieties due to further migration; thus, for instance, whereas Lovari was originally spoken in western Romania, it is possible to talk about Hungarian and Austrian Lovari, which coexist and interfere with the Romungro and the Burgenland Romani varieties, respectively (both belong to the Central dialect group as opposed to Lovari, which is a member of the Vlax dialects). There can be so many differences between two, fairly distinct dialects that the speakers might switch to a language that they both speak alongside Romani (Boretzky 1995). In the Romani lexicon and declension, there is a very clear-cut distinction between inherited and borrowed vocabulary in terms of morphological patterns and paradigms. No morphophonological reason exists which would justify the difference between the oblique stem of \bar{aro} 'flour', which is \bar{ares} -and the oblique stem of foro 'town', which is foros-.¹ There is, however, an ostensible blurring or mixing of the two different inflectional patterns, as noted most prominently by Elšík (2000), and the anomalies cannot possibly

¹ This unique feature is called "thematicity" in the terminology of Romani linguistics. The term itself is borrowed from Indo-European linguistics but with an unrelated meaning: in Romani linguistics, it does not refer to the presence or absence of a thematic vowel or consonant; it simply refers to the difference between the ways inherited and borrowed lexical items are treated morphologically. The former is referred to as "thematic", while the latter is called "athematic".

be explained with a traditional rule-based approach. There are several references to analogy in Matras (2002) and in Elšík (2000), but only as a secondary phenomenon. However, if we examine the blurring from a primarily analogical perspective, we can see how language changes in real time and how categories based on earlier historical developments become obscure or fuzzy.

The erosion of the distinction between the inherited and the borrowed parts of the lexicon is even more interesting if we take into consideration that this has been one of the few but very solid features that all dialects have shared. We will have a closer look at these processes through the example of the Lovari dialect as spoken in Hungary, while giving examples from other dialects as well.

2 Analogy

The term analogy will be used in the broad, Saussurean sense throughout the paper: "an analogical form is a form made on the model of one or more other forms" (Saussure 1966: 161). His classic example is the eventual spread of the rhotacised oblique onto the nominative in Latin:

$\bar{o}r\bar{a}t\bar{o}rem : \bar{o}r\bar{a}tor = hon\bar{o}rem : x$ x = honorFigure 2: The Saussurean analogical proportion (Saussure 1966: 161)

Here, the former nominative singular *hono*s is replaced by a more regular form, *honor*. This pattern, also called four-part or proportional analogy (*cf. e. g.* Kraska-Szlenk 2007) is the very pattern we encounter in Romani declension, where forms such as *foros*- are replaced by the more regular and, strangely enough, also more conservative *fores*-.

Rung (2011) gives a very detailed overview of analogy-based approaches. He notes that structural linguists (Sapir 1921 and Bloomfield 1933 in particular) maintained that analogy had a great significance: new utterances are created based on an analogy with previously uttered or heard words and sentences. In other words, patterns and exemplars, already existing in our minds, serve as bases for new forms or old ones undergoing some sort of change. The governing forces of language may thus be seen as surface patterns, or, in other words, "closely related surface forms" (Bybee 1985: 49-50): constructions or abstract schemas, which are complex instances of form and meaning (Goldberg 1995; Booij 2010), similarly to the notion of sign taken in the original, Saussurean sense: "the linguistic sign unites ... a concept and a sound-image" (Saussure 1966: 66). While constructional approaches maintain

the creative aspect of language, they "generally recognise that grammars don't generate sentences, speakers do" (Goldberg 2006: 22). The same is true for word formation, where "patterns can be seen as abstract schemas that generalize over sets of existing complex words" (Booij 2007: 34). But patterns don't only exist in syntax and word formation; they are also present in inflectional paradigms, as can be seen in Romani.

Paradigms, that is, a set of forms belonging to the same lexeme (*cf.* Wurzel 1989: 52), form an important part of analogical theories as the similarity of combinations of form and function is a significant characteristic of paradigms, and analogy maintains paradigmatic uniformity (Albright 2009; Eddington 2006). Similarity in grammatical function involves similarity in form, or, in other words, "it is natural for related concepts to be designated by related sounds" (Humboldt 1999: 71). Similarity can be measured in terms of the surface forms (the actual identity of phoneme sequences) or based on the extent to which the defined features of words are alike (Rung 2011).

Analogy is closely related to the concept of patterns and (ir)regularity in that it "supposes a model and its regular imitation" (Saussure 1966: 161) and these models, or patterns, can also be seen as constructions. Analogy rests on statistical evidence; analogical force depends on the frequency of the pattern in question. A pattern with higher type or token frequency is more powerful, and competing patterns result in instability. On the other hand, less frequent forms are more prone to undergo analogical change. The variation caused by competing patterns is further enhanced by the diversity of dialects, the lack of a written standard and frequent borrowing in Romani.

If we dismiss the dichotomy of underlying and surface representations and do not try to force one and single origin on the different surface forms, considering the surface forms themselves instead, the variation within a paradigmatic schema will be much less problematic and exceptions can easily be integrated in the system. In an analogy-based framework, we can disregard another dichotomy, namely the opposition between the diachronic and the synchronic approach to language, as language change, which is usually part of diachronic descriptions, is palpable in synchronic terms through the unstable and variegated forms. We can also incorporate the notion of rules, although not taken in the generative sense, but following van Marle (1990), who bears upon analogy as a synchronic force and argues that "the speakers of a language have the capacity to construct rules on the basis of the existing words" (van Marle 1990: 267), called "rule-creating creativity".

A practical model for measuring analogical effects is Analogical Modelling of Language (*cf.* Skousen 1989; Skousen 1992; Skousen *et al.* 2002), where patterns are represented by a dataset of exemplars, that is, a set of words whose similarity to the form in question is strictly predefined. The exemplars are arranged into supracontexts based on the distribution of previously chosen

variables. Homogeneous supracontexts form the analogical set which is used to predict the outcome.

3 Inflection in Romani

Matras (2002) presents an analytical model of Romani nouns, in which the surface form of an inflected noun is assumed to consist of layers, similarly to Indo-Aryan as described by Masica (1991) or the blocks of realisation rules in Paradigm Function Morphology (Stump 2001). We will see, however, that these layers are simply unnecessary for Romani.

Three layers are posited in this framework, namely Layer I, composed of nominal and oblique endings,² Layer II, which comprises the actual case endings, and a set of adpositions named Layer III. Layer I endings are attached directly to the nominal base. The declension class of a noun can traditionally be seen from the Layer I oblique ending, which is in turn determined by several factors to be discussed below, but as we will see, many questions arise related to these factors as well as the different declension classes. Layer II endings are case suffixes attached to the Layer I form of a noun. All in all, Romani distinguishes eight different cases: nominative, accusative, dative, genitive, ablative, locative, instrumental and vocative. Most of the case suffixes are fixed in form (although they are subject to some variation among the dialects), showing only voice assimilation, and are added to the oblique stem.

category	Form	function / meaning				
lemma	ānró	'egg'				
nominal base	ānr-	stem				
Layer I	- <i>es</i> -	oblique marker				
Layer II	-te	Locative				
Layer III	ande	Locative?				

Table 1: The general layout of a Romani noun

² For the sake of simplicity, I will adopt the terms "nominal" and "oblique" used in Romani linguistics in the paper, although they might as well just be dubbed short and long stems.
As we can see from the example, the terms used are ambiguous. The use of the term "nominal" is redundant if we say that there is no other stem. That is what we apparently see, as both the nominal and the oblique (and the vocative, for that matter) endings attach to this. Strictly speaking, the oblique "stem" is not a stem, but it is derived from the nominal base.

 $bakr + -o \rightarrow bakró$ 'sheep' Nom. $bakr - + -es \rightarrow bakrés$ - 'sheep' Obl. $bakr - + -a \rightarrow bakrá$ 'sheep' Voc. Figure 3: An overview of the relations among stems in Romani

It would therefore be sufficient to posit one single stem which serves as the basis for all other forms of the given noun. Elšík (2000), on the other hand, proposes to differentiate between BSA (base-stem affixation) and OSA (oblique-stem affixation) languages. The former refers to languages where the cases are marked with individual suffixes; the latter means that the case suffixes are attached to an oblique stem. Romani belongs to the OSA languages, as opposed to Hungarian, for example, where case suffixes directly follow the stem, i.e. the nominative form, without mediation (bárány 'sheep' Nom. $\rightarrow b \dot{a} r \dot{a} n y b a n$ 'sheep' Loc.). If we treat the nominative and the vocative independently, this could indeed be a possible analysis. Blake (2000; 2001) also make reference to an oblique stem "which serves to set the nominative off from the other cases" (Blake 2001: 42). A similar example is Lezgian (Blake 2000 based on Mel'čuk 1986), where the bare oblique stem functions as the ergative case. Elšík (2000) also mentions Daghestanian languages, where the ergative case is unmarked, similarly to Romani, where the unmarked case is the accusative, which is derived from the oblique stem by an identity process. As for the vocative, Matras (2002) notes that the vocative forms can be found "alongside" the three layers and "connect directly to the nominal base" (Matras 2002: 80). But even then, the nominative-oblique dichotomy remains.

From a synchronic perspective, it is more economical to say that there are two bases: the nominal base, marked by a zero morpheme: $bakr + \emptyset$, and an oblique base, marked by a suffix: bakr + es and sokr + os. These could then in turn serve as bases for the nominative and the vocative on the one hand, and the rest of the cases on the other. Thus, we could treat all the cases in the same way, even if historically the vocative is of different origin, and we would not have to deal with all the diverse nominative endings as Layer I elements.

base	case	Form
nominal	nominative	-Ø, -o, -i, -a
nominal	vocative	- <i>a</i>
oblique	accusative	-Ø
oblique	dative	-ke
oblique	locative	-te
oblique	ablative	-tar
oblique	instrumental	-sa
oblique	genitive	- <i>k</i> -

Table 2: singular case markers in Lovari

The exact status of the Romani genitive (it behaves like an adjective in many aspects) is not relevant here but it is discussed in detail, among others, by Elšík (1997), Grumet (1985) and Koptjevskaja-Tamm (2000).

The term Layer III is also misleading, and its use may only be justified by historical reasons. Layer III in Romani derives from Indo-Aryan Layer III indeed, but while the latter is "*potentially* mediated" (Masica 1991: 234) by a Layer II element, the former has a "preposed position" (Matras 2002: 80), and so it has no contact with the other layers. In actual fact, it consists of prepositions, which are used completely independently of Layers I and II and by now they are always followed by the nominative: *ando kher* 'in the house'.³ For the purposes of our analysis, they do not need to be considered as an inflectional layer.

³ With regard to Austrian Lovari, we find the form *ande bute beršende* 'in many years' in Cech & Heinschink 1999a, which testifies the diachronic development of the layers and the existence of variation among different varieties of the same dialect. They also note the existence of the form *ande bute beršen*; here, the disappearance of the locative case marker points towards the erosion of redundancy finally achieved in Hungarian Lovari, where the use of the nominative case after Layer III adpositions eliminates the "double" locative.

preposition	definite article	noun Nom.				
ande 'in'	0	kher 'house'				
Table 3: the current state of Laver III in Lovari						

 Table 3: the current state of Layer III in Lovari

We have now arrived at the conclusion that – if we disregard the concept of layers derived from the Indo-Aryan heritage for a moment – it looks economical to say that there are two base forms and a set of suffixes, some of which attach to the nominal base, while others attach to the oblique base. This is in line with the notion of stem space as described by Bonami & Boyé (2006). There is one single lexeme equipped with a stem space with two slots. We will now turn to the factors determining which declension class a given noun belongs to. These factors are mostly hypothetical and have little to do with what information native speakers might store in their minds (*cf.* Blevins & Blevins 2009). If one is aware of all the background information concerning a word, one can inflect the noun according to the "rules", but if only the surface similarities and differences are considered (*cf. e.g.* Kálmán *et al.* 2012), the variation and the erosion of the inherited-borrowed dichotomy can be explained more easily.

Sixteen declension classes are listed for Hungarian Lovari in Hutterer & Mészáros (1967) based on gender, animacy, the nominative ending and the inherited or borrowed nature of the word. Some grammars, *e. g.* Matras 2002, consider palatalisation as an additional, separate criterion. The high number of classes serves as a motivation for the revision of these criteria, after which we will see that the information regarding animacy and the nominative ending is redundant, and the declension is determined by grammatical gender, and, to a lesser extent, the above-mentioned inherited-borrowed distinction. The tables below show the reduced system.

Masculine	bakró	'sheep'	sókro 'father-in-law'		
	(inhe	erited)	(borr	rowed)	
	singular	plural	singular	plural	
N	bakró	bakré	sókro	sokrurá	
А	bakrés	bakrén	sokrós	sokrón	
D	bakréske	bakrénge	sokróske	sokrónge	
L	bakréste	bakrénde	sokróste	sokrónde	
Abl	bakréstar	bakréndar	sokróstar	sokróndar	
Ι	bakrésa	bakrénca	sokrósa	sokrónca	
G	bakrésk-	bakréng-	sokrósk-	sokróng-	
V	bakrá	bakrále	sokrá	sokrále	

Table 4: masculine declension in Lovari

E in in .	1.1	6 + 2	a = a (dec als)		
Feminine	KIRI	ant	raca duck		
	(inhe	erited)	(be	orrowed)	
	singular	plural	singular	plural	
N	kirí	kirjá	rāca	rācí	
А	kirá	kirán	rācá	rācán	
D	kiráke	kiránge	rācáke	rācánge	
L	kiráte	kiránde	rācáte	rācánde	
Abl	kirátar	kirándar	rācátar	rācándar	
Ι	kirása	kiránca	rācása	rācánca	
G	kirák-	kiráng-	rācák-	rācáng-	
V	kirá	királe	rācá	rācále	

Table 5: feminine declension in Lovari

With regard to the paradigms, that is, the forms in the cells, the nominative ending is not important, as it is always dropped, be it a vowel, as can be seen in the tables, or a zero morpheme, and replaced by the oblique ending; thus nom. sing. *manúš* 'man' \rightarrow obl. *manušés-/manušén-*, following the pattern of *bakró* 'sheep', and nom. sing. *phen* 'sister' \rightarrow obl. *pheňá-/pheňán-*, following the pattern of *kirí* 'ant'.

The forms in the dark grey cells depend on the animacy status of nouns. The accusative form depends on the animacy of the noun: that of inanimate nouns is syncretic with the nominative (nom. and acc. *kher* 'house'). The accusative form of animate nouns is the oblique stem (nom. $b\bar{a}l\delta$ 'pig' and acc. $b\bar{a}l\delta$ s). The ending of the oblique stem does not depend on the animacy status of the given noun. The animacy status of a noun follows from a possible animacy

hierarchy (Matras 2002 based on Holzinger 1993 and Hancock 1995) among the nouns, with humans at one end and body parts at the other. Lovari maintains the broadest animacy split, with body parts inflected as animates as well (nom. *vast* 'hand' \rightarrow acc. *vastés*, as opposed to Romungro, a Central dialect spoken in Hungary: nom. *va* \rightarrow acc. *va*).⁴

Elšík 2000 gives a thorough account of the historical development of Romani nominal paradigms, where he notes that "two criteria are general for all nouns: the gender, and the shape of the base-form suffix" (Elšík 2000: 14). Hutterer & Mészáros 1967 also claim that the classification of a noun relies heavily on three factors: the meaning of the word (natural gender), the nominative ending (grammatical gender) or the meaning and the ending together. The first factor may be easily dismissed and replaced overall by grammatical gender, which is indeed a determining factor, as we can see from the tables; this is only complicated by the inherited-borrowed dichotomy in the masculine, as can be seen from Table 4.

a. ānró m. 'egg' → obl. sing. stem ānrés b. coló m. 'blanket' → obl. sing. stem colós-⁵

Gender neutralisation appears in many cases in the nominative; therefore, the nominative endings cannot be considered a determining factor of the declension of a given noun. The consonantal stems or, in other words, the stems ending in a zero morpheme, and words ending in *-i* may either be masculine or feminine. Again, the latter has diachronic reasons: it either goes back to the Proto-Romani inheritance (Elšík 2000) or later loanword adaptation in Lovari using an *-i* marker (Hutterer & Mészáros 1967). However, the declension of these loanwords has always been ambiguous. The words derived by the borrowed agentive suffix *-āri* also inflect according to the inherited pattern.

⁴ There have been attempts to formalise this hierarcy. Matras (2002) explains the role of the oblique stem as the marker of the accusative and other functions (the "Independent Oblique") in terms of topicality. Elšík (2000) claims that the direct object in case of inanimate nouns is expressed by the nominative form and suggests to treat the subject/direct object split as "hyper-paradigmatic". Thus, there would be no need to postulate separate paradigms based on the animate-inanimate distinction. This idea leaves the choice to pragmatic aspects.

⁵ The data presented in the paper mostly come from reliable written sources, which are free from the desire to codify or unify the language (Cech & Heinschink 1999a and 1999b; Hutterer & Mészáros 1967; Vekerdi 2000), but all the data were also confirmed and attested by the informant I worked with while writing the paper: Mária Nagy, an elderly, less educated lady, originally from Nagykálló in the north-eastern part of Hungary, currently residing in Budapest.

(2) a. phrāl m. 'brother' → obl. sing. stem phrālésb. phen f. 'sister' → obl. sing. stem pheňác. juhāsí m. 'shepherd' → obl. sing. stem juhāsósd. vitēzí m. 'brave warrior' → obl. sing. stem vitēzése. romňí f. 'woman' → obl. sing. stem romňáf. butjārí m. 'worker' → obl. sing. stem butjārés-

In Early Romani, there was a distinction between two different kinds of consonant-final masculine classes, whose nominative plural forms were different: *kher* 'house' \rightarrow nom. pl. *kherá*, as opposed to *vast* 'hand' \rightarrow nom. pl. *vast*. This identical plural form has disappeared by now from Lovari, and we find that the plural of *vast* is *vastá*.

Nouns ending in -o are exclusively masculine, nouns ending in -a are exclusively feminine.⁶ The former group is split again as for the oblique stem, because there are inherited and borrowed items among them. The masculine nouns ending in a consonant contain a special subgroup of "abstract nouns, which are characterized by a specific derivational suffix" (Elšík 2000), -ipen. Matras (2002) notes that its formants treasure an old oblique form -ip(e)nasin many Romani dialects instead of the expected *-ip(e)nes. He adds that the form nevertheless appears in Burgenland Romani "by analogy to the general masculine oblique formation" (Matras 2002: 84, and see also Halwachs 1998) in the form -ipes. However, -ipes is probably not an alternative formant of *ip(e)nes. Rather, the loss of the final nasal in Burgenland Roman as well as Lovari and more generally south of the Great Divide (a bundle of isoglosses in central Europe, for more detail cf. Matras 2005) resulted in the surface form -ipe (the abstract nouns thus becoming the only group with an -e as the nominative singular ending), to which the old oblique could not be applied, and so the general, inherited oblique spread onto the abstract nouns containing this particular suffix.

ending	Ø)		-0	- <i>i</i>			-е	-a
gender	m	f		m	m		f	m	f
inherited status	+	+	+	-	+	-	+	+	-
oblique singular	- <i>es</i> -	-a-	- <i>es</i> -	-os-/-es-	- <i>es</i> -	-os-/-es-	-a-	- <i>es</i> -	-a-

 Table 6: the pattern of nominative endings in Lovari after the historical developments described by Elšík (2000)

⁶ Not even this is completely unambiguous as Elšík (2000) mentions Romungro loanwords borrowed from Hungarian such as *komuništa* 'communist', which is masculine.

There are two important conclusions we can draw from the table: 1. the inherited or borrowed status of feminine nouns is irrelevant with regard to their declension; 2. there is variation in the oblique form of borrowed masculine nouns.

The small number of loans ending in -u resulted in the change of the final vowel to -o (e. g. original papu 'grandfather' $\rightarrow papo$) in some varieties. A similar phenomenon can be seen in the Lovari verbal system where verb stems ending in an -u-, which are exclusively made up loan verbs, are prone to losing their category and are recategorised as -i- stem verbs, which constitute the bigger class of loan verbs (Baló 2011 and Baló 2012a). The analogical effect based on frequency is conspicuous here and it heavily affects the low number of stems containing an -u- all over the Lovari morphology.

A very interesting dichotomy has existed within inherited feminine nouns ending in a consonant, namely that some of them are palatalised⁷ in the plural and in the oblique cases, while others are not. The examples below are taken from Elšík (2000).

(3) a. $\underline{z}uv$ 'louse' \rightarrow obl. sing. $\underline{z}uv\dot{a}$ b. suv 'needle' \rightarrow obl. sing. $suvj\dot{a}$ c. $p\bar{i}r\dot{i}$ 'pot' \rightarrow obl. sing. $p\bar{i}rj\dot{a}$ -

According to Elšík (2000) and Matras (2002), this is of Proto-Romani origin and a result of the infiltration of palatalised forms from other feminine paradigms. We can see double neutralisation here, between the nominative forms like žuv and suv on the one hand, and between the oblique forms suvjáand pīrjá- on the other. Elšík (2000) suggests that feminine nouns jotated in the oblique constitute a mixed class, where the nominative form resembles that of the consonant-final feminines, while the other forms are taken from feminines with a stem-final -i, where jotation is obligatory. Matras (2002) adds that "with pre-European feminines ending in a consonant, jotation is analogous, and hence often irregular" (Matras 2002: 83). Regularity is in the process of being reinstated, however; Elšík (2000) also adds that there is a tendency of de-jotation so as to avoid mixed paradigms, and that is why the originally jotated oblique singular form suvjá- 'needle' becomes suvá-, at least in certain dialects. This is justified by newly collected data, which proves that the process has not stopped, and the words where jotation could easily be triggered by the stem-final -i lose the palatalisation: oblique singular pīrá- instead of pīrjá-. This regularisation process consists of a simple

⁷ This phenomenon is called "jotation" in Romani linguistics.

analogy again, where the unpalatalised forms spread all over the feminine, except for words where the stem-final *-i* is preceded by a palatalised consonant, like, for example, *romnji* 'woman', *raklji* 'non-Roma girl', *angrustji* 'ring', *brādji* 'bucket'.⁸

4 The inherited-borrowed dichotomy

The difference between the inherited and borrowed inflectional patterns is most conspicuous in the masculine oblique. Diachronically, even the borrowing pattern was borrowed from Greek (Bakker 1997), and many of the first borrowings also came from Greek.⁹

Therefore, the declension of borrowed words in Early Romani showed a variety of oblique affixes, as far as the vowel is concerned: -os, -is, -us (Elšík 2000).¹⁰ The most obvious effect of analogy is the almost complete disappearance of the different vowels in Lovari (*e. g. sapúj* 'soap' \rightarrow obl. sing. *sapujos*- from Early Romani **sapuní(s)* \rightarrow obl. sing. **sapunís*-), preserving only the oblique in -os, which has thus become the only oblique affix of borrowed words.¹¹ The literature regarding the *-is* ending is slightly contradictory. Elšík 2000 considers the oblique singular in *-is* to be "a later development" at one point (Elšík 2000: 18), replacing the originally inherited form **sapunés*-. On the other hand, he also says (Elšík 2000: 23) that, at a later stage in the history of Romani declension, the original *-is* ending was replaced by *-os*, e.g. *doktorí* 'doctor' \rightarrow obl. sing. *doktorós*-, instead of the original **doktorís*-. This only justifies that "forms are [...] preserved because

⁸ Palatalisation in this position seems to be very common for alveolar stops, nasals and approximants.

⁹ Borrowing on the structural level is not rare in Romani: similarly to the borrowing pattern which was borrowed from Greek, there is a plural marker borrowed from Romanian, namely *-uri*, which often appears in the form *-ura* and attaches to borrowed nouns, but never to inherited ones: *juhāsí* 'shepherd' \rightarrow pl. *juhāsurá*, *sókro* 'father-in-law' \rightarrow pl. *sokrurá*.

¹⁰ Elšík 2000 also adds that the marker has thus lost its monomorphemic nature. However, as Baló 2012 notes with regard to the Lovari verbal system, the bimorphemic or monomorphemic nature of these markers becomes insignificant if we look at the analogy-based processes which have taken place and are taking place in the language. For instance, in Estonian where the partitive singular *lukku* of *lukk* 'lock' implies the short illative singular *lukku*, alongside *lukusse*, "even though neither *lukk* nor *-u* can be associated with the grammatical meaning "partitive" or "illative" (Ackerman, Blevins and Malouf 2009: 56). Their inference is that the deduction of new forms is facilitated by the knowledge of other forms.

¹¹ Although see the remarks regarding words with a stem-final -u.

they are constantly renewed by analogy" (Saussure 1966: 172).

Loanword markers, such as *-i* and *-o* can be used simultaneously, thus $tud\bar{o}šo'$ 'scientist' may coexist with $tud\bar{o}ši$, from Hungarian tudos. The oblique endings may vary, too, but not necessarily related to the nominal ending: $tud\bar{o}šo'/tud\bar{o}ši \rightarrow obl.$ sing. $tud\bar{o}šo'$ - as well as $tud\bar{o}šo'/tud\bar{o}ši \rightarrow obl.$ sing. $tud\bar{o}šo'$ - as well as $tud\bar{o}so'/tud\bar{o}si \rightarrow obl.$ sing.

As a well-known example, let us take a look at the case of the Greek-derived word $f\bar{o}ro$ 'town'. We learn that the forms of both the singular and the plural oblique stems are ambiguous: they may be $f\bar{o}r\acute{e}s$ - $f\bar{o}r\acute{o}s$ - and $f\bar{o}r\acute{e}n$ - $f\bar{o}r\acute{o}n$ -, respectively. As Elšík 2000 states, diachronically, $f\bar{o}r\acute{o}s$ - replaced $f\bar{o}r\acute{e}s$ -, so that the oblique form could resemble the nominative singular. This process, however, goes against the basic layout of the inherited inflection, where the oblique singular stem ends in *-es*-, no matter what the nominative ending is (for example nominative singular $b\bar{a}l\acute{o}$ 'pig' and oblique singular stem $b\bar{a}l\acute{e}s$ -). The case is more likely to be that the loss of the word-final consonant resulted in a form similar to many inherited nouns, and the oblique form is slowly taking on the inherited pattern, too – or at least re-acquiring it.

bāló : bālés- = fōró : x x = fōrés Figure 4: the Saussurean analogical proportion applied to the Lovari oblique stem

As for the oblique plural stem, Matras 2002 claims that *fōrén*- became *fōrón*-, possibly based on an analogy to the nominative singular. On the other hand, Hutterer & Mészáros (1967) mentions that the original form of the oblique plural stem of *sokro* 'father-in-law', from Romanian *socru*, is *sokrón*-, but it appears more and more frequently in the form *sokrén*-, and even the oblique singular stem can be *sokrés*- instead of *sokrós*-, as attested by informants. A similar process to the singular might have taken place in the plural as well.

marker	- <i>es</i> -	- <i>OS</i> -	-en-	- <i>0</i> 17-	- <i>a</i> -	-an-
function	oblique singular	oblique singular	oblique plural	oblique plural	oblique singular	oblique plural
possible gender	masculine	masculine	masculine	masculine	feminine	feminine
inherited or borrowed?	inherited borrowed	inherited borrowed	inherited borrowed	inherited borrowed	inherited borrowed	inherited borrowed
possible nominative singular ending	Ø, -0, -i	Ø, -o, -i	Ø, -o, -i	Ø, -o, -i	Ø, -a, -i	Ø, -a, -i
possible nominative plural ending	-e, -a, ura	-ura, -a	-e, -a, ura	-ura, -a	-a, -i	-a, -i

Table 7: the matrix of Lovari oblique endings

The feminine oblique plural was historically -en-, which is renewed in the Vlax dialects, possibly by analogy to the nominative plural, which ends in an -a (Matras 2002 based on Boretzky 1994). This is not generally true if we consider the data from Austrian Lovari; Cech & Heinschink (1999a) note that -en- is possible, too, for inherited words, and only -en- is possible for borrowed words. Considering this fact from a synchronic point of view, we might say that the -en- appears because it is "typical" of the oblique plural (*cf.* the masculine).

(4) a. *romňí* 'woman' → obl. pl. *romňén/romňán*b. *vórba* 'word' from Romanian *vorbă* → obl. pl. *vorbén*

If we look at the masculine now, we can see that it is also fairly uniform. Some anomaly only occurs among the nominative plural endings. Elšík (2000) mentions a similar anomaly with regard to Burgenland Romani: inherited and borrowed masculine nouns with a stem-final -o differ only in their oblique singular and nominative plural forms there and calls it "interaction between thematic and athematic classes" (Elšík 2000: 23). In Lovari, the situation is somewhat different: even if the nominative plural ending is -*ura* (*e. g. forurá*), the oblique endings may be either -*es-/-en-* or -*es-/-os-*.

5 Conclusion

In summary, we can say that basically there are two declension classes in Lovari: one masculine and one feminine. There is only one factor which slightly alters this: the inherited-borrowed dichotomy, described in detail by Boretzky (1989) and Bakker (1997) among others, However, as Elšík (2000) already notes, it is becoming blurred. Let us conclude by summing up the possible processes behind this phenomenon.

In the case of Romani, it is difficult to determine what the original forms of a certain word were exactly. What we can see here is that the forms $f\bar{o}r\acute{e}n$ -, $f\bar{o}r\acute{e}s$ -, sokrén- and sokrés- are in use, and they are spreading, weakening the role of the forms in -os-/-on-, which suggests that the inherited classes seem to exert an analogical force on the borrowings, at least as far as the masculine is concerned.

This can be related to the fact that many borrowings become obscure; for a bilingual speaker of Hungarian and Lovari, the words *tudōšó* and *juhāsi* might be transparent borrowings;¹² older borrowings, like *fōro* and *sócro* might become more integrated into the system. Generalisations may be made based on surface patterns that are "stronger" in some aspect; this might be the case for the historically deeper-rooted inherited pattern which is in constant opposition with the borrowed pattern, which is not as old but has become well-established due to the high degree of contact Romani has been exposed to.

References

- Ackerman F., J. P. Blevins, R. Malouf 2009. "Parts and Wholes: Implicative Patterns in Inflectional Paradigms" in J. P. Blevins, J. Blevins (eds), *Analogy in Grammar: Form and Acquisition*, Oxford: Oxford University Press, 54-82.
- Albright A. 2009. "Modelling analogy as probabilistic grammar" in J. P. Blevins, J. Blevins (eds), Analogy in Grammar: Form and Acquisition. Oxford: Oxford University Press, 185-213.

¹² As one Romungro informant remarked to the author during his fieldwork, referring to a lexical item borrowed from Hungarian but inflected in Romani: "we can only say this in Hungarian".

- Bakker P. 1997. "Athematic morphology in Romani: The borrowing of a borrowing pattern" in Y. Matras, P. Bakker, H. Kyuchukov (eds), The Typology and Dialectology of Romani. Amsterdam: John Benjamins, 1-21.
- Bakker P., Y. Matras 1997. "Introduction" in Y. Matras, P. Bakker, H. Kyuchukov (eds), *The Typology and Dialectology of Romani*. Amsterdam: John Benjamins, VII-XXX.
- Baló A. M. 2011. "A current trend or a historic remnant? The case of a Lovari verb-forming suffix". Poznań Studies in Contemporary Linguistics 47: 264-282.
- 2012. "Arguments from Lovari loan-verb adaptation for an analogy-based analysis of verbal systems" in F. Kiefer, P. Siptár, M. Ladányi (eds), *Current Issues in Morphological Theory: (Ir)Regularity, Analogy and Frequency*. Amsterdam: John Benjamins, 3-22.
- Blake B. J. 2000. "Case" in G. Booij, C. Lehmann, J. Mugdan (eds), Morphologie/Morphology. Ein internationales Handbuch zur Flexion und Wortbildung/An International Handbook on Inflection and Word-Formation. Berlin: Mouton De Gruyter, 1073-1090.
- 2001. Case. Cambridge: Cambridge University Press.
- Blevins J. P., J. Blevins (eds) 2009. *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.
- Bloomfield L. 1933. Language. New York: Holt.
- Bonami O., G. Boyé (2006). "Deriving Inflectional Irregularity." in S. Müller (ed.), Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar, Varna. Stanford: CSLI Publications, 361– 380.
- Booij G. 2007. "Construction Morphology and the Lexicon" in F. Montermini, G. Boyé, N. Hathout (eds), *Selected Proceedings of the 5th*

Décembrettes: Morphology in Toulouse. Somerville: Cascadilla Proceedings Project, 34-44.

— 2010. Construction Morphology. Oxford: Oxford University Press.

- Boretzky N. 1989. "Zum Interferenzverhalten des Romani. (Verbreitete und ungewöhnliche Phänomene.)". Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 42: 357-374.
- 1994. Romani. Grammatik des Kalderaš-Dialekts mit Texten und Glossar. Wiesbaden: Harrassowitz.
- 1995. "Interdialectal interference in Romani" in Y. Matras (ed.), Romani in Contact. The History, Structure and Sociology of a Language. Amsterdam: John Benjamins, 69-94.
- Bybee J. L. 1985. *Morphology. A Study of the Relation between Meaning and Form.* Amsterdam: John Benjamins.
- Cech P., M. F. Heinschink 1999a. Basisgrammatik. Arbeitsbericht 1a des Projekts Kodifizierung der Romanes-Variante der österreichischen Lovara (hrsgg. von Dieter W. Halwachs). Vienna: Verein Romano Centro.
- 1999b. Wörterbuch. Arbeitsbericht 3 des Projekts Kodifizierung der Romanes-Variante der österreichischen Lovara (hrsgg. von Dieter W. Halwachs). Vienna: Verein Romano Centro.
- Choli-Daróczi J., L. Feyér 1988. *Cigány nyelvkönyv*. [Romani Textbook] Budapest: Magyarországi Cigányok Kulturális Szövetsége.
- Eddington D. 2006. "Paradigm Uniformity and Analogy: The Capitalistic versus Militaristic Debate". *International Journal of English Studies* 6: 1-18.
- Elšík Viktor. 2000. "Romani nominal paradigms: their structure, diversity, and development" *in* V. Elšík, Y. Matras (eds), *Grammatical relations in Romani: The noun phrase*. Amsterdam: John Benjamins, 9–30.

Goldberg A. E. 1995. *Constructions*. Chicago: The University of Chicago Press.

- 2006. Constructions at Work. Oxford: Oxford University Press.

- Grumet J. 1985. "On the genitive in Romani." *in* J. Grumet (ed.), *Papers from the fourth and fifth annual meetings: Gypsy Lore Society, North American Chapter.* New York: Gypsy Lore Society, 84-90.
- Halwachs D. W. 1998. Amaro vakeripe Roman hi Unsere Sprache ist Roman: Texte, Glossar und Grammatik der burgenländischen Romani-Variante. Klagenfurt: Drava.

Hancock I. 1995. A handbook of Vlax Romani. Columbus: Slavica.

- Holzinger D. 1993. Das Romanes: Grammatik und Diskursanalyse der Sprache der Sinte. (= Innsbrucker Beiträge zur Kulturwissenschaft, 85.)
 Innsbruck: Verlag des Instituts für Sprachwissenschaft der Universität Innsbruck.
- Humboldt W. von. 1999. On Language. On the Diversity of Human Language Construction and Its Influence on the Mental Development of the Human Species. Cambridge: Cambridge University Press.
- Hutterer M., G. Mészáros 1967. A lovári cigány dialektus leíró nyelvtana. [A descriptive grammar of the Lovari Gypsy dialect] Budapest: Magyar Nyelvtudományi Társaság.
- Kálmán L., P. Rebrus, M. Törkenczy 2012. "Possible and impossible variation in Hungarian" in F. Kiefer, P. Siptár, M. Ladányi (eds), Current Issues in Morphological Theory: (Ir)Regularity, Analogy and Frequency. Amsterdam: John Benjamins, 23-49.
- Koptjevskaja-Tamm M. 2000. "Romani genitives in cross-linguistic perspective" in V. Elšík, Y. Matras (eds), *Grammatical relations in Romani: The noun phrase*. Amsterdam: John Benjamins, 123–149.
- Kraska-Szlenk I. 2007. Analogy: The Relation between Lexicon and

Grammar. Munich: Lincom Europa.

- Marle J. van. 1990. "Rule-creating creativity: analogy as a synchronic morphological process" in W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, J. R. Rennison (eds), *Contemporary Morphology*. Berlin: Mouton De Gruyter, 267-273.
- Masica C. P. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.
- Matras Y. 2002. *Romani: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- 2005. "The classification of Romani dialects: A geographic-historical perspective" in D. Halwachs, B. Schrammel (eds), *General and applied Romani linguistics*. Munich: Lincom Europa, 7-26.
- Mel'čuk I. A. 1986. "Toward a Definition of Case" in R. D. Brecht, J. S. Levine (eds), *Case in Slavic*. Columbus: Slavica, 35-85.
- Miklosich F. 1872-80. Über die Mundarten und Wanderungen der Zigeuner Europas. Vienna: Karl Gerold's Sohn.
- Rung A. 2011. Magyar főnévi alaktani jelenségek analógiás megközelítésben. [Certain phenomena of Hungarian nominal morphology in an analogybased framework] PhD dissertation. Eötvös Loránd University, Budapest.
- Sapir E. 1921. Language. New York: Harcourt, Brace & World.
- Saussure F. de. 1916/1966. *Course in General Linguistics*. New York: McGraw-Hill Book Company.
- Skousen R. 1989. *Analogical modeling of language*. Dordrecht: Kluwer Academic Press.
- 1992. Analogy and structure. Dordrecht: Kluwer Academic Press.
- Skousen R., D. Lonsdale, D. B. Parkinson 2002. *Analogical Modeling: An Exemplar-Based Approach to Language*. Amsterdam: John Benjamins.
- Stump G. 2001. Inflectional Morphology. Cambridge: Cambridge University

Press.

- Vekerdi J. 2000. A Comparative Dictionary of Gypsy Dialects in Hungary. Budapest: Terebess.
- Wurzel W. U. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer Academic Press.

A CROSS-LINGUISTIC INSIGHT ON AGENTIVE NOUN FORMATION IN ITALIAN AND FRENCH1

Bruno Cartoni Département de Linguistique, Université de Genève Fiammetta Namer ATILF-Université de Lorraine Stéphanie Lignon ATILF-Université de Lorraine

Abstract

The objective of this paper is to describe a contrastive dictionary-based study on word-formation and to show the advantages of the contrastive approach to improve the description of lexical morphology cross-linguistically. The present article focuses on the formation of agentive nouns. We first describe the methodology used to extract the contrastive data from a bilingual dictionary, then, through different case studies of specific agentive nouns formation, we show the benefit from such contrastive analysis and the new light that it sheds onto each morphological system taken individually.

1 Contrastive approach to morphology

Contrastive lexical morphology is the study of word-formation in two or more languages in parallel in order to highlight cross-linguistic morphological similarities and differences and to shed new light on the languages considered individually (see Lefer, 2011 for an exhaustive overview of contrastive morphology research). It is anchored in the recent tradition of contrastive linguistics (e.g., see James, 1980; Fisiak, 1983; Ringbom, 1994), which addresses different linguistic phenomena in a contrastive perspective.

Studying languages in contrast has often an "applied linguistic" objectives, such as the setup of second language learning method or the constitution of

¹ This work was supported in 2012 and 2013 by the Switzerland-France Joint Research Program EGIDE 'Germaine de Staël' Grant n°26433YL: '*MOCOCO* - MOrphologie, COrpus et analyse COntrastive'.

bilingual lexicography works. But the systematic comparison of two languages can also bring new theoretical insight. In the present study, we show that contrastive approach can ease a "meaning-based approach" (sometime referred as onomasiological) to the study of word-formation, i.e. an approach that first takes an object of study that is defined by semantic types or features, and then looks at the way different languages realize the object formally (through word-formation processes or not). Contrary to classical studies in morphology that usually focus on one particular wordformation process at a time (through the collection of potential complex lexemes coined with the considered process), the bi-directional contrastive analysis proposed here allows gathering all the complex lexemes that take part of a given semantic class, coined or not by a word-formation process. As a consequence, this approach allows considering the lexicon as a whole, and questions interfaces between lexical morphology and other linguistic means of denotation (syntactic paraphrases, simplex lexemes, borrowed lexemes, etc.). Thus, the boundaries of lexical morphology in the constitution of the lexicon are also assessed.

2 Word-formation processes under study: the agentive nouns

In this paper, we focus on nouns denoting agents, as described in lexical semantic studies such as Anscombre 2003, Busa, 1997, Cruse 1973, Van Valin & Wilkins 1996, or within a morphological perspective as in Booij 1986 and Grossmann 1998. Nouns denoting a human gave rise to different propositions of classification. The hierarchy of human nouns being multidimentional (according to various factors, such as gender, relation, profession, property, role), we focus here exclusively on characterizing agents, that are named following their behavior, or the ideal they claim to follow, and on classifying agents, or "actors", that regularly take part of an activity. (see Lo Duca (2004) and Roché (2011 a)).

Two main aspects have been taken into account in order to design the classification of agent nouns. The former is based upon lexical properties of nouns, the latter upon morphological ones. The identification of lexical properties of human nouns is guided by syntactic tests partly inspired by G. Gross's so-called object classes (see Gross 2009, 2011), and involving verb operators. For instance, the noun used to identify the agent in a clause headed by the French verb 'pratiquer' (to practice) or the Italian verb 'praticare' denote the actor of a professional or leisure activity (1), whereas the noun compatible with fr : 'exercer la fonction de' / it : 'esercitare/svolgere la funzione di' (to act as or to hold an office of) refers to a distinguished member of a professional or social hierarchy (2). In (1) and (2), tests and examples are given for French. In (2a), RelA stands for 'relation adjective'.

(1) a Qui est X? – X est quelqu'un qui **pratique**_{fr} N1 \rightarrow X est un NH_{[+} prof/leisure]

[Who is X? – X is someone who practices N1 \rightarrow X is a NH_[+ prof/leisure]

 $b N_{FR}$ = course, pêche à la ligne, médecine, journalisme, rugby \rightarrow NH_{FR} = coureur, pêcheur à la ligne, médecin, journaliste, rugbyman

(2) a Who is X? - X est quelqu'un qui exerce la-une fonction_{fr} (RelA + de NH) → X est un NH_[fonction]
[Who is X? - X is someone who (acts as / holds an office (of)) (RelA + NH) → X is a NH_[fonction]]

b RelA_{FR} : dirigeante, politique, ministérielle, officielle, présidentielle, exécutive, judiciaire \rightarrow NH = un dirigeant, homme politique, ministre, officiel, président, exécutif, magistrat

 $c NH_{FR}$: capitaine, ministre, procureur, chef, secrétaire, directeur, président, officier, juge, adjoint

By adapting and extending the initial set of Gross's operator verbs, we come to the hierarchy presented in Figures 1 and 2. Following Lo Duca (2004) partition, the first distinction is done between characterizing agents (Figure 1), who can be defined by 'support, are proponents/supporters/followers of X' for followers, and by 'have the habit of / are used to X' for behaviors, where X is morphologically related to the human noun. Actors (Figure 2) are sub-classified in the same way: for instance, among nouns who can be seen as agent of a 'salient activity', testcrossings allow to store those matching (1a) with the class 'Employee or manager', 'Manual activity or selling', 'Agriculture, livestock farming' or 'Sport and leisure' according to further characterizations.



Figure 1: Agent Nouns Classification: characterizing agents



Figure 2: Agent Nouns Classification : Actors

The second device which contributes to the distribution of agent nouns within this classification makes use of word formation knowledge obtained by following a Word Based approach (Aronoff 1994, Fradin 2003); it consists in defining each complex word with respect to its base(s), and consequently, to group words whose definition instantiates the same semantic pattern, and whose base belongs to the same semantic type. Table 1 shows what kind of properties can be inferred from morphological analysis. Approximately 70 semantic relations are needed to account for the set of agent nouns we are dealing with in this study.

Agent noun definition wrt its base	NBase Semantic type	Base	Agent Noun _{LG}	Formation pattern
He/she whose work or occupation consists of building	object	fiore _{IT} fleur _{FR}	fioraio _{IT} fleuriste _{FR}	-aio _{IT} -iste _{FR}
and/or selling Nbase	5	gioiello _{IT} bijou _{FR}	gioielliere _{IT} bijoutier _{FR}	-iere _{IT} -ier _{FR}
He/she whose work/occupation	location	posteggio _{IT} 	posteggiatore _{IT} gardien de parking _{FR}	-ore _{IT} N p N _{FR}
place is Nbase		bar _{FR}	barista _{IT} barman _{FR}	-ista _{IT} borrowing _{FR}
Musician who uses Nbase	Music Instrument	arpa _{IT} harpe _{FR}	arpista _{IT} harpiste _{FR}	-ista _{IT} -iste _{FR}
He/she whose work or occupation consists of doing Nbase	activity	maratona _{IT} marathon _{FR}	maratoneta _{IT} marathonien _{FR}	-eta _{IT} -ien _{FR}

Table 1: Semantic features provided by the morphological analysis

The multi-dimentional classification described above, including wordformation process, semantic reference to the base and semantico-referential criteria is then applied to the contrastive data extracted from the bilingual dictionary. The extraction methodology is described in the next section.

3. Bilingual dictionary as a source of data

Morphological studies usually rely on the collection of complex lexemes, either gathered from large language repository (such as dictionary) or collected in textual corpora. Contrastive morphology follows the same methodology, but the multilingual aspect with such an approach is particularly challenging.

While many contrastive morphology studies are based on multilingual corpora (either comparable or parallel, see for example Cartoni & Lefer 2001, and Lefer 2011), we decide here to focus on a bilingual dictionary, for several reasons. First of all, bilingual dictionaries (like monolingual ones) are exhaustive to some extend: they represent a stable representation of the state of the lexicon at a certain point, and do not depend on the context of production of texts in corpora. And this can be particularly crucial for some word-formation processes. For example, when looking into the French-Italian

parallel corpus extracted from the Europarl corpus (Cartoni and Meyer 2012), - just as (Cartoni & Lefer 2011) did for English-French-Italian contrastive analysis of negative word-formations, only 69 different occurrences of complex lexemes coined in *—ista* were found in the Italian side (compared to 954 occurrences in *—ista* of the bilingual dictionary used in this study). In addition, most of them were of the "characterizing agent type" (*communista*, "communist", *monopolista* "monopolist", probably because of the nature of such corpus (parliamentary debates).

The second advantage of bilingual dictionary is that they can be manipulated more easily than corpus data, where the extraction can require important manipulations.

However, bilingual dictionary have also important drawbacks that one should keep in mind when analyzing the extracted data. First of all, bilingual dictionaries do not contain neologisms, i.e. lexemes that were coined recently and that could bear witness to the productivity of the word-formation rule that coined it (in this respect, extraction from corpus combined with frequency or productivity factor would be better used). The second disadvantage depends on the way bilingual dictionaries are conceived. Rarely based on parallel corpora, bilingual dictionaries' quality strongly relies on the lexicographers' work, the way cross-linguistic equivalence is established. In the same vein, practical factors (such as the size of the dictionary and hence, of each entry) may also intervene with the quality of the dictionary. Nonetheless, as we will present in this paper, bilingual dictionaries represent an inestimable source of data for the study of two languages in contrast.

3.1 Extraction methodology

In this study, we rely on the Italian-French bilingual dictionary (Garzanti, 2006), that contains 65 308 entries in the It \rightarrow Fr direction, and 62 046 entries in the Fr \rightarrow It direction. The extraction methodology that was set in place to acquire bilingual data is a recursive one, where every cycle contains four steps.

First, we extracted French entries denoting human nouns and constructed with one of the most frequent suffixation rule *iste*, *-eur*, *-ien* and *-ier* (these suffixes were chosen because they were the most frequent translation of Italian entries suffixed in *-ista*, Cartoni&Namer 2012). In a second step, we duplicate entries that are polysemic and that give rise to different translations. For instance, *accessoiriste*_{fr} is translated by Italian *trovarobe* or *accessorista*, whether it denotes an employee working in a theater or in a garage.

In the third step, we categorize the nouns according to the classification described in section 2. In parallel, the translations proposed by the dictionary are also classified according to their formal structure. We distinguish lexemes that are no more analyzable (*athlète*_{fr} translation of *agonista*_{it}),

morphologically complex lexemes, distinguishing the word formation process that produced it, such as suffixation rules (e.g. suffixed in *-aire*: $mousquetaire_{\rm fr}$, translation of $moschettiere_{\rm it}$), or compounding rules (guardasigilli_{it}, translation of chancelier_{fr}). We also distinguished translation formed following syntactic pattern (it: datore di lavoro, translation of employeur_{fr}), and translation provided as a definition (fr : personne qui fait du marché noir, as a translation of borsanerista_{it}). Finally, in the fourth step, we reiterate the preceding steps for each Italian suffix that was the most frequently used in the translation of French lexemes in *-ien*, *-iste*, *-ier* et *-eur*.

At the end of this second cycle, we found the initial French suffixes, and also some new suffixes that are used to coin agentive nouns. If they are frequent enough, step 1 to 4 is reiterated.

Figure 3 sketches the iterative process to acquire the parallel data (only most frequent relations are displayed). From a selection of French suffixes, Italian suffixes are individualized. From these, parallel data in the other direction of translation (It \rightarrow Fr) are acquired. From double extraction, we also uncover non-morphological processes that are also included in the analysis (labeled as "other" in Figure 3).



Figure 3: Iterative process for gathering data (French → Italian → French)

Thanks to this methodology, we obtained an exhaustive set of nouns in French and Italian. In total, 2134 entries were gathered for the $Fr \rightarrow It$ direction, and 2429 for the Italian \rightarrow French direction. Some entries containing more than one translation, a total of 2608 translations were found into French, and 3008 into Italian. In those case, we reduplicate the entries to obtain 1:1 translation pairs.

Each pair of noun is provided as a translation equivalent by the dictionary, and so can be considered as cross-linguistically equivalent. Each pairs is then classified according to the semantico-referential classification (see Figures 1 and 2). Then, each part of the pair is classified according to their word-formation processes (or other means of denotation) and their semantic relation to the base.

4. Benefit from contrastive morphology.

In this section, we present the results of the extraction and classification procedures explained in Section 3. We first provide an overview of the figures that allows the quantification of noun formation processes in the two languages (Section 4.1). This first overview precedes a more detailed presentation of some interesting cases of divergences (Section 4.2).

4.1 Quantifying divergences between languages

The first quantitative analysis is presented in Tables 2 and 3, where we distinguish morphological and non-morphological translation (i.e. translation that is realized though a syntactic construction, a definition, or a non-constructed word). For the sake of clarity, Tables 2 and 3 display only the most frequent configuration.

En 🕹 It	Morphological
Fr 7 II	translation
-iste (732)	
actors (482)	76.3%
followers (222)	87.4%
behaviours (28)	96.4%
<i>-eur</i> (1482)	
actors (1216)	76.4%
followers (256)	76.2%
-ier (319)	
actors (310)	88.4%

 Table 2: Formal type of translation for suffixed nouns in *-iste*, *-ier*, *-eur* according to their semantic type.

It → Fr	Morphological translation
-ista (1096)	
actors (727)	70.6%
followers (310)	86.8%
behaviours (59)	69.5%
-ore(1175)	
actors (1065)	84.9%
behaviours (93)	78.5%
-iere (210)	
actors (209)	87.5%

 Table 3: Formal type of translation for suffixed nouns in *—ista*, *-iere*, *-ore* according to their semantic type.

This first distinction highlights discrepancies between the two morphological systems, discrepancies that vary according to the type of agents. For example, actors nouns displays more difference between Italian and French, for suffixed noun in *–iste, -eur, -ista* and *-iere* (resp. 23.7%, 23,6%, 29,4 and 22,5% of non-morphological translations), while followers are the most homogeneous in their formation in *-iste* or *–ista* (resp. 12,6 et 13,2 % only of non-morphological translations).

Another interesting measure to compute is to quantify the divergences of « coverage » of morphological process between languages. The notion of "mutual correspondence" (inspired by the one proposed by Altenberg 2002) measures the differences between two elements that are supposed to be equivalent. Table 4 shows the mutual correspondence for representative pairs of suffixes that are supposed to be equivalent because they are cognate (they more or less share the same forms). For each pairs, we give the proportion (in %) in which one suffix is translated by the other in the two directions of translation, the mean of this two proportions (the mutual correspondence) and the difference between the two proportions. This mutual correspondence allows evaluating the distance in comparison to a null hypothesis, according to which one suffix is always translated by it cognate (a mutual correspondence of 100%) and the difference shows the degree of discrepancies between the two suffixes of the pair.

	IT→FR	FR→IT	Mututal correspondance	Diff
ista⇔iste				
adepte	82.3%	82%	82.15	0.3
Actor	39.5%	54.4%	46.95	14.9
behavior	59.3%	85.7%	72.50	26.4
ore⇔eur				
Actor	72.5%	56.2%	64.35	16.3
behavior	65.6%	29.2%	47.40	36.4
iere⇔ier				
Actor	50.7%	25.8%	38.25	24.9

Table 4: Mutual correspondances for three pairs of suffixes IT-FR

These quantitative results help triggering several qualitative analyses, and completing or confirming previous studies such as (Fradin, 2003; Lignon et Roché, 2011; Roché 1997, 2004, 2011a,b) for French and (Bisetto, 1996, Dardano 1978, Lo Duca 2004) for Italian. Data in Table 2 also shows in which semantic area morphological divergences are the more important. While *followers* appear to be coined in a very homogenous way in the two languages with the suffixes *-ista* and *-iste* (diff. of 0.3), the differences are much more important for "actors" coined with pairs *-iere/-ier*, and *-ista/-iste*. The low proportion (25.8%) of French nouns in Xier translated in Xiere reflects a large availability of the Italian suffix and much more important constraint on the French rules. Another case in point is the important gap in *behaviour nouns* with the pairs *-eur/-ore*, which can be explained by the competition in Italian with the suffixe *-one*, which is very frequently an equivalent of *-eur* for this type of nouns (*critiqueur*_{ft}, *criticone*_{it}).

4.2 Qualitative analysis

From the quantitative analysis presented above, deeper qualitative analysis can be performed. In this section, we overview some of these analysis, with the objective of showing interesting insight contrastive analysis can uncover.

4.2.1 Semantico-referential Divergences

Onomasiological approach provides global picture on semantico-referential classes and the kind of items that are used to denote such class. In Table 5 below, we provide the figure of the different construction in French (by a derivational process, a borrowing (*borr.*), a compound (*comp*), etc. - see classification below) according to the semantic type of the Italian source entries. Over-abundant phenomena are highlighted in bold. In the rest of this

Italian Saliant		French									
Activities	eur	iste	ant	ien	ier	aire	borr	Def.	Sim plex	comp	NP
Artistic and cultural activities	62	87	1	11	7	1	3	0	12	7	54
Hunting and Fishing	5	0	0	0	0	0	0	0	0	0	3
Illicit activities	47	2	4	0	2	1	2	1	9	1	8
Manual activities selling	280	53	14	11	152	2	4	10	9	10	98
Sport and Leisure	85	39	1	0	10	0	12	1	4	3	50
Agriculture – Livestock farming	54	4	0	1	10	0	0	1	5	1	15
Employee or manager	115	44	13	2	44	11	8	6	18	15	49
Specialist	16	80	8	27	0	0	0	3	3	14	37

paper, we will focus on two specific categories, highlighted in grey in the table.

 Table 5: Distribution of the French denotation means according to the semantic-referential type.

In the table 5, the semantico-referential classes come from the classification performed on the Italian entries, and the other column described the different kind of French translations that are found in each class. In this section, we will focus on the "sport and leisure" category. For this category, several observations can be made. First, the two most frequent suffixes are *-eur* and *-iste*, both mainly on nominal bases. Second, an important amount of noun phrases are used to denote a sport activity (50 cases), which echoes the lack of "plasticity" of French *-iste* compared to its Italian counterpart *-ista*. Finally, there are many borrowings from other languages, and compare to the other semantico-referential classes; "sports and leisure" are the one where the "borrowing" is the most frequent. With respect to these loan words, it is interesting to stress that source Italian lexemes are morphologically constructed on borrow the actors noun in the foreign language (*crossman, judoka*).

But looking more closely at the data, it has to be noticed that these borrowings should be considered as "pseudo borrowing", because they did not exist *per se* in the source language (mostly English).

Similarly, we also noticed an important amount of *-eur* nouns (85/205), which are derived from nouns. In many cases, the base noun is a loaned word (e.g. *bridgeur*, *footballeur*, *hockeyeur* are respectively suffixed on *bridge*, *football*, *hockey*).

Finally, there is an important amount of NP bases (50/205), as shown in examples (3) a to e.

- (3) a. coureur de bobsleigh_{FR} (\leq bobbista_{IT})
 - b. lanceur de javelot_{FR} (\leq giavellottista_{IT})
 - c. joueur de poker_{FR} (\leq pokerista_{IT})
 - d. joueur d'échecs_{FR} (\leq scacchista_{IT})
 - e. joueur de tennis_{FR} (<tennista_{IT})

The interesting point here is that there are morpho-phonological constraints that prevent the coinage of suffixed lexemes in French (**tennisiste*, **javelotiste*), cf. 4.2.2. More interestingly, the construction of the French NP follows a very regular pattern, as shown in example (4).

(4) Xista_{IT} \rightarrow N1 prep N2_{FR} where N1=supertyp of Xista and N2 = X

Example (4) shows the regular pattern that coins a French noun phrase to translate the Italian complex lexem in *-ista*. The first noun (N1) of the NP is always a supertyp of *Xista (joueur, courreur, ...)* and the second noun (N2) is the translation equivalent of the base of the Italian complex lexeme (X).

4.2.2. Morpho-phonological constraints

Morpho-phonological constraints can affect the distribution of affixes in the semantic categories. The hierarchy of these constraints differs in various languages. Nouns of specialists provide a good illustration of this phenomenon. As can be seen from Table 5, French and Italian differ widely in the distribution of the suffixes that are used to coin lexemes belonging to the semantic category of 'specialists'. In Italian, the suffix *-ista* is almost the only affix in use. But French displays several suffixes as *-iste*, of course, but nouns also may end in *-ien*, *-eur*, and *-isant*. Since semantics cannot be invoked to explain such variations, we turn towards a morpho-phonological explanation.



Actually, the onset distribution before suffixes is very different according to the affix value. Figure 5 shows the percentage of the onset distribution before each affix in French.

From this table, various observations can be put forward:

_

- Some onsets seem unfavorable to *-iste* (/[, k, s, t, z/);
- In those cases, the compensation is not homogeneous:
 - For the onset /, k, t/, the suffix –*eur* takes over;
 - \circ For /s/, it is the suffix –*ien*;
 - \circ And, at least, for /z/, the suffix *-ant*.

The existence of a dissimilation constraint, which tends to prevent the consecution of two identical or nearly identical phonemes allows us to understand that *–iste* avoids sibilants.

A quite noticeable fact is that French uses mainly the suffix *-ien* after /s/ for specialist nouns, whose formation meets dissimilation constraint by employing the suffix *-ien* instead of the expected *-iste* (for more details: Lignon & Plénat, 2009)

In synchrony, the use of *-ien* to form specialist nouns occurs only when the base noun ends with a sibilant; the only specialist nouns built in *-ien* from bases whose onset value is not /s/ in earlier stages are prior to 16th century (*grammairien, historien, chirurgien*, etc.). Indeed, before that time, the suffix *-ien* was the only one available to build specialist nouns.

The sibilant |z| does not behave like /s/. On the one hand, *-iste* suffixed nouns from bases ending in |z| are in minority, due to the dissimilative

constraint. But on the other hand, the suffix used is not *-ien*. More precisely, such derived nouns are all *-isant* ending nouns (see $\S.4.2.3$).

Another onset after which the suffix *-iste* is in minority is /t/: the *-iste* suffix is not only refractory to sibilants, it is also sensitive to the presence of a /t/, still for dissimilative reasons. The suffix *-ant* also contains this consonant, even if it occurs only in the feminine forms (*-ant/-ante*). However, this is probably not the major reason why *-*ant is never found after a stem ending with /t/, cf. section below.

As Figure 5 shows, the suffix *-ien* is also conflicting with /t/, but for other reasons than dissimilation. Indeed, /t/ can be assibilated before *-ien*, but this assibilation is not regular. The speaker can hesitate on the way to pronounce it (/sj $\tilde{\epsilon}$ / or /tj $\tilde{\epsilon}$ /?) and generally tends to avoid this hazardous configuration by choosing the only affix capable of bearing with this onset, i.e. the suffix *-eur*. Furthermore, when applying to verb bases, *-eur* combines frequently with verbs' learned stem (also called 'stem 13', according to Bonami, Boyé & Kerleroux 2009). The most common form of this 'stem 13' ends with /at/. Hence, *-teur* endings for *-eur* suffixed nouns are very frequent sequences in the lexicon, which can have a ripple effect on the choice of this suffix.

4.2.3 Impact from "morphological series"

Most Italian nouns of specialist are suffixed with *-ista*, and, to a lesser extent, with *-ore*. On the other hand, French has at its disposal a more complete set of affixes to form specialist nouns: *-iste* is the most frequent one, but these nouns can also end with *-isant*, *-eur* and *-ien*.

Affix selection is sometimes ruled by morpho-phonological constraints (e.g. with *-ien*), but in other occasions, morpho-phonology does not interfere: series effects are at play, and their identification has been made possible through the comparison in French and Italian of semantically homogeneous classes. Namely, this is what happens with *-isant*. In fact, French nouns denoting specialists of a language or a civilization massively end with *-isant* (5), whereas the corresponding semantic class of nouns in Italian are always formed by *-ista* suffixation (6).

(5) Hébraïsant, latinisant, arabisant, italisanisant

(6) Ebrai(ci)sta, latinista, arabista, italianista

In French, two cases occur with this semantic class of nouns, for a given base X: either both Xiste and Xisant exist (*basquisant / basquiste; hébraïsant / hébraïste*), or only Xisant is attested (*arabisant*).

In the latter case, Xisant refers without any doubt to a specialist, and *-isant* fulfils the role normally plaid by *-iste*. But in the former case, Xiste and Xisant do not always share the same meaning: sometimes Xiste refers to the very specialist, whereas Xisant may denote a scholar learning the language in question, or even an (unskilled) amateur (*latinisant, latiniste*). Otherwise, the

-iste noun denotes an adept (a "basquiste" supports the independence of Basque Country) and therefore the *-isant* rival noun refers to the specialist.

5. Conclusion

In this paper, we have presented the advantages of the contrastive approach for the study of lexical morphology. We have shown that, through a sound empirically based methodology, contrastive data can be extracted, classified an analyzed. Through this classification, interesting insights are uncovered.

The main innovative aspect entailed by the approach is the onomasiological perspective of morphological analyses: the contrastive method allows to study one language through the lens of the other language, and so to uncover denotation means in a broad sense, including morphology, syntactic/idiomatic expression and lexicalized words. In particular, the lexicon of the two languages can be assessed through different perspectives.

With this methodology, we have built an organized set of lexical data both morphologically and semantically related. The case studies described in this paper have shown the important plasticity of some suffixes in Italian, and the stronger tendency in French to use loan words. Hence, the boundaries of lexical morphology processes in the construction of the lexicon can be drawn for each language, and the data shown in this study reveal that this boundary is not located at the same place in Italian and in French.

With the onomasiological approach, we can also measure gaps and overlaps between word formation processes, and thus highlight their specificities in each language. The cases that we have studied have shown, for instance, that *—ista* and *—iste* are the suffixes expected to form nouns denoting specialists. However, the application of this dominant rule is sometimes compromised by other parameters, such as well-formedness morpho-phonological constraints or series effects. In case of mismatch between French and Italian supposedly semantically equivalent processes, contentious cases are brought to light, which offers new insights to the morphological analysis in each compared language.

The proposed methodology, when reiterated for other morphological process, can open avenues for gathering parallel data and setting the stage for large scaled contrastive analysis of lexical morphology. If applied to many different language pairs, this can help in setting a map of word-formation in different languages.

References

- Altenberg, B. 2002. "Concessive connectors in English and Swedish". In Hasselgård, H., Johansson, S., Behrens, B. & Fabricius-Hansen, C. (Eds.) Information Structure in a Cross-Linguistic Perspective. Rodopi.
- Ansombre J.-C. 2003. "L'agent ne fait pas le bonheur : agentivité et aspectualité dans certains noms d'agent en espagnol et en français", *Thélème, Revista Complutense de Estudios Franceses*, 11, 11-27.
- Aronoff, M. 1994. Morphology by Itself. Cambridge: MIT Press.
- Bisetto, A. 1996. "Il suffisso -tore". *Quaderni patavini di linguistica* 14, 39-71.
- Bonami O., G. Boyé & F. Kerleroux. 2009. "L'allomorphie radicale et la relation flexion-construction". in B. Fradin, F. Kerleroux & M. Plénat, *Aperçus de morphologie*. Presses Universitaires de Vincennes. Saint-Denis: 103-125.
- Booij, G. 1986. "Form and Meaning in Morphology : the case of Dutch "agent-nouns"". *Linguistics* 24, 503- 518.
- Busa, F. 1997. "The Semantics of Agentive Nominals in the Generative Lexicon". In P. Saint-Dizier (ed) *Predicative Forms in Natural Language*. Amsterdam : Kluwer.
- Cartoni B. & Lefer M.-A. 2011. "Negation and lexical morphology across languages: insights from a trilingual translation corpus" in Fernandez-Dominguez J., Lefer M-A. & Renner V (eds). Poznan Studies in Contemporary Linguistics. Special issue on 'English Word-Formation in Contrast'
- Cruse, D. A. 1973. "Some thoughts on agentivity". *Journal of Linguistics* 9, 11-23.
- Dardano, M. 1978. *La formazione delle parole nell'italiano di oggi* (primi materiali e proposte). Roma: Bulzoni.

- Fisiak, J. 1983. "Present trends in contrastive linguistics". In K. Sajavaara
 (ed) Cross-language analysis and second language acquisition, Jyväskylä: University of Jyväskylä. 9-38.
- Fradin, B. 2003. *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.
- Gross, G. 1995. "À propos de la notion d'humain". In J. Labelle and C. Leclère (eds), *Lexiques-Grammaires comparés en français*. Amsterdam: John Benjamins: 71-80
- Gross, Gaston. 2009. "Sur le statut syntaxique des substantifs humains". In J.Cl. Ancombre, ed. Danielle Leeman (eds) *Mélanges*. Chambéry: Presses de l'Université de Savoie : 27-41
- Gross, Gaston. 2011. "Classification sémantique des noms humains collectifs" *Cahiers de Lexicologie* 98:65-82.
- Grossmann, M. 1998. "Formazione dei nomi di agentie, strumento e luogo in catalano". XVIII congresso internazionale di linguistica e filologia romanza:383-392.
- James, C. 1980. Contrastive analysis. Longman. Harlow
- Lefer M.-A. 2011. "Contrastive word-formation today: Retrospect and prospect". *Poznan Studies in Contemporary Linguistics*, 47(4). 645-682.
- Lignon, S. & Plénat, M. 2009. "Echangisme suffixal et contraintes phonologiques. " Aperçus de morphologie du français. B. Fradin, F. Kerleroux and M. Plénat (eds). Paris, Presses Universitaires de Vincennes: 65-81.
- Lignon, S, & Roché, M. 2011. "Entre histoire et morphophonologie, quelle distribution pour -éen vs -ien ? " In *Des Unités Morphologiques au Lexique*, éd. Michel Roché, 191-250. Paris: Hermès.
- Lignon, S. 2008. "Les noms de spécialistes en -iste et en -ien : le chimiste perturbé ou comment le physicien se réajuste". *Perturbations et*

Réajustements, Eds B. Vaxelaire, R. Sock, G. Kleiber, F. Marsac, Presses Universitaires de l'Université de Marc Bloch, Strasbourg, pp. 287-296.

- Lo Duca Maria G. 2004. "Nome di agente". La Formazione delle parole in italiano. M. Grossman et F. Rainer (éds), Max Niemeyer Verlag, Tubingen. 191-215.
- Ringbom, H. 1994. "Contrastive Analysis". In R.E. Asher (ed): The Encyclopedia of Language and Linguistics, Pergamon Press. Oxford. 737-742
- Roché, M. 1997 "Briard, bougeoir et camionneur : dérivés aberrants, dérivés possibles", in D. Corbin, B. Fradin, B. Habert, F. Kerleroux et M. Plénat (éd.), *Silexicales* 1, pp. 241-250.
- Roché, M. 2004. "Mot construit ? mot non construit ? quelques réflexions à partir des dérivés en -ier(e) " *Verbum* 26:459-480.
- Roché, M. 2011a. "Quel traitement unifié pour les dérivations en -isme et en iste ?" in M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique*, pp. 69-143.
- Roché, M. 2011b. "Quelle morphologie? " In M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique*, 15-39. Paris: Hermès.
- Van Valin, R D., and Wilkins, D P. 1996. "The case for 'Effector' : Case Roles, Agents, and Agency Revisited". In M. Shibatani & S. A. Thompson (eds.). *Grammatical Constructions*. Oxford: Clarendon Press. 289-322.

Dictionary

Garzanti francese-italiano, italiano-francese. 2006. I grandi dizionari Garzanti. Milano, Garzanti Linguistica

VERB STEM ASPECT IN MIAN

Sebastian Fedden University of Surrey

Abstract

Languages differ in how central a role aspect plays in their grammatical systems. In some languages aspect distinctions are central to the system, for example the derivationally related imperfective-perfective aspect pairs in Slavonic languages. In other languages there are only some instances in which aspect is relevant. I confront this with new data from Mian, a Mountain Ok language of Papua New Guinea. Perfective-imperfective stem aspect is an important part of Mian grammar because it imposes restrictions on how the stem can be further inflected. Similar systems can be found in all Mountain Ok languages; they are however rare in Papuan languages in general. A corpus study based on 456 Mian verbs (lexemes) shows that 27% form aspect pairs, derived by a range of diverse morphological processes: various suffixes (sometimes with concomitant tone change), one infix, stem change (in the form of various apophony patterns) or suppletion. About 40% are monoaspectual verbs, which lack either of the stems, in many cases for semantic reasons. There is a large number of biaspectual verbs (29%), which are the default morphological pattern. The evidence from Mian shows that aspect distinctions can be important grammatically, yet are restricted to a subset of verbs.

1 Introduction¹

Languages differ in how central a role aspect plays in their grammatical systems. On the one hand, aspect distinctions can run throughout the system. The derivationally related imperfective-perfective aspect pairs in Russian, for example *pisat*' (IPFV) vs. *na-pisat*' (PFV) 'write', and in other Slavonic languages are a familiar example. In these languages the aspectual opposition is essentially a lexical phenomenon, but it is so pervasive and systematic that it has to be considered part of the grammar (Bertinetto and Delfitto 2000: 210; Dahl 1985: 89). On the other hand, in some languages aspect is restricted to certain areas of the grammar, for example the Imperfect in Romance, which has an imperfective meaning, *e.g.* French *je mangeais* 'I was eating', but where the aspectual contrast is tied to the past.

In this chapter I confront this possibility space with new data on aspect from the lesser known Papuan language Mian, a Mountain Ok language of Papua New Guinea. In this language, as in all other Mountain Ok languages, many verbs show a perfective (PFV) vs. imperfective (IPFV) contrast encoded in the verb stem. The Mian system is noteworthy because of the wide range of morphological processes which are employed in aspect marking. In a few cases there are suppletive or near-suppletive aspectual stems. Other than that, aspectual stems are distinguished through various suffixes (sometimes with concomitant tone change), one infix or a stem change (in the form of various apophony patterns). There is also a large number of biaspectual verbs, which do not show any aspectual stem alternation.

Whenever a Mian verb stem is cited here its aspect value is given in brackets if a given stem is unequivocally perfective or imperfective, such as *wen* 'eat (IPFV)' and *dowôn*' 'eat (PFV)', *ga* 'say (IPFV)' and *ge* 'say (PFV)', or *unê* 'go

¹ Acknowledgements: I am grateful to the editorial committee and an anonymous reviewer whose insightful comments greatly improved the chapter. Versions of this chapter were presented at Les Décembrettes 8 - International Conference on Morphology, Bordeaux, December 6-7, 2012 and the South Eastern Morphology Meeting (SEMM) at the University of Surrey, Guildford, January 25, 2013. I would like to thank the respective audiences for helpful comments and discussion. My thanks go to Grev Corbett and Matthew Baerman for helpful discussion of multiple aspects of this chapter, and to Alexander Krasovitsky for discussion of aspect in Russian and for help with Russian examples and the search based on Zalisnjak's online dictionary. They are not responsible for the views expressed here. The work reported here was supported by the Arts and Humanities Research Council (UK), partly under grant AH/H500251/1 and partly (since April 2013) under grant AH/K003194/1. I thank this funding body for its support. Correspondence address: Sebastian Fedden, Surrey, Guildford GU2 7XH, UK. E-mail: s.fedden@surrey.ac.uk.
(IPFV)' and *un* 'go (PFV)'. For biaspectual verbs the gloss is given without any specification of aspect, *e.g. bali* 'bear fruit'.

As an introductory example from a traditional narrative consider the sentence in (1). All verb stems appear in boldface.

(1) Mian (Fedden 2011: 515, 523)

a.	gwáab=i	īmaye	dowôn' unê-b-ib=a
	small=PL.AN	themselves	eat.PFV go.IPFV-DS.SIM-3PL.AN.SBJ=MED

b.	haleb	ē-ta	te-s-e=a	[]	
	wild.boar	SG.M-EMPH	come.pfv-ds.seq-3sg.m.sbj=med	[]	l

bali-s-e=ta
bear.fruit-DS.SEQ-3SG.N1.SBJ=MED
'the small ones (*i.e.* young boars) themselves ate and were leaving when a wild boar came [...] it (a plant) bore fruit'

There are two events in clause (1a) in a serial verb construction, an eating and a leaving event. While the former is expressed as a bounded event with the perfective stem of 'eat' dowôn', the latter is conveyed as an unbounded event with the imperfective stem of 'go' $un\hat{e}$. While the leaving event is on-going the coming event in clause (1b) takes place, again expressed with a perfective stem for a bounded event, *i.e. te* 'come' (PFV)'. Clause (1c) occurs later in the text. A plant which has grown as a consequence of the plot of the narrative finally bears fruit. The stem *bali* 'bear fruit' is biaspectual. In (1c) the meaning is perfective but the same stem form could also appear in a form with imperfective meaning.

Stem aspect is an important category in Mian because it has ramifications for further inflectional possibilities of a verb stem. Stem aspect determines whether a verb stem can be directly inflected for various TAM categories or whether it needs to enter a construction with an auxiliary. For example, direct inflection with *-s* 'remote past' is possible for perfective stems, shown in (2a), while imperfective stems require an auxiliary in order to be inflected for remote past tense, as in (2b).

(2) a. Mian (Fedden 2011: 286) yōle éil=e a-nâ'-s-ib=e? well pig=SG.M 3SG.M.OBJ-kill.PFV-RPST-2/3PL.AN.SBJ=Q 'Well, did they kill the pig?'
b. Mian wen-bi-s-e=be eat.IPFV-AUX.IPFV-RPST-3SG.M.SBJ=DECL 'He was eating.' Ungrammatical: *wen-s-e=be

Biaspectual verbs do not show any such inflectional restrictions. Before we look at the wide range of morphological means employed in Mian aspect distinctions some typological and background information on the language is in order.

Mian belongs to the Ok family of languages, which is named after the widespread word *ok* 'river, water' (Healey 1964; Voorhoeve 2005) in these languages. The Ok family belongs to the larger Trans New Guinea (TNG) family (Wurm 1982; Ross 2005; Pawley 2005). The Ok family tree is given in figure 1.



Figure 5: The Ok family (based on Healey 1964 and Voorhoeve 2005)

Mian is spoken in Telefomin District of Sandaun Province in Papua New Guinea. The eastern dialect has approximately 1,400 speakers and forms the basis of a comprehensive grammatical description of the language (Fedden 2007; 2011). Most speakers under 75 also speak the variety of Neo-Melanesian Pidgin, Tok Pisin, spoken in Papua New Guinea. Most young speakers have some knowledge of English. Older male speakers above 50 years of age also speak the closely related neighbouring language Telefol. Mian is a word tone language. The domain in which five tonal melodies

In the examples, the five tonal melodies are written as follows: $m\bar{e}n$ 'child' (H), men 'string bag' (LH), $kl\hat{a}$ 'fix' (LHL), $ngun\hat{u}$ 'spread out (IPFV)' (HL). Low tone is unmarked, *e.g. fu* 'cook' (L). Mian is head-marking (Nichols 1996). Unmarked word orders are SV and AOV but constituent order is relatively free with the restriction that the verb always has to be clause-final and is only followed by an illocutionary particle. Word order in the noun phrase is more fixed. The language is strongly zero-anaphoric, *i.e.* noun phrases are mostly elided, if referent identity is retrievable from context or world knowledge. The syntax of the language is characterized by the frequent use of serial verb constructions and clause chaining.

All data presented in this chapter were collected by the author during fieldwork in Mianmin. Any Mian materials without a source are previously unpublished elicited examples.

This chapter has six sections. Following the introduction, §2 is an outline of stem aspect in Mian. §3 is a corpus study based on 456 Mian verb lexemes, which shows that despite the importance of aspect in the language there is a large number of biaspectual verbs, which are not concerned with aspect at all. Then I give an overview of how aspect is marked typologically (§4) and in Papuan languages more specifically (§5). Finally, in §6 I give a summary and offer some conclusions about the Mian system. The appendix is a complete list of all 456 verbs from the corpus study.

2 Aspect in Mian²

For about three quarters of the Mian verbs the stem encodes an aspectual alternation between perfective and imperfective. Dahl (1985: 78) defines perfective aspect as follows:

"A PFV verb will typically denote a single event, seen as an unanalysable whole, with a well-defined result or end-state, located in the past. More often then not, the event will be punctual, or at least, it will be seen as a single transition from one state to its opposite, the duration of which can be disregarded."

Perfective stems in Mian are used for describing a situation as a complete whole without making the internal temporal structure or duration of the situation explicit. The perfective stem in Mian can be used to describe complex situations comprising several phases, which can take some time, for example, making a fire, weaving a string bag, or building a house. The perfective stem has to be used for punctual, non-iterative situations. These have no internal structure and are thus incompatible with imperfectivity. While situations which are referred to by a perfective verb stem can be either punctual (like coughing) or durative (like building a house), in neither case does the perfective stem focus on the internal temporal structure of the situation. Mian shows the common correlation of perfective aspect and past time reference, which is stated in the definition above, but allows the perfective stem to also appear in irrealis forms with future time reference.

² This section draws heavily on Fedden (2011, ch. 8).

Use of the imperfective stem makes the internal temporal structure of a situation explicit. Imperfective aspect is used for non-bounded situations, *i.e.* for situations which hold habitually and for 'on-going' or continuous situations, whose duration is the focus of attention. Imperfective aspect can have habitual or continuous meanings, and it seems sensible to make a distinction here because Mian has a special habitual construction (involving the habitual form of the existential verb *bina*). This is never used for continuous, non-habitual situations. I use the term 'continuous' rather than 'progressive' because stative verbs, such as the existential verb, show the same perfective-imperfective contrast. Figure 2 summarizes the Mian aspectual distinctions.



Figure 2: Mian aspectual oppositions (adapted from Comrie 1976: 25)

What is striking about stem aspect in Mian is not that the language makes a perfective-imperfective distinction in the stem but rather that it uses a wide range of morphological means to do so, namely affixation, stem change (apophony), suppletion, and suprasegmental change (tone change). This is also true of the other Mountain Ok languages Telefol, Faiwol, Bimin (Healey 1964: 68) and Tifal (Healey & Steinkraus 1972). We do not have enough data on the Lowland Ok languages to gauge the extent to which they mark aspect in a similar way.

Table 1 gives an overview of the slots in the Mian verb template and the features expressed in each slot (both the morphosyntactic features person, number and gender, and the morphosemantic features aspect, tense and mood). Three segmented and glossed example verbs are provided.

$(OBJ-)^3$	Stem	-TAM	-SBJ	Translation
Person	Aspect	Tense	Person	
Number		Aspect	Number	
Gender		Mood	Gender	
	dowôn'	-Ø	-е	
	eat.PFV	REAL	-3.SG.M.SBJ	'he ate'
	wen	-b	-е	
	eat.IPFV	IPFV	-3.SG.M.SBJ	'he is eating'
i-	nâ'	-S	-е	
	hit.PFV	RPST	-3.SG.M.SBJ	'he hit them'

Table 1	:	Slots	in	the	Mian	verb	temp	late
I GOIC I	•	DICED		vii v	TATCOLL		COMP	

As the Mian tense, aspect and mood system is fairly complex, I can only illustrate the important points here. For a detailed treatment I refer the reader to the description in Fedden (2011: 282-314). The TAM slot can accommodate exactly one suffix from a set of tense, aspect and mood suffixes. These are the tense suffixes $-b^{(+H)}$ 'non-hodiernal past'⁴ and -s 'remote past', the aspect suffix -b 'imperfective', and the mood suffix - $n\sim-\emptyset$ 'realis' (-n after vowel, zero after consonant) and -amab 'irrealis'.⁵ In the following I show how stem aspect interacts with these categories.

A perfective stem suffixed with $-n \sim -\phi$ 'realis' has a default temporal interpretation as an immediate past, as in (3):

(3)Mian

imen=e

dowôn'-Ø-e=be taro=SG.N1 eat.PFV-REAL-3SG.M.SBJ=DECL 'He ate taro.'

The realis suffix $-n \sim -\emptyset$ can only be directly appended to perfective or biaspectual stems. The same holds for the two tense suffixes $-b^{(+H)}$ 'nonhodiernal past' in (4a) and -s 'remote past' in (4b).

³ While all finite verbs have a subject cross-referencing suffix, only seven (transitive) verbs are lexically specified to also index their object with a prefix, hence the brackets.

⁴ The superscript (+H) indicates that while all non-hodiernal past forms have the suffix -b most (but not all) of them have an additional exponent in the form of a tonal change.

⁵ Irrealis marking and stem aspect is more complicated. I refer the reader to the description in Fedden (2011: 292-294).

- (4) a. Mian dowôn'-b^(+H)-e=be eat.PFV-NHODPST-3SG.M.SBJ=DECL 'He ate (but not today).'
 - b. dowôn'-s-e=be eat.PFV-RPST-3SG.M.SBJ=DECL 'He ate (in the remote past).'

For these three categories to be expressed with imperfective stems they need to enter a construction with an auxiliary. This is illustrated for realis mood and the remote past in (5a) and (5b).

(5) a. Mian wen-bi-n-e=be eat.IPFV-AUX.IPFV-REAL-3SG.M.SBJ=DECL 'He was eating.'

b. wen-bi-s-e=be eat.IPFV-AUX.IPFV-RPST-3SG.M.SBJ=DECL 'He was eating (in the remote past).'

Only imperfective or biaspectual stems can be inflected with -b 'imperfective', as shown in (6), perfective stems cannot.

(6)	Mian				
	imen=e	wen-b-e=be			
	taro=sg.n1	eat.IPFV-IPFV-3SG.M.SBJ=DECL			
	'He's eating taro.'				

As expected, only imperfective stems can be used in the habitual construction, which requires the use of the habitual auxiliary *bina* attaching to the imperfective stem, as in (7).

(7) Mian imen=e wen-bina-b-e=be taro=SG.N1 eat.IPFV-AUX.HAB-IPFV-3SG.M.SBJ=DECL 'He habitually eats taro.'

For biaspectual verbs there is no aspect distinction in the stem. It is possible to use a biaspectual stem in all examples from (3) to (7). Two selected examples are given in (8a) and (8b):

(8) a.	Mian	
	bín=o	we-s-e=be
	floor=N2	sweep-RPST-3SG.M.SBJ=DECL
	'He swept the	floor (in the remote past).'
	1	

b. bin=o we-b-e=be floor=N2 sweep-IPFV-3SG.M.SBJ=DECL 'He's sweeping the floor.'

The perfective-imperfective contrast is not only important for knowing how a given verb stem can be further inflected for various TAM categories; it also plays a role in a valence-increasing operation. In the perfective, a recipient or benefactive argument has to be introduced by means of compounding a verb with the perfective stem of 'give' $-\hat{u}b'$ -, which has a quasi-applicative function (cf. Foley 2000: 380).⁶ This is illustrated in (9a). The indices merely show that a reflexive interpretation is not possible. In the imperfective, the recipient suffix is appended to the verb stem directly, as in (9b).

(9) a. Mian (Fedden 2011: 279)
éil=e mak=e
pig=SG.M other=SG.M
a-nâ'-ûb'-e-Ø-ib=a
3SG.M.OBJ-kill.PFV-give.PFV-PL.AN.R.PFV-DS.SEQ-2/3PL.AN.SBJ=MED
'They_l killed another pig for them_k, and then someone else ... (where l≠k)'

 b. Mian (Fedden 2011: 110) nakamín=e imen=o éil=e man=SG.M taro=PL.N1 pig=SG.M wen-ha-b-e=a eat.IPFV-3SG.M.R.IPFV-DS.SIM-3SG.M.SBJ=MED 'While a pig was eating a man's taro (, the man...)'

This section showed how important stem aspect is in Mian. It determines the inflectional potential of any given verb. Depending on stem-aspect value certain TAM categories are either impossible to be expressed or they can only be expressed with the help of an auxiliary. Furthermore, the way recipients/benefactives are introduced depends on stem aspect.

⁶ For reasons to say that $-\hat{u}b'$ - 'give (PFV)' has not yet fully grammaticalized into an applicative suffix, see Fedden (2010: 463).

3 The corpus study

The corpus study reported here is based on a sample of 456 Mian verbs. This is an exhaustive sample based on the wordlist provided in the appendix, which reflects the current state of description of the language. These 456 verbs are individual lexemes which are either imperfective, perfective or biaspectual. Some of these will be grouped into aspect pairs below, but the initial number of 456 verbs is not based on pairings.

3.1 Aspect pairs

Of these 456 verbs, we can group 124, *i.e.* a good quarter, into 62 aspectual pairs. Table 2 gives an overview of all morphological processes involved in relating the aspectual stems to each other. For suffixation it is not the case that one aspectual value is consistently associated with the suffixed form. Rather, sometimes the perfective stem is suffixed, sometimes the imperfective stem, sometimes both. The direction of derivation is indicated in table 2 by arrows. Note that in several cases there is also a tone change. This will be taken up briefly below. Although there is a plethora of different processes involved in aspect marking, some trends can be discerned. Affixation is by far the most common means and suffixation by *-ka* is the most frequent process of marking the imperfective. If aspect is marked by apophony, the pattern is most frequently /a/ in the perfective and /u/ in the imperfective.

While reduplication is not involved in relating perfective and imperfective stems to each other, it is associated with the imperfective aspect in that there are a few verb stems, such as *fufun* 'blow (IPFV)' or *sasan* 'moan (IPFV)', which are monoaspectual and imperfective-only, *i.e.* for these there are no perfective counterparts.

Suprasegmental changes in the form of tone changes do occur but they are never the sole exponent of stem aspect; at least not in the aspect pairs in the current corpus. It is entirely possible that such examples turn up in an expanded corpus. About one third of all aspect pairs also involve a tone change. I have so far not been able to find any consistent patterns, so it is difficult to say that a certain tone melody or tone change is an exponent of a certain stem aspect.

The aspect affixes (or at least some of them) are possibly verbs etymologically which have grammaticalized into suffixal stem aspect markers. Presumably, these verbs were originally used in a serial verb construction consisting of a lexical verb and a functional verb expressing aspect. This sequence later underwent univerbation, resulting in a single verb with an aspectual stem suffix.

Process	PFV		IPFV	Gloss		Count	
	fa	\rightarrow	fa-ka	'make fire'	10		
	ge	\rightarrow	ge-n	'build'	2		
	-têm'	\rightarrow	-tem-ê '	'see'	3		
	dei-lâ'	←	dei	'remove hair'	5		
Suffixation	nge-la	\leftrightarrow	nge-n	'beg'	3	34	
	halbu-a	←	halbû'	'fold'	2		
	ulilò	←	uli	'roll thread'	1		
	me-le	\leftrightarrow	me-n	'touch'	1		
	he-na	\leftrightarrow	he-n	'seek'	7		
Infixation	fu <el>a</el>	$fu < el > a \leftarrow$		'bathe'		4	
	ifa		ifu	'serve (food)'	11		
Apophony	ge		ga	'say'	1	13	
	biki		bika	'close, squeeze'	1		
	dowôn'		wen	'eat'			
Suppletion	baa		0	'say'		11	
	-ma		-san	'plant'			
						62 pairs	

Table 2: Morphological processes relating aspectual stems to each other

For at least one suffix the verbal origin is still apparent: *-lo* marking perfective aspect, which presumably comes from the verb $l\partial$ 'hit (PFV)'. This etymology is plausible because in many Papuan languages which express aspect by means of a serial verb construction perfective aspect is typically marked by verbs of contact, such as 'hit' (Foley 1986: 145). A second stem aspect suffix in Mian with a verbal origin might be *-ka* for imperfective aspect, whose possible verb etymon is *ka* 'put (IPFV)'.

In addition to these aspect pairs we find more complex relations in which three or more stems are related to each other. These are given in table 3, which is an exhaustive listing.

PFV	IPFV	
-nâ, -lò	-е	'hit, kill'
te~tl, tlaa(n)	te, tle	'come'
un, on, unaa(n)	unê	'go'
п	bi~bl, biaa, bina, biaan	'stay, exist'

Table 3: Sets of three or more stems

For 'hit, kill' there is a single imperfective stem, -e 'hit, kill (IPFV)' opposite two perfective stems, $-n\hat{a}$ ' 'hit, kill (PFV)' and $-l\hat{o}$ 'hit, kill (PFV)', which each have the full paradigm of a perfective verb.

For 'come' there is a perfective stem $te \sim tl$. Allomorph choice depends on the following segment, tl is used before /i/, and te elsewhere. The perfective stem tlaa(n) is only used in medial verbs in clause chaining constructions where it also expresses 'same subject' and sequentiality of events. The imperfective stem is te. The stem tle is used for iterative or habitual situations.

For 'go' there are two perfective stems: *un* and *on*. The former is used for all tense and mood forms except the remote past (*on-s-io=be* [go.PFV-RPST-2/3PL.AN.SBJ=DECL] 'They went (in the remote past).') and the non-hodiernal past (*on-b*^(+H)-*i=be* [go.PFV-NHODPST-1SG.SBJ=DECL] 'I went (but not today).'), for which *on* must be employed. The perfective stem *unaa*(*n*) is only used in medial verbs in clause chaining constructions where it also expresses 'same subject' and sequentiality of events.

The existential verb has one perfective stem n. In the imperfective bi (bl before /i/) is used except for habituals (bina) and the non-hodiernal past (biaa). Finally, the stem biaan is used only in medial verbs in clause chaining constructions where it also expresses 'same subject' and simultaneity of events. For details on the inflection of the existential verb, see Fedden (2011: 299-303).

3.2 Monoaspectual verbs

Mian has a large number of monoaspectual verbs, which exist either only in the perfective or only in the imperfective. A subset of 162 verbs out of the total of 456 verbs in the corpus only have a perfective stem. They typically have punctual meanings, *e.g.* $-\dot{a}$ 'let go (PFV)', *bina* 'shoot (PFV)', *kilo* 'begin (PFV)', *kimi(n)* 'go out (of fire) (PFV)', and *mâa*' 'stand up (PFV)'. Due to the lack of an imperfective stem the possibility of further inflecting these verbs for aspect is restricted. For instance, inflection with -b 'imperfective', a reduced (grammaticalized) form of the existential verb *bi* 'stay' (Fedden 2011: 288), requires an imperfective or a biaspectual stem. Monoaspectual perfective stems cannot take this suffix. This is illustrated in (10):

Mian
 *mâa'-b-e=be
 stand.up.PFV-IPFV-3SG.M.SBJ=DECL
 Intended: 'he's standing.'

In order to express the continuation of the result of the event described by the perfective stem $m\hat{a}a'$ 'stand up (PFV)' the imperfective auxiliary *-bi* is required, as in (11):

 Mian mâa'-bi-Ø-e=be stand.up.PFV-AUX.IPFV-IPFV-3SG.M.SBJ=DECL 'He's standing.'

The meaning of these forms is precisely the continuation of the result of the event described by the perfective stem rather than a continuous form, as can be seen from the meaning of (12):

(12) Mian as=e kimin-bi-Ø-e=be fire=SG.N1 go.out.PFV-AUX.IPFV-IPFV-3SG.N1.SBJ=DECL 'The fire went out and stays (out).' *'The fire is going out.'

There are also monoaspectual verbs with only an imperfective stem, but much fewer: only a subset of 22 verbs out of the total of 456 verbs in the corpus. They typically have durative meanings, *e.g. dlan* 'last (of money or supplies) (IPFV)', *ei* 'fly (IPFV)', *en*- 'hurt (IPFV)', *gen* 'be sick (IPFV)', and *un* 'hum (IPFV)'. As with monoaspectual perfective stems, there are restrictions regarding the possibility of further inflecting these verbs for aspect. For instance, inflection with $-n\sim-\emptyset$ 'realis' requires a perfective or a biaspectual stem. Monoaspectual imperfective stems cannot take this suffix directly. This is illustrated in (13):

(13)	Mian			
	*wan=e	ei-n-e=be		
	bird=sg.m	fly.ipfv-real-3sg.m.sbj=decl		
	'Intended: The bird flew.'			

In order to allow inflection with $-n \sim -\emptyset$ 'realis' or the tense suffixes which can attach directly only to a perfective or a biaspectual stem, such as -s 'remote past' or $-b^{(+H)}$ 'non-hodiernal past', use of the auxiliary is required, as in (14):

(14) Mian wan=e ei-bi-n-e=be bird=SG.M fly.IPFV-AUX.IPFV-REAL-3SG.M.SBJ=DECL 'The bird was flying.'

The fact that monoaspectual verbs only have a perfective stem is in many cases related to the lexical semantics of the verb. Often these verbs have punctual semantics and it is therefore expected that they do not produce imperfective forms since they do not need them. I would claim that such a restriction of aspectual distinctions based on verb semantics does not apply to verbs with durative semantics, which produce perfective as well as imperfective forms. This might be the reason why there is a large number of perfective-only stems but only a small number of imperfective-only ones.

3.3 Biaspectual verbs

Apart from aspect pairs and monoaspectual verbs there is a sizeable subset of 132 biaspectual verbs, which have only a single stem. Biaspectual verbs are not subject to any restrictions with respect to further inflection. These verbs typically have durative meaning but allow perfective meaning, where internal temporal constituency is disregarded. Examples are: *bu* 'hunt', *dli* 'dance', *ein* 'burn', *fu* 'cook', *gâala* 'destroy', *haa* 'weave', *hebâ* 'lean', *ki* 'measure', *klâ* 'fix', *singa* 'pour', *waa* 'swim', and *yo* 'initiate'. There are a few biaspectual verbs with punctual meaning, *e.g. tila* 'flash (of lightning)', which would typically be perfective but the same stem can as well be used in the imperfective with an iterative meaning. In the case of *tila* 'flash' this would be an iteration of flashes.

A comparison with Russian, which has derivationally related aspectual opposites like Mian might be instructive. Only about 3% of Russian verbs are biaspectual.⁷ This is what we would expect in a language where aspect plays such a central role. Consider the biaspectual verb *issledovat*⁻ 'investigate', which can have an imperfective (15a) or a perfective reading (15b).⁸

(15) a. Russian

On	vsju	žizn′	issledoval	tvorčestvo	Puškina.
he	all	life	investigate.PST	works	of.Puškin
'All h	is life he	has been	investigating Push	kin's works.'	

b.	Russia	1					
	Posle to	ogo kak	on	issledov	al	ètu	problemu,
	after	-	he	investig	ate.PST	this	problem
	on	poterjal	k	nej	interes.		
	he	lost	to	it	interest		
	'After l	he had inv	estigated	this pro	blem he l	ost any i	nterest in it.

⁷ Andrej Zaliznjak's online Russian dictionary (available at http://starling.rinet.ru/cgibin/query.cgi?root=/usr/local/) lists a total of 24,874 verbs; 24,142 out of 24,874 verbs, i.e. 97% (rounded to full numbers), are either perfective or imperfective. This means the proportion of biaspectual verbs is only 3%. On a detailed treatment of biaspectual verbs in Russian, see Anderson (2002) and Janda (2007).

⁸ Both examples were provided by A. Krasovitsky, p.c.

Russian has many hundred biaspectual verbs, but compared to its whole verbal vocabulary the proportion is very small. The list of biaspectual verbs provided here is based on Zaliznjak & Šmelev (2000) and Es'kova (2009). First, there are 18 native Russian verbs, *e.g. bežat* 'escape', *kaznit* 'execute' and *obeščat* 'promise'. In addition there are 714 biaspectual verbs all of which involve one of the following suffixes: *-ova-*, *e.g. issledovat* 'investigate', *-irova-*, *e.g. likvidirovat* 'liquidate', *-izova-*, *e.g. organizovat* 'organize', or *-ficirova-*, *e.g. klassificirovat* ' classify'. All except the ones in *-ova-* are loans. Anderson (2002) reports that 95% of all biaspectual verbs in Russian are in fact loans (based on Zaliznjak 1977).

The perfective-imperfective contrast applies across the whole Russian lexicon. The proportion of biaspectual verbs is very low and biaspectuality is mainly a feature of loan words in Russian.

3.4 Results of the Mian corpus study

Before looking in more detail at how aspectual distinctions are morphologically expressed cross-linguistically I summarize the results of the corpus study in table 4.

	Part of	Part of complex	Monoaspectual	Total
	aspect pair	relation	verb	
Perfective	62	8	162	232
Imperfective	62	8	22	92
Total	124	16	184	324
	T 11 4 D	14 641 347	4 1	

Table 4: Results of the Mian corpus study

The cells in the row labelled 'total' in table 4 mean the following: 124 verbs can be grouped into aspect pairs. 16 verbs form more complex aspect relations involving more than two stems. This is the case for the stems of 'come', 'go', 'hit, kill' and the existential verb. 184 verbs are monoaspectual. 232 verbs have a perfective form. 92 verbs have an imperfective form. A total of 324 verbs out of 456 verbs, *i.e.* 71% (rounded to full numbers) are either perfective or imperfective. The remaining verbs, *i.e.* 132 (or 29% of 456), are biaspectual.

The following section takes a look at how aspect is marked typologically. This will put us in a better position to appreciate the complexity of the Mian system, which uses almost all available morphological processes.

4 Aspect marking typologically

Typologically, aspect is most commonly expressed periphrastically or inflectionally (Bybee *et al.* 1994; Bybee & Dahl 1989) and there is a

tendency for certain aspect categories to be expressed either periphrastically or inflectionally (Dahl 1985). For instance, a periphrastic construction is typically used for the progressive, as in familiar Indo-European languages, *e.g. I am working*, and its Italian and French equivalents *Sto lavorando* and *Je suis en train de travailler*. A similar observation can be made about the perfect. Perfectives and imperfectives, on the other hand, are typically expressed with a bound form.

This form-meaning correlation is related to the degree of grammaticalization. A gradual generalization of meaning is paralleled by a gradual reduction in form and fusion with the verb (Bybee & Dahl 1989: 56). The progressive as a less general meaning shows less grammaticalization of form, whereas perfective and imperfective are more abstract meanings and hence show a stronger degree of grammaticalization of form. Furthermore, the perfective-imperfective distinction is typically expressed by more complex means than are found in other areas of morphology, including other areas of the tense-aspect system (Dahl 2000: 16). The degree of lexical idiosyncrasy is high and it is often not predictable from one verb to another how the opposition is realized. This is exactly the situation we found in Mian.

In the following I give a brief overview of the means of encoding the perfective-imperfective distinction cross-linguistically, which employs the full inventory of morphological marking, namely affixation, suppletion, stem change, suprasegmental means (tone or stress contrasts), and reduplication.

Marking by affixation is common. In Russian (Comrie 1976: 90), simple (*i.e.* non-affixed) verbs are imperfective, *pisat* 'write (IPFV)', while prefixed verbs are perfective, *na-pisat* 'write (PFV)'. For many prefixed – and thus perfective verbs – suffixation with *-iva* is possible to derive an imperfective verb, *e.g. vy-pisat* 'write out (PFV)' vs. *vy-pisyvat* 'write out (IPFV)' (see figure 3 below).



Figure 3: Russian aspect pairs (Comrie 1976: 90)

Marking can be by suppletion, for example in Russian *brat* 'take (IPFV)' vs. *vzjat*' 'take (PFV)' or Georgian *xedav* 'see (IPFV)' vs. *naxav* 'see (PFV)'

(Comrie 1976: 98). Another possibility is stem change, found for example in Modern Greek *graf*- 'write (IPFV) vs. *graps*- 'write (PFV) (Comrie 1976: 96). Suprasegmental means such as stress or tone can play a role in aspect marking. In Russian, stress can have an effect (together with a vowel change), *e.g. 'brosit'* 'throw (PFV)' with stress on the first syllable vs. *bro'sat'* 'throw (IPFV)' with stress on the second syllable (G. Corbett, p.c.). Tone is important in the Gur language Gulimancema, spoken in Burkina Faso, in which the perfective-imperfective contrast is marked with a difference in tone for a subset of verbs (Delplanque 2009), *e.g. nè* 'drink (IPFV)' vs. *né* 'drink (PFV)' or *kpésè* 'wash (IPFV)' vs. *kpésé* 'wash (PFV)'.⁹ To use only a change in tone to express a different aspect is certainly rare.

Finally, reduplication frequently expresses meanings associated with the imperfective (Bybee *et al.* 1994: 168), such as continuous, habitual and progressive, as in the language Mwera, a Bantu language from Tanzania, *e.g.* simple $ta\hat{w}a$ 'tie' and reduplicated $ta\hat{w}a$ - $ta\hat{w}a$ 'tie over and over again' (Harries 1950: 77, cited in Bybee *et al.* 1994: 160). Repetition can have a similar function. Consider the following example from Tok Pisin in (16), in which the verb is simply repeated. Unlike reduplication, which is a morphological process resulting in a single word, repetition yields two separate words:

(16)	Tok Pisin (Mühlhäusler 1985: 383; segmentation and glosses mine)					
	yu	fait-im	pig,	fait-im	fait-im	
	2sg	strike-TR	pig	strike-TR	strike-TR	
	'You strike the pig, and keep on striking it.'					

To summarize, cross-linguistically, the marking of the perfectiveimperfective distinction uses all available morphological processes, concatenative and non-concatenative. Mian uses a wide range of these in its stem aspect system. In the following section I narrow the focus somewhat and situate Mian against aspect systems commonly found in Papuan languages.

5 Aspect marking in Papuan languages beyond Mian

Multiple means of aspect marking in the verb stem and the existence of biaspectual verbs is a property of all Mountain Ok languages. Table 5 gives one example each of suffixation, suffixation with tone change, apophony, and suppletion for the Mountain Ok language Telefol (Healey 1964: 68). An

⁹ I thank Matthew Baerman for bringing this language to my attention.

	Perfective	Imperfective	Gloss
Suffixation	dàá/daa	dàá-kà/—	'put'
Suffixation plus tone	ùn/un	ún-è/un-ê	ʻgo'
change			
Apophony	bók <i>ò/baa</i>	bákà/(o)	'say'
Suppletion	ùndú/(beilò')	wèè/—	'prepare'
Biaspectual	bíkí/biki	bíkí/—	'pierce'

example of a biaspectual verb is given as well. Telefol forms are given first Mian cognates follow after the slash.¹⁰

Outside of Ok, aspect distinctions in the stem can be found in the Papuan languages Marind (Drabbe 1955) and Kiwai (Ray 1932), in Korafe (Farr 1999: 22-25) and in Abui (Kratochvíl 2007: 82-86). While Marind, Kiwai and Korafe are Trans New Guinea languages like Mountain Ok (Ross 2005), Abui (a member of the Alor-Pantar family) probably is not (Holton *et al.* 2012). Apart from the absence of tone, the morphological processes which account for the derivational relation between aspect-specific stems in Korafe are almost as complex as in Mian. Farr (1999: 22) lists stem change through final vowel shift, partial reduplication of the stem, suppletion and suffixation of *-ut* to derive the imperfective stem, the only obvious differences to the Mian system being that reduplication is a more common process in Korafe and that suffixation is restricted to a single suffix. In the other languages mentioned above the range of means is more limited. Marind uses only suffixes and Kiwai only stem change (Foley 1986: 146-148). Abui likewise only employs stem change (Kratochvíl 2007: 83).

Most Papuan languages, especially Trans New Guinea languages, convey aspectual distinctions periphrastically by means of serial verb constructions with verbs like 'hold', 'take', or 'hit' for perfective aspect and 'stay', 'stand', 'lie', or 'do' for imperfective aspect (Foley 1986: 145). This is the case in Kalam (Pawley 1993; 2008). Example (17) illustrates the expression of continuous aspect by means of a verb serial construction with *md*- 'stay':

(17) Kalam

b	yob	ag	md-p-ay
man	big	sound	stay-prs-3pl
'The b	ig men a	re talking	

¹⁰ Non-cognate forms appear in brackets; — indicates that the form is lacking in Mian, *e.g.* Mian *daa* 'put (PFV)' is a monoaspectual verb.

High incidence of serial verb constructions in the expression of aspect is related to the high frequency of serial verb constructions in these languages in general (Foley 1986: 143). Other Trans New Guinea languages in which serial verb constructions play a role in aspect marking are Fore (Scott 1978), Enga (Lang 1973), Dani (Bromley 1981), and Barai (Olson 1981). Outside of Trans New Guinea this is the case in Iatmul (Sepik; Staalsen 1972) and Vanimo (Skou; Ross 1980).

Inflectional systems of aspect marking can be found "in the south central area of New Guinea, from southern Irian Jaya, through the Western Province of Papua New Guinea" (Foley 1986: 146). For example, Marind and Kiwai have sets of inflectional aspect suffixes in addition to the aspect-specific stems mentioned above. So does Oksapmin (Loughnane 2009). Nen makes use of aspect-sensitive number augments, which are suffixed to the verb root, and verbs are classed into biaspectual, imperfective-taking and perfective-taking according to their inflectional potential (Evans 2011).

Suprasegmental means of indicating aspect are rare in New Guinea languages, as they are world-wide. A language which exploits tone for aspect marking is Iau (Lake Plains family, West Papua) (Bateman 1986; 1990). Iau is a syllable tone language (cf. Donohue 1997: 356-357) with eight distinct tonal melodies. In the contrastive pair tai^5 'has fallen' vs. tai^2 'was falling' (tone indicated by superscript numbers) the difference in aspect is purely marked by a difference in tone.¹¹

Summing up, the wealth of morphological means of aspect marking found in Mian and in Mountain Ok more generally is certainly rare in the New Guinea context.

6 Summary and conclusion

Mian uses a wide range of morphological means of marking stem aspect: suffixation, infixation, apophony and suppletion. In about a third of all aspect pairs, there is a concomitant change in tone. Suffixation is quite unordered. There are nine different patterns. What is more, sometimes the perfective stem bears the suffix, sometimes the imperfective stem.

Not only are the means of aspect marking in Mian diverse, they are also unpredictable which strongly suggests that we are dealing with derivationally related verbs, rather than one verb which is inflected for aspect. The wealth of means of aspect marking is also in contrast with how the other morphosemantic categories, tense and mood, are formed in the language,

¹¹ The Iau aspect system is complex involving a multitude of contrasts. Only a small number of examples can be given here. For a summary of the Iau aspect system, see Foley (2000: 381-382).

which exclusively use affixation (with tonal change being restricted to the non-hodiernal past). This means that stem aspect in Mian is more lexical compared to other categories, as is indeed typical of perfective-imperfective oppositions cross-linguistically. While all Mountain Ok languages share this wealth of means, it is rare in Papuan languages in general, in which verb serializations are very often the means of choice in aspect marking.

Stem aspect in Mian is very likely an old system, which is suggested by the fact that it is mainly an unsystematic lexicalized system of oppositions, grammaticalized to a certain degree, in the sense that stem aspect impinges upon the inflection potential of the verb. The remaining sub-regularities suggest the existence of a more productive earlier system which then broke down. As systems of this type are not created ex nihilo, the existence of many different morphological processes would be unexpected, if at some point different lexical items had simply become associated as aspectual pairs.

What is especially curious about the Mian system is that while stem aspect is an important grammatical category we find that only about a quarter of Mian verbs enter into morphologically distinct aspect pairs. For many verbs the lack of an imperfective stem can be explained straightforwardly by their punctual semantics. Almost one third of the verb vocabulary is biaspectual, *i.e.* there is only one aspect-neutral form. This means that a substantial minority of verbs are not concerned with aspect at all.

Nowadays the biaspectual pattern is clearly the default in Mian, as almost 30% of the verbs follow this pattern, which is by far the most common of any pattern of aspect alternation. The question why Mian has so many biaspectual verbs is difficult to answer synchronically. In further research diachronic work within Mountain Ok might show that these are reconstructable, in which case the past situation is comparable to the present.

Abbreviations

1 - first person, 2 - second person, 3 - third person, AN - animate, AUX - auxiliary, DECL - declarative, DS - different subject, DUR - durative, EMPH - emphatic, GPST - general past, HAB - habitual, HPST - hesternal past, IPFV - imperfective, IRR - irrealis, M - masculine, MED - medial, N1 - neuter 1, N2 - neuter 2, NHODPST - non-hodiernal past, OBJ - object (theme or patient), PFV - perfective, PL - plural, PRS - present, PST - past, PUNCT - punctual, Q - question, REAL - realis, R - recipient (object), RPST - remote past, SBJ - subject, SEQ - sequential, SG - singular, SIM - simultaneous, TR - transitive.

References

- Anderson C. 2002. Biaspectual Verbs in Russian and Their Implications on the Category of Aspect. Honors Thesis. University of North Carolina, Chapel Hill.
- Bateman J. 1986. *Iau Verb Morphology*. (NUSA: Linguistic Studies of Indonesian and Other Languages in Indonesia, 26.) Jakarta: Badan Penyelenggara Seri NUSA, Universitas Atma Jaya.
- 1990. "Iau segmental and tone phonology" in B. K. Purwo (ed), Miscellaneous Studies of Indonesian and Other Languages in Indonesia, Part X (NUSA: Linguistic Studies of Indonesian and Other Languages in Indonesia, 32). Jakarta: Badan Penyelenggara Seri NUSA, Universitas Atma Jaya, 29-42.
- Bertinetto P.M., D. Delfitto 2000. "Aspect vs. Actionality" in Ö. Dahl (ed), Tense and Aspect in the Languages of Europe. Berlin: Mouton de Gruyter, 189-225.
- Bromley H.M. 1981. A Grammar of Lower Grand Valley Dani. (C-63.) Canberra: Pacific Linguistics.
- Bybee J.L., Ö. Dahl 1989. "The creation of tense and aspect systems in the languages of the world". *Studies in Language* 13-1: 51-103.
- Bybee J.L., R. Perkins, W. Pagliuca 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Language of the World*. Chicago: University of Chicago Press.
- Comrie B. 1976. Aspect. An Introduction to the Study of Verbal Aspect and Related Problems. Cambridge: Cambridge University Press.
- Dahl Ö. 1985. *Tense and Aspect Systems*. Cambridge: Cambridge University Press.
- 2000. "The tense and aspect systems of European languages in a typological perspective" in Ö. Dahl (ed), *Tense and Aspect in the Languages of Europe*. Berlin: Mouton de Gruyter, 3-25.

- Delplanque A. 2009. "Identité des langues Gur du Burkina Faso". Département de Linguistique Université de Ouagadougou, Université de Tours. Ms.
- Donohue M. 1997. "Tone systems in New Guinea". *Linguistic Typology* 1: 374-386.
- Drabbe P. 1955. *Spraakkunst van het Marind* [A Grammar of Marind]. Studia Instituti Anthropos 11.
- Es'kova N.A. 2009. "Formal'nye sootnoš enija meždu členami vidovyh par v russkom jazyke" [The formal relationship between members of aspectual pairs in Russian] *in* Z. Saloni and J.D. Apresjan (eds), *Metody Formalne w Opisie Języków Słowiańskich* [Formal methods in the description of Slavic languages]. Białostok.
- Evans N. 2011. "Some puzzles of aspect in Nen". Paper presented at the Research Centre for Linguistic Typology (RCLT) Seminar, LaTrobe University, Melbourne, Jan 13, 2011.
- Farr C.J.M. 1999. The Interface Between Syntax and Discourse in Korafe, a Papuan Language of Papua New Guinea. (C-148.) Canberra: Pacific Linguistics.
- Fedden S. 2007. *A grammar of Mian, a Papuan language of New Guinea*. PhD Dissertation. University of Melbourne. [Available at:
- http://dtl.unimelb.edu.au/R/Q3IMDPAXP5X5YN6AL1J9ATXP9XF6YJS5Q 66UV7UFSAFGH1HFAG-00483?func=dbin-jumpfull&object id=67554&local base=GEN01&pds handle=GUEST]
- 2010. "Ditransitives in Mian" in A. Malchukov, M. Haspelmath, B. Comrie (eds), *Studies in Ditransitive Constructions: A Comparative Handbook*. Berlin: De Gruyter Mouton, 456-485.
- 2011. A Grammar of Mian. Berlin: De Gruyter Mouton.

- Foley W.A. 1986. *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
- 2000. "The languages of New Guinea". Annual Review of Anthropology 29: 357-404.
- Harries L. 1950. *A Grammar of Mwera*. Johannesburg: Witwatersrand University.
- Healey A. 1964. A Survey of the Ok Family of Languages, Reconstructing Proto-Ok. PhD Dissertation. Australian National University, Canberra.
- Healey P., W. Steinkraus 1972. A Preliminary Vocabulary of Tifal With Grammar Notes. Santa Ana, CA: Summer Institute of Linguistics.
- Holton G., M. Klamer, F. Kratochvíl, L.C. Robinson & A. Schapper 2012."The historical relations of the Papuan languages of Alor and Pantar". Oceanic Linguistics 51: 86-122.
- Janda L. A. 2007. "Aspectual clusters of Russian verbs". *Studies in Language* 31: 607-648.
- Kratochvíl F. 2007. A Grammar of Abui. PhD Dissertation. Faculteit der Letteren: Universiteit Leiden.
- Lang A. 1973. *Enga Dictionary, With English Index* (C-20.) Canberra: Pacific Linguistics.
- Loughnane R. 2009. *A Grammar of Oksapmin*. PhD Dissertation. University of Melbourne.
- Mühlhäusler P. 1985. "Syntax of Tok Pisin" *in* S. Wurm and P. Mühlhäusler (eds), *Handbook of Tok Pisin*. Pacific Linguistics C-70. Canberra: Australian National University, 341-421.
- Nichols J. 1996. "Head-marking and dependent marking grammar". *Language* 62: 56-119.

- Olson M. 1981. Barai Clause Junctures: Towards a Functional Theory of Interclausal Relations. PhD Dissertation, Australian National University, Canberra.
- Pawley A. 1993. "A language which defies description by ordinary means" *in*W.A. Foley (ed), *The Role of Theory in Language Description*. Berlin: Mouton de Gruyter, 87-129.
- 2005. "The chequered career of the Trans New Guinea hypothesis: Recent research and its implications" in A. Pawley, R. Attenborough, J. Golson, R. Hide (eds), Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-speaking Peoples. Canberra: Pacific Linguistics, 67-108.
- 2008. "Compact versus narrative serial verb constructions in Kalam" in G. Senft (ed), Serial Verb Constructions in Austronesian and Papuan Languages. Canberra: Pacific Linguistics, 171-202.
- Ray S. 1932. *A Grammar of the Kiwai Language, Fly Delta, Papua*. Port Moresby: Government Printer.
- Ross M. 1980. "Some elements of Vanimo, a New Guinea tone language". Pacific Linguistics A-56. Canberra: Pacific Linguistics, 77-109.
- 2005. "Pronouns as a preliminary diagnostic for grouping Papuan languages" in A. Pawley, R. Attenborough, J. Golson, R. Hide (eds), Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuanspeaking Peoples. Canberra: Pacific Linguistics, 15-66.
- Scott G. 1978. *The Fore Language of New Guinea*. (B-47.) Canberra: Pacific Linguistics.

Smith J., P. Weston 1987. "Mian-English wordlist". Ms.

Staalsen P. 1972. "Clause relationships in Iatmul". Pacific Linguistics A-31. Canberra: Pacific Linguistics, 45-69.

- Voorhoeve B. 2005. "Asmat-Kamoro, Awyu-Dumut and Ok: An enquiry into their linguistic relationships" in A. Pawley, R. Attenborough, J. Golson,
 R. Hide (eds), Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-speaking Peoples. Canberra: Pacific Linguistics, 145-166.
- Wurm S. 1982. Papuan Languages of Oceania. Tübingen: Narr.
- Zaliznjak A.A. 1977. *Grammatichskij Slovar' Russkogo Jazyka* [Grammatical dictionary of Russian]. Moscow: Russkij yasyk. [Online version available at http://starling.rinet.ru/cgibin/query.cgi?root=/usr/local/share/starling/morpho&morpho=1&basena me=morpho\zaliznia\dict]
- Zaliznjak A.A., A.D. Šmelev. 2000. Vvedenie v russkuju aspektologiju [Introduction into Russian aspectology]. Moscow: Jazyki russkoj kul tury.

Appendix - List of verbs

This appendix provides a Mian-English wordlist consisting of all the 456 verbs on which the study reported in this article is based. The list is based on my own fieldwork and has been complemented with Smith and Weston (1987), which is their dictionary of the Mian language. I have carefully checked the material from Smith and Weston (1987) with one speaker and added tonal and grammatical information.

An entry consists of underlying tone specification, word class plus information of transitivity, meaning, and stem aspects. Any specification that is doubtful at this stage is preceded by a question mark. Any specification that is unknown has a question mark in its place.

The following abbreviations are used:

BIASP	biapectual	SS	same subject
HAB	habitual	TP	Tok Pisin loan
IPFV	imperfective	v/ambitr.	ambitransitive verb
joc.	jocular	v/ditr.	ditransitive verb
MONOASP	monoaspectual	v/func.	functional verb
PFV	perfective	v/intr.	intransitive verb
PL	plural	v/tr.	transitive verb
SG	singular		

All obligatory argument-indexing affixes are indicated with dashes on verb stems as follows:

- V_{stem}: The verb does not index the object or does not have one, *e.g. fu* 'cook (transitive)', *un* 'hum, drone (IPFV, intransitive)'.
- -V_{stem}: The verb obligatorily indexes its object with a prefix, *e.g.* -*têm*' 'see (PFV)', or -*ô* 'take (PFV)'.
- -Vs_{tem}-: The verb obligatorily indexes two objects, *i.e.* the theme (with a prefix) and the recipient (with a suffix), *e.g.* -*ûb*'- 'give (PFV)'.
- V_{stem}-: The verb obligatorily occurs compounded with the verb 'give', which is followed by an object suffix in the perfective and requires an object suffix (from a somewhat different set) but no compounding with 'give' in the imperfective, *e.g. fote-* 'chase away'.
- A dash in brackets, *e.g.* (-)*ba* 'put into (PFV)', indicates that the affix is optional and can be left out without changing the valency of the verb. Even if the argument prefix is not there the verb is still transitive.

А

- -a L. v/tr. hurt (PFV, MONOASP).
- -à' HL. v/tr. let go, leave, lose, send (PFV, MONOASP).
- -aa L. v/tr. rouse (e.g. prey), set off (PFV, MONOASP).
- *aal ge* L/L. *v/intr*. be ashamed (BIASP).
- aalob olaketou L/L. v/intr. be uncowed (PFV, MONOASP).
- áan LH. v/intr. lie, sleep (SG subject; áala LH. with PL subject) (BIASP).
- afen L. v/intr. be awake, be alive (IPFV, MONOASP).
- afetâ LHL. v/tr. divide (BIASP).
- al tlia- L/L. v/tr. be angry (PFV).
- *ale-* L. *v/tr*. show, teach (BIASP).
- ali L. v/tr. squeeze pandanus sauce (onto taro dough) (BIASP).
- andaakbû LHL. v/tr. squash, weigh down (BIASP).
- angkikî LHL. v/intr. be alert, be forearmed (PFV, MONOASP).
- atdî LHL. v/tr. throw into fire (PFV, MONOASP).

atli- L. v/tr. be angry (IPFV).

atomâa' LHL, v/tr. join (PFV). atosaan L. v/tr. join (IPFV).

В -ba ?L. v/tr. put inside (PFV). ba L. v/intr. dry up (bodies of water, i.e. rivers, lakes, puddles) (BIASP). ba L. v/intr. grow (of plants) (BIASP). -ba L. v/tr. fill; cover (of liquids) (PFV, MONOASP). baa L. v/tr. say; tell (PFV). bafu L. v/tr. boil (BIASP). bai-L. v/tr. cut out (SG object) (PFV, MONOASP). baka L. v/tr. accompany (PFV, MONOASP). balì HL. v/intr. come up (of plants), bear fruit (BIASP). HL. v/ambitr. cut, split (SG object) (PFV, MONOASP). balò bām tabâ H/LHL, v/tr. unfold (BIASP). bām tou H/?L. v/intr. unfold, open up like a flower (BIASP). batlâa' LHL. v/ambitr. tear apart (vine, leaf, or bark) (PFV, MONOASP). be L v/tr. pull back; masturbate (BIASP). be L. v/intr. walk (IPFV, MONOASP). HL. v/tr. prepare, pack (PFV, MONOASP). beilò' beitaalô LHL. v/intr. be weak, be lazy (PFV, MONOASP). LHL. v/tr. accompany (BIASP). beke LHL. v/tr. almost cut off, cut so that the 'cut-off' bit is still attached (e.g. lid of can, bekelâ bit of wood) (PFV, MONOASP) belâ LHL. v/ambitr. cut alongside, operate, open up (PFV, MONOASP). L. v/ambitr. open (IPFV). beta betelâ LHL. v/ambitr. open (PFV). bi~bl L. v/intr. exist, stay, remain (IPFV). bî' L. v/tr. close, shut (eyes) (BIASP). -bià HL. v/ambitr. throw, fall off (PFV, MONOASP). biaa^H LH. v/intr. stay, exist, remain (Non-hodiernal past) (IPFV). biaan L. v/intr. exists, stay, exist remain (SS IPFV). L. v/intr. swell (PFV, MONOASP). bibila bika L. v/ambitr. close, squeeze; pierce (e.g. insect bites), nail; explode (IPFV). biki L. v/ambitr. close, squeeze; pierce (e.g. insect bites), nail; explode (PFV). LHL. v/ambitr. sew; be blocked (by vegetation) (PFV, MONOASP). bikî ' bikilâ LHL. v/tr. wring out (BIASP). L. v/intr. be burnt (PFV, MONOASP). bila bilâ(ka) LHL. v/tr. cut to pieces (IPFV, MONOASP). L. v/intr. stay, exist, remain (HAB) (IPFV). bina L. v/tr. shoot, penetrate (PFV, MONOASP). bina LHL. v/ambitr. fell, push down, hit; fall (PFV, MONOASP). blelâ ' bo L. v/tr. search for tracks (BIASP). L. v/intr. come together forming a crowd (PFV, MONOASP). bobola bokâ LHL. v/tr. marry many (women) (BIASP). bou L. v/tr. beat (with palm); strum (guitar) (BIASP) bouwâ' LHL. v/tr. wait in vain, search in vain (PFV, MONOASP). -bù HL. v/tr. bury (PFV, MONOASP). bu L. v/tr. hunt (BIASP). -bu L. v/tr. plant (tobacco, sugar cane) (BIASP).

-bu L. v/tr. put inside (IPFV).

bû LHL. v/tr. hook, grip, catch (PFV, MONOASP).

bukowâ' LHL. v/intr. swell to the point of bursting (PFV, MONOASP).

bukù ' HL. v/tr. spread widely, take over (PFV, MONOASP).

D

da- L. v/tr. help (BIASP). daa ?L. v/intr. dwell, abide (BIASP). daa L. v/tr. put/leave somewhere (PFV, MONOASP). dà' HL. v/ambitr. break off (PFV). daha L. v/tr. peel wood or palm bark (BIASP). LHL. v/tr. throw out (embers or ashes) (PFV, MONOASP). dabiâ dafama L. v/intr. be bumpy (PFV, MONOASP). L. v/tr. take down (e.g. from a hook); play (e.g. recorded music or video) (PFV, dai-MONOASP) daka L. v/ambitr. break off (IPFV). dalbi L. v/intr. molder (of fruit) (BIASP). dalò HL. v/intr. stop breathing, die (PFV, MONOASP) HL. v/tr. break off (SG object, e.g. a banana) (PFV, MONOASP). dalò dama L. v/intr. grow up (of people), grow big (of plants), thrive (PFV, MONOASP). LHL. v/tr. close a hole (e.g. in the ground, a fence, tree bark, etc.) (PFV). datanâ' LHL. v/tr. remove by pulling out (PFV, MONOASP). datlâa datôu LHL. v/intr. sit down (PFV, MONOASP). datunû LHL. v/tr. close a hole (e.g. in the ground, a fence, tree bark, etc.) (IPFV). dê' LHL. v/?tr. not want, not like, desist, stop (PFV, MONOASP). LHL. v/intr. wait a while (PFV, MONOASP). defâ' dei L. v/tr. leave, avoid, deny (PFV, MONOASP) dei L. v/tr. pick (leaves); pluck, remove (hair) (IPFV). LHL. v/tr. pick (leaves); pluck, remove (hair) (PFV). deilâ LHL. v/ambitr. break apart, separate (PFV, MONOASP). delâ delaakma L. v/tr. pour out (PFV). delaaksaan L. v/tr. pour out (IPFV). delò HL. v/tr. break off (PL object, e.g. bananas) (PFV, MONOASP). dena L. v/intr. clear off (PFV, MONOASP). deskî' LHL. v/intr. turn (BIASP) di L. v/tr. give (breast), serve (BIASP) di L. v/tr. tie (rafters) (BIASP) dì' HL. v/tr. tighten (bowstring) (BIASP). difibma L. v/intr. warm up (PFV). difibsaan L. v/intr. warm up (IPFV) HL. v/tr. do garden work (IPFV). dìk' dîk' LHL. v/tr. make a hole (IPFV, MONOASP) dika L. v/tr. dig (BIASP). dikihâ' LHL. v/intr. diffuse, spread (of liquid) (PFV, MONOASP). dikila ?L. v/tr. do garden work (PFV) dilbî' LHL. v/ambitr. scatter (BIASP). L. v/intr. become numb (of limb), become cold (of fire) (PFV, MONOASP). dimila diwatdî LHL. v/tr. sweep into the fire (BIASP). L. v/intr. last (of money or supplies, not temporal) (IPFV, MONOASP). dlan dli L. v/intr. dance (TP singsing) (IPFV, MONOASP). dli- L. v/tr. push. (BIASP). dò HL. v/tr. sew (IPFV, MONOASP). dobô LHL. v/intr. topple, fall down (PFV, MONOASP). dobô

- LHL. v/tr. feel, taste, affect (PFV, MONOASP).
- dogi L. v/tr. lead pig or child on a leash (BIASP). HL. v/tr. spread out (a flat object) (BIASP). dogunù

doi-L. v/tr. untie (SG object) (PFV, MONOASP). dokaa L. v/tr. behold (BIASP). doketòu HL. v/tr. remove (IPFV, MONOASP). doko-L. v/tr. forget (PFV, MONOASP). dokô' LHL. v/intr. change (place) (PFV, MONOASP). dokomaa L. v/tr. dance in the spirit house, joc. rock the house (PFV). dokomsaan L. v/tr. dance in the spirit house, joc. rock the house (IPFV). dolâ LHL. v/tr. carve (e.g. a picture or symbol), write (BIASP). doli L. v/tr. plant (e.g. pineapples, sago) (BIASP). dou HL. v/ambitr. close (door or pot) (BIASP). dowâ' LHL. v/tr. pull out a handful of objects (e.g. the prongs of a geim-arrow) (PFV, MONOASP). dowôn' LHL. v/tr. eat, drink (PFV). L. v/intr. nod off, doze (IPFV, MONOASP). dumun

Е

L. v/tr. hit, kill (IPFV). -e -êb LHL. v/tr. take (in order to carry) (PFV, MONOASP). ei L. v/intr. fly (IPFV, MONOASP). êi LHL. v/ambitr. accumulate (water); impound (water) (BIASP). ein L. v/ambitr. be cooked, burn (BIASP). eintunu L. v/tr. heat up stones for leaf oven (IPFV). eitana L. v/tr. heat up stones for leaf oven (PFV). en- L. v/tr. hurt, pain (IPFV, MONOASP). enâ' HL. v/intr. do thus (as shown) (BIASP).

F

fa L. v/tr. lay (egg) (BIASP).

- fa L. v/tr. make fire (PFV).
- -fâ LHL. v/tr. put, put asleep, look after, give birth (PFV).
- faa L. v/tr. make body paint (BIASP).
- -fâa LHL. v/tr. lift, raise (PFV, MONOASP).
- faka L. v/tr. make fire (IPFV).
- fibâ LHL. v/ambitr. tremble, shake (BIASP).
- fofola L. v/intr. wither, become parched (by the sun) (PFV, MONOASP).
- L. v/tr. paint the body (IPFV). fofou
- L. v/tr. paint the body (PFV) fofoula
- fote-LHL. v/tr. expel, rout (BIASP).
- fu L. v/tr. smoke; cook (BIASP).
- -fu- L. v/ditr. send (PFV, MONOASP)
- -fû' L. v/tr. grab, grip (PFV, MONOASP).
- fua L. v/intr. wash body (IPFV).
- fubâ LHL. v/tr. wash (hands, body, clothes) (PFV).
- LHL. v/tr. wash (hands, body, clothes) (IPFV). fubâ(ka)
- fuela L. v/intr. wash body (PFV).
- L. v/intr. blow into the fire (IPFV, MONOASP). fufun
- fumentlaa L. v/intr. ponder, brood (PFV, MONOASP).
- fun L. v/intr. think (IPFV)
- funa L. v/intr. think (PFV).
- G
- ga L. v/func. say (IPFV).
- ga L. v/tr. cook in a leaf oven (BIASP).
- gâala LHL. v/tr. tear down, destroy (house or fence) (BIASP).

L. v/tr. pass, surpass, bypass (PFV, MONOASP). gaige L. v/func. say (PFV). ge L. v/tr. build, fasten, roll; fight (PFV). gekâ LHL. v/intr. line up (PFV, MONOASP). gelà ' HL. v/tr. build a ladder to reach a house on stilts (PFV, MONOASP). gen L. v/intr. be sick (IPFV, MONOASP). gen L. v/tr. build, fasten, roll; fight (IPFV). LHL. v/tr. scratch (BIASP). gengâ HL. v/tr. tie again and again (IPFV, MONOASP). gengkà getei-LHL. v/tr. lack, miss (PFV, MONOASP). gi L. v/intr. laugh. gî LHL. v/tr. tie, hang up (BIASP). gibâ LHL. v/tr. bring up, rear (BIASP). gibba L. v/intr. get wet (PFV). gibbu ?L. v/intr. get wet (IPFV). gila L. v/intr. laugh (PFV) glitâ LHL. v/tr. wipe off (BIASP). glukowâa' LHL. v/intr. slacken, come loose (PFV, MONOASP). go- L. v/tr. like (PFV, MONOASP). gò' HL. v/tr. cut skin or flesh (PFV). gobtou L. v/?tr. pull together (i.e. pull hands and feet towards the body); do chin-ups; huddle up for sleep (PFV, MONOASP). L. v/intr. shrivel up (PFV, MONOASP). gogola goholo L. v/ambitr. coil up (PFV, MONOASP). goi-L. v/tr. smash (PFV, MONOASP). gokà HL. v/tr. cut skin or flesh (PFV). L. v/tr. put handle of string bag around forehead so that the bag hangs down the back goki (PFV, MONOASP). gokilêb LHL. v/tr. put on head (in order to carry) (PFV, MONOASP). LHL. v/ambitr. sear skin (PFV). golâ LHL. v/intr. burn (like ginger) (PFV, MONOASP). golâ golâ LHL. v/tr. cut and clear (BIASP). golâ(ka) LHL. v/ambitr. sear skin (IPFV). gububma ?L. v/intr. collide (PFV) gububsaan ?L. v/intr. collide (IPFV). gungglù' HL. v/tr. knot together, tie together (BIASP). LHL. v/?tr. bend knee, kneel (PFV, MONOASP). gwagwelô' LHL. *v/tr*. cut out bowels (PFV, MONOASP).

gwi L. v/tr. use black magic (BIASP).

Н

- hà' HL. v/ambitr. break, dig (PFV).
- hàa HL. v/tr. catch (fish) (BIASP).
- haa L. v/tr. weave (BIASP).
- hâa' LHL. v/intr. walk around, wander, roam (IPFV, MONOASP).
- -hâa' LHL. v/tr. chase (IPFV, MONOASP).
- hai- L. v/tr. cut off (a long protruding object) (PFV, MONOASP).
- haka L. v/ambitr. break, dig (IPFV).
- hake L. v/tr. break through (PFV, MONOASP).
- halâ LHL. v/ambitr. break (PFV, MONOASP).
- *halà'* HL. *v/tr*. abstain, prohibit (BIASP).
- halbì HL. v/tr. weed (BIASP).
- halbû' LHL. v/tr. fold (IPFV).

halbua ?L. v/tr. fold (PFV). L. v/tr. worry about (PFV). (-)halila (-)halin L. v/tr. worry about (IPFV). halò LHL. v/ambitr. cut, break, help a friend in a fight (PFV, MONOASP). L. v/intr. become old (of things) (PFV, MONOASP). hamila hana L. v/intr. get up, rise (PFV, MONOASP). LHL. v/tr. pull towards oneself (PFV, MONOASP). hatelâ' LHL. v/intr. lean (PFV, MONOASP). hebâ heitda L. v/intr. shake hands (BIASP). helâLHL. v/ambitr. break, traverse (PFV, MONOASP). helò HL. v/ambitr. break (PL subject/object) (PFV, MONOASP). hen L. v/tr. seek (IPFV). L. v/tr. seek (PFV). hena hetanâ LHL. v/tr. meet (PFV). LHL. v/tr. meet (IPFV) hetunû L. v/intr. feel bad, suffer (PFV, MONOASP). hota

I

iba L. v/tr. pour (PFV). *ibu* L. v/tr. pour (IPFV). *ifa* L. v/tr. scrape ash off baked taro (IPFV). *ifa* L. v/tr. serve food (PFV). *ifela* L. v/tr. scrape ash off baked taro (PFV). *ifu* L. v/tr. serve food (IPFV). *inà* ' HL. v/intr. do thus (BIASP). *isa* L. v/tr. string bow (PFV).

K

-ka-L. v/ditr. give (IPFV). -ka L. v/tr. put (IPFV). kaan L. v/intr. die (PFV, MONOASP). kakibi L. v/tr. join together (BIASP). L. v/tr. remove fire before making a leaf oven (BIASP). kamaa ke L. v/func. do (BIASP) kè HL. v/tr. cut taro dough (BIASP). L. v/tr. go towards midday (of the sun) (PFV, MONOASP). kela ?L. v/tr. extinguish (PFV, MONOASP). kemela ki L. v/tr. align, read, measure, point (BIASP). L. v/intr. stir up trouble (BIASP). kika kikekâ' LHL. v/tr. rub (BIASP). kiki L. v/tr. share (BIASP).

- kilo L. v/intr. begin (PFV, MONOASP).
- *-kimà* HL. *v/tr*. put in the fire (PFV).
- *kimâa*' LHL. *v/tr*. care for, watch out for (PFV, MONOASP).
- *kimi(n)* L. *v/intr*. go out (of fire, lamp, tobacco) (PFV, MONOASP).
- -kimsan L. v/tr. put in the fire (IPFV).
- klâ LHL. v/tr. make, work, build; fix complete (BIASP).
- klaa L. v/intr. rot, decay (PFV).
- klaan L. v/intr. rot, decay (IPFV).
- klafâ LHL. v/tr. put on back (in order to carry); climb (PFV, MONOASP).
- klen L. v/intr. rustle (of leaves, etc.) (IPFV, MONOASP).
- klolâ' LHL. v/tr. bare, expose, lay open (by removing skin or bark) (PFV, MONOASP).

klutâLHL. v/tr. break, shatter (PFV).klutâ(ka)LHL. v/tr. break, shatter (IPFV).komôuLHL. v/tr. sit on shoulder (BIASP).kou L. v/tr. have sexual intercourse, copulate (BIASP).kubu?L. v/tr. smoke (food), cure (food) (BIASP).kun L. v/intr. emanate smell (BIASP).kweko?L. v/tr. mix (BIASP).

L

-lò HL. v/tr. hit, kill (PFV).

Μ

-ma L. v/tr. plant, grow (bananas, taro, sugarcane) (PFV). L. v/tr. cut (meat) (BIASP). таа LHL. v/intr. stand up (PFV, MONOASP). mâa maanafa L. v/tr. cut (meat) (PFV, MONOASP). maanafu L. v/tr. lacerate (PFV, MONOASP). L. v/intr. swagger (PFV, MONOASP). тата mamlêya LHL. v/intr. turn around on the spot (while hopping/dancing) (PFV, MONOASP). mangglom L. v/intr. wail, cry (only PL subject) (PFV, MONOASP). me L. v/intr. cry (IPFV). LHL. v/intr. fall, drop out (PFV, MONOASP). -mêin L. v/tr. hang up (PFV, MONOASP). -meki mela L. v/intr. cry (PFV). L. v/tr. touch (PFV). melemelekala L. v/intr. work hard, drudge (PFV, MONOASP). men- L. v.tr. touch (IPFV). mengâ LHL. v/tr. pull taut, force (PFV, MONOASP). mî' LHL. v/tr. meet, gather (PFV, MONOASP). miba ?L. v/tr. close by putting down lid (BIASP). HL. v/tr. hold child in arm, brood, give birth (BIASP). -mikì miki L. v/tr. put in mouth (BIASP). mo L. v/intr. come and go, pass by (PFV, MONOASP). HL. v/tr. like (to be with a person) (PFV, MONOASP). mokôb' mokonga L. v/intr. become emaciated (PFV, MONOASP) HL. v/intr. become ripe (of banana) (PFV, MONOASP). molà molâ LHL. v/intr. become full (of moon) (PFV, MONOASP). HL. v/tr. break banana (off stalk) (PFV, MONOASP). molò motomâa LHL. v/tr. ascertain, confirm (PFV, MONOASP). LHL. v/tr. put child or pig over shoulders (in order to carry) (PFV, MONOASP). -môu moukowâ' LHL. v/intr. noise a pig makes when running away (PFV, MONOASP).

N

n L. v/intr. exist, stay, remain (PFV).

na L. v/tr. do, make (BIASP).

-nâ' LHL. v/tr. hit, kill (PFV).

- *nema* L. *v/tr*. peel (taro, sweet potato) (IPFV).
- nantana L. v/tr. lick (PFV).
- nantunu L. v/tr. lick (IPFV).
- nemelâ' LHL. v/tr. peel (taro, sweet potato) (PFV).
- ngaan L. v/tr. sing, call out (IPFV).
- ngaana L. v/tr. sing, call out (PFV).
- nganà HL. v/tr. spread out (leaves, feathers, blanket, tree bark) (PFV).

L. v/tr. beg, ask for persistently (PFV). ngela ngen L. v/tr. beg, ask for persistently (IPFV). L. v/tr. bark, howl (of dogs) (BIASP). ngoun HL. v/tr. spread out (leaves, feathers, blanket, tree bark) (IPFV). ngunù L. v/tr. scrape taro (BIASP). nini LHL. v/tr. bite (PFV, MONOASP). -ntamâ '

0

-ò HL. v/tr. take, pick up (PFV, MONOASP). 0 L. v/tr. say, talk, tell, call (IPFV). obâ LHL. v/tr. play (ball) (BIASP). obdî LHL. v/tr. fetch water (BIASP). obtanà HL. v/tr. put fire to, light (PFV). obtunu L. v/tr. put fire to, light (IPFV). oli L. v/intr. almost stop (of rain only) (BIASP). LHL. v/tr. watch for, be on the lookout for (BIASP). olibâ omflebâ LHL. v/tr. miss (with missile), come close (PFV, MONOASP). on L. v/intr. go (PFV). ou L. v/tr. assemble arrow by putting the head into the shaft (BIASP).

S

L. v/intr. sleep (PFV, MONOASP). S L. v/tr. shoot (PFV, MONOASP). saan (-)san L. v/tr. plant, grow (bananas, taro, sugarcane, Hong Kong taro) (IPFV). L. v/intr. moan (IPFV, MONOASP). sasan LHL. v/tr. strengthen, support (PFV, MONOASP). sbalmâ ?L. v/intr. be happy, rejoice (PFV). seila sein L. v/intr. be happy, rejoice (IPFV). sengela ?L. v/tr. pour (PL object) (BIASP) LHL. v/tr. remove the bark of Gnetum gnemon (to make thread for string bags) (PFV, seselô MONOASP). sibbe L. v/intr. fill up (PFV, MONOASP). silâ LHL. v/tr. scrape dirt or taro corm, rub the skin with stinging nettle (BIASP). -silêb LHL. v/tr. set out after, follow directly (PFV, MONOASP). simaan L. v/intr. be upset, be ashamed (IPFV). simaana L. v/intr. be upset, be ashamed (PFV). L. v/tr. pour (SG object) (BIASP). singa singgila ?L. v/tr. shake (liquid) (BIASP). sità HL. v/tr. insist, keep asking (with wéng 'talk' as object) (BIASP). sita L. v/tr. watch over; guard (BIASP). sitâ' LHL. v/tr. try to loosen, jiggle (unsuccessfully), keep asking (BIASP). situbû LHL. v/tr. knead (taro dough) (BIASP). -ski ?L. v/tr. turn (BIASP). LHL. v/tr. push (PFV, MONOASP). slelêb -suan L. v/tr. be angry with, hate (IPFV). L. v/tr. be angry with, hate (PFV). -suana

Т

- ta L. v/tr. deny (BIASP).
- ta L. v/tr. sharpen (BIASP).
- tà' HL. v/ambitr. cut between, tear.
- taa L. v/intr. spit (BIASP).
- taa L. v/tr. open (the mouth) (BIASP).

taalwaa L. v/intr. worry (IPFV, MONOASP). tai- L. v/tr. cut off (SG object) (PFV, MONOASP). taka-L. v/tr. avoid (PFV, MONOASP). L. v/tr. cut off (IPFV). taka LHL. v/tr. make a snare, set a trap (PFV, MONOASP). takakâ talò HL. v/tr. cut off (SG object) (PFV, MONOASP). HL. v/tr. lock up, pen in, imprison (PFV, MONOASP). -tamà -tamâa LHL. v/tr. step on (PFV, MONOASP). L. v/tr. comb (PFV). tana -tanà HL. v/tr. light (of fires only) (PFV, MONOASP). -tangâa' LHL. v/tr. hang up piece of clothing (to dry) (PFV, MONOASP). te L. v/intr. come (IPFV) L. v/intr. come (PFV). te~tl tefù' HL. v/?intr. warm oneself by the fire. (BIASP). tei- L. v/tr. cut off (PL object) (PFV, MONOASP). LHL. v/tr. stretch (the body) (PFV, MONOASP). tekêi' telâ LHL. v/tr. cut between (PFV, MONOASP). telò HL. v/tr. cut off (PL object) (PFV, MONOASP). têm' LHL. v/intr. have a look (PFV). -têm' LHL. v/tr. see (IPFV). temdei-L. v/tr. leave (PFV, MONOASP). temê ' LHL. v/intr. look (IPFV). LHL. v/tr. look at (IPFV). -temê -tên ' LHL. v/tr. see (PFV). tenà ' HL. v/tr. make string bag (BIASP). L. v/tr. crack (nuts), knead (dough) (BIASP). teya ?L. v/tr. hide (BIASP). ti tibila ?L. v/tr. cover (BIASP). tibtà ' HL. v/tr. stack (BIASP). tikà' HL. v/intr. wriggle (BIASP). tila L. v/intr. flash (of lightning) (BIASP). tila L. v/tr. remove, loosen, undo (BIASP). L. v/intr. be in one's final throes (IPFV, MONOASP). tilen tlaa(n) L. v/intr. come (PFV). tlâa' LHL. v/intr. be(come) sad (PFV, MONOASP). -tlâa' LHL. v/tr. remove (PFV, MONOASP). tlamàn' HL. v/intr. become silent (PFV, MONOASP). tlanhaa L. v/intr. play (BIASP) tle L. v/intr. come (iterative) (IPFV). tli L. v/tr. chew (BIASP). tobtlin L. v/intr. be confused (IPFV). tobtlina L. v/intr. be confused (PFV). tolâ LHL. v/tr. peel off skin (of banana) (BIASP). L. v/tr. be afraid of (IPFV). tosian tosiana L. v/tr. be afraid of (PFV) HL. v/tr. put over fireplace (PFV, MONOASP). -tòu tou HL. v/tr. set down (BIASP). tou L. v/intr. sit down (PFV, MONOASP). toufa ?L. v/tr. put food on leaves (BIASP). L. v/intr. sit (BIASP). toula LHL. v/tr. put on arm (in order to carry) (PFV, MONOASP). -toulêb L. v/intr. sit down (PFV, MONOASP). toun

tubu L. *v/tr*. pour, sprinkle (BIASP).

tubû' L. v/tr. shine (of the sun) (BIASP).

- LHL. v/tr. grab (a person) (PFV, MONOASP). tubunâ
- tulâ LHL. v/tr. tap blossom for nectar (of birds and insects), poke a stick in a hole in order to bring forth an animal (BIASP).
- L. v/tr. comb (IPFV). tunu

U

- L. v/tr. cut (wood) (BIASP) u LHL. v/ditr. give (PFV) -ûb-
- ulà ' HL. v/intr. stop, cease (PFV, MONOASP).
- ulâ' LHL. v/tr. burn (PFV, MONOASP).
- ulâa LHL. v/tr. open (PFV, MONOASP)
- HL. v/tr. pull out (taro corms, PL) (PFV). ulelò
- uli L. v/tr. make a thread or rope by rolling fibre (IPFV).
- HL. v/tr. make a thread or rope by rolling fibre (IPFV). ulilò
- un L. v/intr. hum, roar (of planes and engines) (IPFV, MONOASP).
- un L. v/intr. go (PFV).
- unaa(n) L. v/intr. go (SS, PFV).
- unê LHL. v/intr. go (IPFV).
- L. v/tr. place (a child) into the arm (so that it can sleep) (PFV, MONOASP). -usa

w

wà HL. v/tr. pick (fruit), cut (PFV). L. v/tr. hide (PFV, MONOASP). waa L. v/tr. make carvings in arrows (BIASP). waa L. v/tr. swim (lit. push (water)) (BIASP). waa LHL. v/intr. fly (of insects and helicopters) (IPFV, MONOASP). waalâ wabolàa HL. v/tr. grab many (PFV, MONOASP). L. v/tr. close (PFV, MONOASP). wai-L. v/tr. cut off (SG object) (PFV, MONOASP). waiwai-L. v/tr. wait (PFV, MONOASP) waka L. v/ambitr. pick (fruit), cut (IPFV). waketòu HL. v/tr. cut short (BIASP). HL. v/intr. stop, cease (PFV, MONOASP). walà walbì HL. v/tr. clear (bush) (BIASP) HL. v/ambitr. break off, cut off (SG object) (PFV, MONOASP). walò LHL. v/intr. become blocked, get stuck (PFV, MONOASP). wambiâ' wamflaa L. v/intr. fly around, fly in circles; chase around (BIASP). LHL. v/intr. take someone's place (PFV, MONOASP). wasêi HL. v/tr. prevent, prohibit (BIASP). watà watwatda L. v/ambitr. break, damage, destroy (PFV, MONOASP). we L. v/tr. sweep (BIASP). webiâ LHL. v/intr. recover (PFV, MONOASP). L. v/tr. cut off (PL objects) (PFV, MONOASP). weiwei-L. v/tr. miss (PFV, MONOASP). welâ LHL. v/tr. cut (PL object) (PFV, MONOASP) LHL. v/tr. cut off (PL object) (PFV, MONOASP). welò L. v/tr. eat (IPFV). wen went L. v/tr. hear, listen, understand (PFV). LHL. v/tr. hear, listen to, understand (IPFV). wentê

wi L. v/tr. cut (wood) (BIASP).

Y ya L. v/tr. cross body of water (BIASP). yaalâa' LHL. v/intr. wriggle, writhe (BIASP). yangke L. v/tr. answer, pay back, take revenge (BIASP). yo L. v/tr. give birth, create, initiate (BIASP). yolyomaa L. v/intr. jump (PFV, MONOASP).

Y

MEASURING MORPHOLOGY: THE TIP OF THE ICEBERG? A RETROSPECTIVE ON 10 YEARS OF MORPHOLOGICAL PROCESSING

Hélène Giraudo & Madeleine Voga (1) CLLE-ERSS (CNRS) & Université de Toulouse, France (2) Université Paul Valéry, Montpellier, France

Abstract

Despite the intensive study of morphological effects with various on-line techniques such as masked priming, psycholinguistics did not manage so far to present a consensual framework, and are still divided on the nature and the locus of morphology in the mental lexicon. In this contribution, we propose to focus on three issues related to morphological effects which have not been given the right importance so far: the implications of studying morphology through nonwords, the role of frequency of the lexical items used as materials, and finally the role of a novel variable measuring the influence of formally related but morphologically unrelated word forms on processing, *i.e.* the pseudo-relatives. The experiment presented here provides evidence in favour of these two variables. We propose a revised model of morphological processing, sensitive to lexical (*e.g.* frequency) and exo-lexical characteristics of the stimuli (*e.g.* pseudofamily size), capable to cope with various effects induced by true morphological relatives and pseudorelatives, as well as for surface effects, such as the pseudoderivation effect.

Over the last 40 years, multiple studies have addressed the issue of morphological processing during word recognition, trying to establish how morphologically complex words are analysed and coded in long-term memory. Until 2000, and while morphological effects have been reported in various languages (mostly English, but also Hebrew, Russian, French, Italian, Spanish or Serbo-Croatian), using different paradigms (mainly priming paradigms) and tasks (lexical decision and naming), psycholinguists were divided: on one hand, tenants of the decompositional approach (*e.g.* Taft & Forster, 1975), based on an affix stripping mechanism intervening during the first stages of lexical access and which can be assimilated to the morphemebased theory of morphology propounded by linguists (*e.g.*, Halle & Marantz, 1993); On the other hand, those who privileged a whole-word-access comparable to the word-based approach (see Chap. 3 in Haspelmath & Sims, 2010). During this period, experimental studies focused on factors
determining complex word recognition and influencing processing of their surface as well as their internal structure. Among these factors, the effects of surface and base frequencies, defining the statistical occurrence of complex words, were extensively studied in languages for which lexical databases were available (*e.g.*, Baayen, Dijkstra, & Schreuder, 1997 for Dutch ; Burani, Salmoso & Caramazza, 1984 for Italian ; Colé, Beauvillain,& Segui, 1989 for French ; Ford, Davis, & Marslen-Wilson, 2010; Taft, 1979; 2004 for English). The fact that recognition latencies depended on both surface and base frequencies was taken as evidence that the reader was sensitive to morphological structure and that a component of morphological processing is related to perceptual sensitivity, suggesting that lexical access strongly depends on whole-word as well as morphemic information.

Priming and masked priming studies went further in examining the role and the representation of morphology within long-term memory. These paradigms are specifically designed to explore the nature of activation transfers from a prime stimulus on target recognition (Forster, 1999) at conscious (long-term priming) and non-conscious (masked priming) levels of processing. The masked priming technique allows for the manipulation of various kinds of relationships between two words, thus rendering it possible to determine the positive or negative effect of a shared linguistic characteristic (phonology, orthography, morphology and semantics). In the case of morphologically related words, it enables the researcher to tease apart the respective part of form and meaning in morphological priming. From the seminal repetition priming study conducted by Stanners et al. (1979) to the most recent investigations combining masked priming to brain activity (e.g., Morris, Grainger, & Holcomb, 2013) morphological priming effects have been extensively studied and have systematically revealed strong facilitation effects. Experimental results exhibiting morphological effect (facilitation) differing significantly from formal and meaning relationships, conduced the authors to conclude that independent morphological representations were coded somewhere within the mental lexicon in a similar way as orthographic, phonological and semantic representations.

Taken together, experimental results suggesting various frequency effects on one hand and demonstrating autonomous morphological effects, independent from semantic and orthophonological relationships on the other, led to the three following options described in literature relative to morphological representation and processing: a) the purely sublexical option, in which morphemes stand as access units, implying an obligatory decomposition mechanism that systematically splits off the affix from its base (Taft, 1994);

b) the intermediate sublexical option, postulating a morphemic access route acting in parallel with a whole-word access route (e.g., Caramazza, Laudanna & Romani, 1988) and c) the supralexical option, positing abstract morphemic

units at the interface of word-form and meaning representations and organizing word forms in terms of morphological families (Giraudo & Grainger, 2001).

The period from the year 2000 to 2005 was marked by studies focusing specifically on the decomposition vs. nondecomposition issue in order to determine the locus of morphological effects. Even if the priming study carried out by Rastle and colleagues in 2000 historically defines the starting point of a series of masked priming studies, the most striking ones were conducted respectively in French by Longtin, Segui, & Hallé (2003) and in English by Rastle and New (2004). Both manipulated a particular type of word pairs, presenting morphological complexity at their surface form, but which are neither synchronically nor diachronically related (e.g., the English word corner cannot be analyzed in corn + er). Using the masked priming paradigm, it was shown that pseudo-derived word primes (e.g., corner) as well as pseudo-derived nonword primes (e.g., corning) composed of two existing morphemes were able to produce significant priming effects on the recognition times of their base (e.g., corn). Moreover, it appears that the quality as well as the magnitude of these priming effects, is comparable to the priming effects produced by genuinely derived words (e.g., banker-bank). Finally, the systematic use of orthographic control primes (*i.e.*, morphologically simple forms for which the first part alone mimics a stem morpheme, such as *brothel* in which -el never functions as a suffix in English) in these studies showed that these surface morphological effects could not be assimilated to mere formal overlap. Consequently, these effects would exclusively result from the surface morphological structure of the primes.

Longtin and Meunier (2005) then explored the "pseudoderivation effect" using pseudowords in order to test the resistance of early morphological decomposition following manipulation of the lexicality of the primes. In their masked priming study, morphologically complex pseudowords (non existing possible words created with two existing morphemes, for instance, the base sport- + the suffix -ation produce sport-ation) were used as primes. The data revealed that pseudo-derived pseudowords (*i.e.*, sportation) facilitated the recognition latencies of their base (e.g., sport) and did not differ from the facilitation effects obtained using transparent primes (e.g., sportif 'sports' which is a legal and semantically transparent derivation of the base sport). Following the same logic, McCormick, Rastle and Davis (2008) manipulated another category of derived stimuli that cannot be segmented perfectly into their morphemic components (e.g., dropper-drop in which there's a duplicated consonant) in order to test the flexibility of the morphoorthographic segmentation process described by morpheme-based models. Once again their results demonstrate the robustness of this segmentation

process in the case of various orthographic alterations in semantically related (*e.g.*, *adorable-adore*) as well as in unrelated prime-target pairs (*e.g.*, *fetish-fete*).

Taken together these data strongly support the robustness of a morphological decomposition effect across languages, stimuli and sensorial modalities. A complete review of the literature related to this question was made by Rastle and Davis (2008) and perfectly summarized the results in claiming: "morphological decomposition is a process that is applied to all morphologically structured stimuli, irrespective of their lexical, semantic or syntactic characteristics" (p. 949). This conclusion seemed to deliver the *coup de grace* to any approach (the supralexical model in particular) that would postulate intermediate lexematic units situated above word units.

It should be noted at this point that, as Giraudo & Voga (2013) notice, a more recent study conducted by Crepaldi, Rastle, Coltheart, & Nickels (2010) opened a breach in this wall of certainty. A series of masked priming experiments were carried out on English irregularly inflected forms (viz. allomorphs). Interestingly enough and in total contradiction to their starting hypothesis, the authors found that allomorphs (e.g., fell) for which the decomposition of the surface form is not relevant for stem recovering, primed their verbal base (e.g., fall) more than orthographically matched (e.g., fill) and unrelated control words (e.g., hope) did. This result had already been found by Pastizzo & Feldman (2002), and discussed enough by morphologists, but it had not been attributed the right importance by the tenants of the sublexical approach because of minor pitfalls in the control conditions (which did not have any incidence on the results, as the results of Crepaldi et al. demonstrate). Crepaldi et al. thus conceded the "existence of a second higherlevel source of masked morphological priming" and proposed a lemma-level composed of inflected words acting "at an interface between the orthographic lexicon and the semantic system" (p. 949).

This breach gave rise to numerous experimental studies whose aim was henceforth to reduce the gap between 10 years of research exclusively focused on finding morphological decomposition everywhere and the necessity to consider (or re-consider) data spotlighting that morphology cannot be reduced to the syntax of words. For instance, research on the impact of letter transpositions that arises at a morpheme boundary (*e.g.*, *boasetr* for *boaster*) has yielded conflicting results that are still in debate. Cross-linguistic differences, task limitations or particular characteristics of the materials used might explain why some authors have found an impact of letter transpositions (Christianson, Johnson, & Rayner, 2007; Dunabeitia, Perea, & Carreiras, 2007; Sanchez-Gutierrez & Rastle, 2013) while others have not (Diependaele, Morris, Serota, & Grainger, 2013; Rueckl & Rimzhim, 2011). Nevertheless it is acknowledged that «there is more than one way in which morphology can influence visual word recognition» (Diependaele et al., 2013, p. 1001) implying that decomposition is not the beall and end-all of morphological processing.

In this contribution, we propose to focus on three, rather neglected but closely related areas of interest in the literature relative to morphological processing: the first one relates to frequency effects, whose study is concomitant with the beginning of psycholinguistic research on lexical access (ex. the serial model, Forster, 1976) and led to a very important number of published studies. Nevertheless, when it comes to morphological processing, frequency effects do not seem to be attributed the role they should. This overlooks the fact that inflected and derived words also exist as free word forms, and not only as analysable units. The corollary of (base and surface) frequency is the residual activation of units, characterising all lexical items, morphologically simple or complex. In experiment 1, we will illustrate the residual activation and some of its implications, through the manipulation of relative frequencies between primes and targets under masked priming conditions. Residual activation is undoubtedly central to interactive activation (McClelland & Rumelhart, 1981), as well as serial frameworks (Forster, 1976), yet the examination of the relevant literature reveals that this factor remains under-exploited in morphological priming protocols.

The second one relates with the fact that morphological processing and representation is not only studied through existing lexical units, as for example in the studies focusing on frequency effects, but also, and in significant proportions, through nonwords and pseudowords of various kinds, as we saw in the introduction (*e.g.* the abundant literature on pseudo-derivation effects). The reason for studying nonwords in most of the cases resides precisely on their non-lexicality, since nonwords are not supposed to have lexical representation(s) or belong to the cluster of a real word. Nevertheless, taking for granted this idea and using it as a foundation when interpreting nonword effects in morphological terms in order to validate or refute morphological accounts of processing real items, may be problematic from a linguistic point of view, both theoretical and experimental.

The third issue which will be addressed here is related to a larger question we can summarize in the following terms: within the general framework of the lexeme approach (Aronoff, 1994) where morphology is not the "syntax of morphemes" but the extension of patterns of existing systematic formmeaning correspondences (Bybee, 1988; 2001, Booij, 2002) research has great interest on focusing on variables and effects coming from the environment of the word-to-be-identified, and not exclusively from its internal characteristics. Substantial evidence in that sense comes from morphological family size effects in a variety of languages, Germanic or Semitic (Dutch: Schreuder & Baayen, 1997; Bertram, Baayen & Schreuder,

2000; English: De Jong, Feldman, Schreuder, Pastizzo, Baayen, 2002; Hebrew: Moscoso Del Prado Martin, Deutch, Frost, Schreuder, De Jong, Schreuder & Baayen, 2005) and reflects the amount of words that will work as "synagonists" during the recognition process. This functioning may work the other way round: this would mean that formally related but morphologically unrelated words act as antagonists, thus inhibiting morphological processing of the word-to-be-identified. The first experiment we present here provides evidence in favor of a novel variable based on exactly this antagonism, taking place inside the word-level but outside the word itself. We coined this variable exo-lexical (c.f. the exolexical workshop we organized in the IMM15) in order to emphasize the fact that the locus of these effects resides outside the word under study and its morphemes, exactly as for the morphological family size variable.

1. Lexical frequency and the role of residual activation.

The masked priming technique implies a prime (ex. taught) and a target (teach), the priming benefit being the difference in the time needed to identify the target, compared to an unrelated condition. Given that the most frequently used task for this kind of protocol is the lexical decision task (yes/no) and that above a certain percentage of errors (in most cases 15-20%, words and nonwords together) the performance of a particular subject is not acceptable (since it would mean that he/she did not really processed the targets), psycholinguists have begun to habitually present as the target the most frequent item, for example the infinitive for French verbs, the present form for English verbs, or, for the experiments examining morphological effects on non-word processing (ex. Meunier & Longtin, 2007) the stem-only form, for example sport. Nevertheless, in interactive activation (McClelland & Rumelhart, 1981) as well as in serial (Forster, 1976) models, the surface frequency of the materials plays an important role, given that it defines the "resting level" or residual activation of a given lexical unit, and consequently, the amount of activation needed to reach the identification threshold. The higher the frequency of the unit, the lower its activation threshold, and consequently, less effort is needed in order to activate it. Masked priming protocols are especially subject to this kind of mechanics, given the very short lapse of time the prime disposes of to activate the target, with SOAs (Stimulus Onset Asynchronies) usually varying between 42 and 57 milliseconds for morphological effects.

What we can observe in the vast majority of masked morphological priming experiments, and this is not surprising, is that the lexical unit taken as the target is the most frequent one, or in the case of morphological effects induced by nonwords, the target is the word and not the nonword. In English protocols, the choice isn't vast: in the verbal system, supposing in a design to study past tense priming, the target can either be the 1st /SG or the 3rd/SG present tense form. In a derivational priming experiment with nonword primes, as the Meunier & Longtin (2007) experiment, the target is the noun sport, a very frequent word, and not a suffixed unit from the great morphological family of sport (ex. sportif 'sports'). If the lexical frequency of the prime and target pairs does not play any role, we should admit that all morphologically complex words are decomposed at the entry of the system, independently of their lexical status (word or nonword) and their activation status (frequent or not frequent). This is indeed the option that the greatest part of the psycholinguistic literature took for some years, as discussed in the introduction and as illustrated by studies as Rastle and Davis (2008) or, more recently, Amenta & Crepaldi (2012, p. 3). Of course things are not so simple, and data suggesting the opposite direction do exist: among the first to be highlighted was Giraudo & Grainger (2000), with French materials, reporting larger effects with high-frequency derived primes than with low-frequency ones, in the same line as Meunier & Segui (1999, with spoken primes).

Our aim here is not to review the data against the decompositional approach, but to insist on a rather surprising lack in the literature: despite the impressive amount of published morphological priming studies, very few of them attempt to reverse the prime and target pair: instead of having the most frequent member of the paradigm (or the morphological family) as target and another less frequent one as the prime, reversing the prime-target pair consists in having the less frequent member as target and the more frequent one as prime. One of the very rare studies operating this inversion is Voga & Giraudo (2009) that we will present here in some detail: as we'll see, manipulating the activation threshold modifies the pattern of effects for real words (verbs).

2. Nonword effects in processing morphology or the distinction between real word and possible word.

Recently, in a paper reviewing the most robust and well-documented morphological priming effects (masked and unmasked), such as the frequency effect as well as the morphological effect on non-word processing, Amenta & Crepaldi (2012) reach the conclusion that "Surely, morphological effects in non-words exclude the possibility that morphological information only comes into play after lexical identification" (p. 9), given that "it is clear that nonwords with a morphological structure are analyzed in terms of their morphemes, thus questioning seriously any theory that suggests morphological processing to kick off upon lexical identification" (p. 7). The experimental effects at stake here concern two types of protocols: a) In

simple (unmasked) lexical decision tasks, where the subject has to decide if the stimulus is a word or not, slower rejection times have been observed for pseudo-inflected nonwords with a real suffix compared to pseudo-inflected words with a real stem and a non-suffix or a non-stem and existing suffix (Burani, Dovetto, Thorton, & Laudanna, 1997; Burani & Thorton, 2003; Caramazza *et al.* 1988; Taft & Forster, 1975); b) In masked priming experiments, where the pseudo-inflected real-stem, real-suffix nonword, jumbled words by letter transpositions facilitates the identification of the target (the stem itself) more than other categories of nonwords (Beyersmann, Dunabeitia, Carreiras, Coltheart, & Rastle, 2013; Christianson, Johnson, & Rayner, 2005; Diependaele *et al.*, 2013; Dunatbeitia, Perea, & Carreiras, 2007; 2008; Longtin & Meunier, 2007, Rueckl & Rimzhim, 2011;Sanchez-Gutierrez & Rastle, 2013).

The underlying principle here is that nonwords used in the experiments, including very word-like nonwords, cannot have a lexical representation at all, since they do not exist as words, neither before nor after the experiment, and thus the effect they induce on facilitating the target cannot be due to their lexical representation at the word level, nor to their frequency, since they don't have any (given the non-existence of the lexical unit). Two remarks seem important to us at this point: first, as Amenta & Crepaldi (2012) acknowledge, nonword morphological priming effects are characterized by great inconsistency; to cite an example, Burani et al. (2002) obtain no difference between rejection times on suffixed nonwords (e.g., donnista 'womanist') and rejection times on orthographically control nonwords that did not contain any morpheme (e.g., dennosto similar to "wemanost" in English); This inconsistency is not found for true morphological effects, we have thus a qualitative difference between the pattern for nonwords and for words. Second, the masked priming technique is nevertheless sensitive to orthographic similarity and this is precisely the reason why an orthographic control is used, at least in studies of type (b), for exemple in Pastizzo & Feldman (2002) or in Giraudo & Grainger (2001). Consequently, one can wonder whether it is acceptable to suppose that the orthographic control created by a real stem and a non-suffix (or a non-stem and an existing suffix) as in Caramazza et al. (1988) in order to match a suffixed nonword (i.e. with a real stem and a real affix, as for exemple *cantevi*, in the Caramazza study, similar to the English *buyed*) is equivalent to a real-word orthographic control such as those used in real-word morphological masked priming experiments (for exemple in Pastizzo & Feldman (2002) or in Giraudo & Grainger (2001) ? In other words, while buyed looks as a real word, the orthographic control to match it does not look as a real word. We cannot answer to what extent this methodological pitfall can change the pattern of results, or whether psycholinguists should banish the use of nonwords for their experiments.

Nonetheless, an indirect answer comes from two groups of experimental data: first, from data on interference on lexical identification, and second from data on neighborhood effects on nonword visual processing.

As far as the first type of experimental evidence is concerned, Bowers, Davis & Hanley (2005a) have shown that having participants learn new words (e.g., BANARA) that were neighbors of familiar words that previously had no neighbors (e.g., BANANA), made it more difficult to semantically categorize the familiar words. This means that interference can also be exerted by items that initially, *i.e.* at the beginning of the experiment, had no lexical status, but acquired it during the experiment. Moreover, as Bowers et al. (2005a) show, this interference was greater the day following initial exposure. In other words, within a mental lexicon dealing every day with novelty, productivity and lexical creation, thus attributing, as Bowers et al. shows, a (probably temporary) lexical status to an item as BANARA, there is no reason to exclude word-like items from the realm of real words, as far as linguistic processing is concerned. This is especially true for experiments where the subject has received the instruction to push the right button if the stimulus is a word and the left button if it isn't, i.e. where the subject has to decide entirely independently the lexicality of the stimuli presented to him/her. (see Grainger & Jacobs, 1996 and Jacobs & Grainger, 1994 for a theoretical account of many empirical findings revealed by lexical decision tasks).

As far as the second type of evidence is concerned, there is substantial work in Italian, a language with shallow orthography that lexical activation is present when processing nonwords. Arduino & Burani (2004) find a facilitatory effect of neighborhood size in naming Italian nonwords (e.g. greno, tegno, darta, Exp. 2) and an inhibitory effect of neighborhood frequency in lexical decision with the same nonwords (when neighborhood size and neighborhood frequency were orthogonally varied, Exp. 1). In the naming task of Arduino & Burani, the evidence in favor of the lexical component in reading nonwords is obvious: nonwords with many neighbors were read aloud faster than nonwords with few neighbors, irrespective of neighborhood frequency. This suggests that even in a language as Italian, where the transparency of grapheme-phoneme correspondences should render the non-lexical print-to-sound conversion the privileged path in reading novel words, reading nonwords can benefit from the activation of the lexicon. The facilitatory effect of neighborhood size on nonword naming latencies with no role for the neighbor's frequency is interpreted by the authors as evidence for lexical activation in the case of newly encountered nonlexical stimuli. Even if the results of the lexical decision are in contrast with the results of the naming task, with respect to the role of neighborhood size and neighbor frequency, it is clear that these two factors induce some kind of influence in processing nonwords. The fact that nonwords with a high frequency neighbor required more time to be rejected, interpreted by the authors inside the dual-route (Coltheart *et al.*, 1993, 2001), and the multiple read-out model (Grainger & Jacobs, 1996), results from lexical activation in the word recognition system. Independently of the particular pattern of results for lexical decision and naming in the study of Arduino & Burani (2004), the point we wish to make here is that nonwords are not identified/read independently and above (or rather below, to illustrate the activations in terms of processing architecture) any participation of the lexicon. Neighborhood size and neighbors' frequency are lexical factors, not pre-lexical.

In light of the above development, it becomes clear that the argument, according to which the sole existence of morphological effects on nonword processing refutes all approaches not based on mandatory decomposition (Amenta & Crepaldi, 2012), seems insufficient. This argument, quite present in psycholinguistic studies demonstrating or reviewing mandatory decompositional effects (Amenta & Crepaldi 2013; Rastle & Davis 2008; Meunier & Longtin 2007) is based on an extremely static view of the mental lexicon and on a very rigid dichotomy between word and non-word, as if the category of "possible word" did not exist. Assuming that a well-formed, morphologically pseudo-derived or pseudo-inflected nonword such as the pseudo-derived sportation of the Longtin & Meunier (2005) study or the cantevi of the Caramazza et al. study, similar to the English buyed, should not induce any priming because it does not have any lexical representation, has little basis. The assumption that a nonword like *buyed* or *sportation* cannot be connected through some kind of link to the stem buy or sport on the lexical or post-lexical level reveals a disregard for several well established facts: those related to language acquisition where children produce these false yet perfectly intelligible forms (buyed, goed, etc) as well as those related to productivity (Hay & Baayen, 2003; Plag, 1999; 2004; Lopez-Villasenor, 2012). Finally, such an argument disregards robust experimental effects arguing in favor of a very thin line between units having a lexical status and intermediate type units that do not, but can acquire it, as in the Bowers et al. (2005) study.

This question may seem not so central to the issue of morphological processing, given that, according to certain logic, the way in which nonwords are behaving is maybe not as important as the way real world words are behaving. Despite this, however, a close look of the literature reveals that this kind of nonword has been extensively used to study and validate the mandatory decomposition hypothesis. To cite only two examples, with the masked priming technique, Rastle, Davis, Marslen-Wilson, & Tyler (2000) show that pseudo-derived nonword primes (*e.g., corning*) composed of two existing morphemes were able to produce significant priming effects on the recognition times of their base (*e.g., corn*). In French, the pseudo-derivation

effect, *e.g. sportation - sport* (Longtin & Meunier, 2005) was found equivalent to the effect induced by semantically transparent true derivations, *e.g. sportif – sport*.

To conclude on morphological priming with nonwords, the question that the literature has to answer is the following: should *cantevi* in Caramazza *et al.* (1988), *donnista* in Burani *et al.* (2002), or *sportation* priming *sport* in Meunier & Longtin (2007) be considered as complete nonwords, deprived of some kind of link to their (very frequent) base form? If we consider this kind of nonwords as *possible words*, linked in some way to the base lexical unit, then the argument according to which nonword effects reflect automatic decomposition loses a lot of its validity. In fact, in this case, the data can very well be interpreted in the opposite way: that the pattern of systematic formmeaning correspondences that we call morphology (Bybee, 1988; 2001, Booij, 2002) is extended to novel words.

3. Exo-lexical variables: the pseudo-family size (Voga & Giraudo, 2009).

The morphological family size variable (Bertram, Baayen & Schreuder, 2000; De Jong, Schreuder & Baayen, 2000) has been shown to influence word processing: complex words with many morphological relatives will be processed faster than those with a poor morphological family, suggesting thus that the locus of morphological effects is not exclusively the word to be processed and that factors outside the word in question intervene on morphological processing. In the case of the morphological family size variable, words from the same family act as synagonists during processing. Nevertheless, in the mental lexicon not only synagonists but also antagonists exist. The role and existence of antagonists are indicated by neuropsychological (Massol, Grainger, Dufau & Holcomb, 2010) as well as behavioral measures. For example, Grainger, Colé & Segui (1991) have found that orthographic similarity of the prime inhibits lexical access of morphologically complex targets, despite (or because of) the absence of any morphological relation between them, e.g. the prime "mûrir" (ripen) inhibits the target "MURAL" (wall) and this inhibition reaches 27ms for words that share their initial letters. This inhibition is accounted for in terms of "preactivation of lexical representations during the processing of the prime, which interferes with the processing of the target" (Grainger, Colé & Segui, 1991, p. 380).

Coltheart's N (Coltheart, Davelaar, Jonasson, & Besner, 1977), another well documented effect related to word processing, refers to the number and relative frequency of neighbors, *i.e.*, words differing by a single letter (such as BANISH and VANISH); Evidence from this type of research has not

always given consistent results, and reviewing them is beyond the scope of this paper. However we stand by the remark of Bowers, Davis & Hanley (2005b) relative to the fact that in competitive network models like Interactive Activation Models (McClelland & Rumelhart, 1981) and SOLAR models (Self-Organizing Lexical Acquisition and Recognition, Davis, 1999) the critical contrast is between words that have no neighbors ("hermits") and words that have one or more neighbors. As noticed by the same authors, (Bowers, Davis & Hanley, 2005b) it is important to have a psychologically accurate definition of what is a neighbor and considering as such only words of the same length that differ by one letter (Coltheart's N) is rather based on simplicity than on perceptual similarity.

Given that morphological processing is dependent on the characteristics of the morphological family, whose members act as synagonists, it is possible that morphological processing also depends on the number and nature of neighbors. These neighbors, when they exist, act as antagonists, leading to interference in target identification, thus delaying morphological processing. Voga & Giraudo (2009) present two experiments exploring a novel variable, coined "pseudo-family size", which is the opposite of the morphological family size. Voga & Giraudo (2009) examined inflectional priming for two kinds of stimuli: verbs coming from big pseudo-families and verbs coming from small or inexistent pseudo-families, *i.e.* what Bowers *et al.* 2005b call "hermits".

By "pseudo-family size" we mean a word as "portons" (meaning "we carry", where "port-" is the stem and "-ons" is the conjugation mark). When this word is presented to the lexical processing system as a prime, it can potentially activate (at least) all words that share its initial letters, *i.e.* the letters of the stem. In other words, portons, has numerous "pseudo-relatives" at the lexical level : portail (portal), porte (door), port (harbour), portier (porter), portion (portion), portique (porch), portrait (portrait), portière (door), portugais (portuguese), but also the actual neighbor, in the sense of BANISH-VANISH, postons (we mail). The working hypothesis in Voga & Giraudo (2009) is that all these pseudo-relatives will behave like competitors at the lexical level. On the other hand, a verb like mourir (infinitive form of the verb *die*) is almost a hermit, since the only pseudo-relative it has is the rare mouron (scarlet pimpernel), and therefore it will receive a very small amount of competition on the lexical – orthographic level. A word can belong to the pseudo-family of another word even if they don't share their stem: for example, portugais (portuguese), under our definition, is a pseudo-relative of portons because the stem of portons is a part of the superset portugais. The decision to include this type of pseudo-relative in the computation of pseudofamily size was based on previous studies emphasizing the role of the beginnings of words in lexical access (Humphreys, Evett & Quinlan, 1990; Grainger, O'Regan, Jacobs & Segui, 1992), as well as on studies on lexical co-activation (Bowers, Davis & Hanley, 2005b). Consequently, this measure of pseudo-family should not be assimilated to stem homographs, such as those of Laudanna, Badecker & Caramazza (1989), ex. colpo – colpa (blow – guilt). In short, we can say that our definition of the pseudo-family size of a lexical entry is the sum of neighbors in the classic sense (Coltheart's N) and of all words sharing their stem with that entry, even if what remains once the stem is removed is not really an affix (*e.g., porter – portugais*). Following this logic, we considered "*mourir*" as a hermit, according to our pseudo-family variable. In the experiment briefly reported below, we oppose inflectional effects obtained with words having no pseudo-family to those obtained with words coming from big pseudo-families, where the prime will activate a legion of lexical competitors.

4. The Experiment

The first experiment briefly reported here (for the complete version, see Voga & Giraudo, 2009) was designed to jointly investigate the role of the pseudofamily size as well as the influence of the relative frequencies of primes and targets. As stated in section 1, what the majority of masked priming studies report as morphological effects is the facilitation induced by a morphologically related prime on the base form target, *i.e.*, the member of the morphological family that already has the greatest residual activation because of its frequency, generally higher than that of other morphologically related forms. As can be seen in Table 1, Experiment 1a studied the classic configuration, where the target is the easiest-to-activate member of the paradigm, while experiment 1b took as targets less frequent inflections, thus reversing the typical design described above. Table 1 provides a global description of the two experiments (1a and 1b).

4.1 Method

4.1.1 Participants. 62 undergraduate students from the University of Aix-en-Provence who reported normal or corrected-to-normal vision participated in the experiment.

4.1.2 Stimuli and design. Fifty-six French words and fifty-six nonwords were used as targets. Targets were always the infinitive form of French verbs, from 4 to 9 letters long (mean: 5.6 letters) with an average frequency of 66.17 occurrences per million (New, Pallier, Ferrand, & Matos, 2001) and consisted of1) 28 verbs, 4 to 9 letters long (mean: 5.6 letters), that had large pseudo-families, and 2) 28 verbs, 4 to 7 letters long (mean: 5.75 letters) that were "morphological hermits", *i.e.* with no or an insignificant pseudo-family (a

pseudo-family consisting of marginal frequency items). These two categories of target word represent the two levels of the pseudo-family size factor (PsFam) that was estimated with the help of a French dictionary (Petit Robert) by exhaustive inspection. Each target was given four types of prime: a repetition prime, two morphologically related primes, and an unrelated prime. These primes define each one of the four experimental conditions for every type of verb, PSFam+ verbs and PsFam- verbs. The two conditions of morphologically related primes were a frequent inflection and a much less frequent one. Table 2 provides examples and prime-target orthographic overlap for each one of the 8 experimental conditions tested in Experiment 1a. 56 French nonverbs were created respecting the orthotactic constraints of the language and were matched for length with the real verbs. The nonword primes were primed in the same way as word primes. Four experimental lists were created by rotating targets across the four priming conditions using a Latin-square design, so that each target appeared only once for a given participant, but was tested in all priming conditions across participants. Participants were randomly assigned to one of the four lists.

	Targets	(Primes)	Infl.	Targets	Infl.	Infl.
	Exp. 1a	Infl. F+	F-	Exp. 1b	F+	F-
PsFam+	monter	monté	montais	montons	monté	montais
verbs	115.4	144.08	4.16	6.27	144.08	4.16
PsFam-	sentir	senti	sentiront	sentons	senti	sentiront
verbs	78.4	95	2.39	4.64	95	2.39

 Table 1. Examples of stimuli and frequencies (in occurrences per million)

 for materials used in experiments 1a and 1b: targets and

 morphologically related primes [frequent inflections (F+) and non

 frequent inflections(F-)] for the two types of verbs, large pseudo-family

size verbs (PsFam+) and small pseudo-family size verbs (PsFam).

4.1.3 Procedure and apparatus. The experiment was conducted on a PC computer using the DMDX software (Forster & Forster, 2003). Subjects were requested to make lexical decisions on the targets as quickly and as accurately as possible, by pressing the appropriate button of the gamepad.

	Primes						
Word	Targets	Rep.	Freq. Infl.	Orth. ovrl.	Non freq. infl.	Orth. ovrl.	Unrel.
PsFam+ verbs	monter (climb)	monter	monté (climbed)	3.75 lt. (64 %)	montais (I was climbing)	3.75 lt. (66 %)	perdre
PsFam- verbs	sentir (feel)	sentir	senti (felt)	4.07 lt. (69 %)	sentiront (they'll feel)	4.21 lt. (69 %)	appeler

	Primes						
Non	Targets	Rep.	Pseudo-	Orth.	Pseudo-	Orth.	Unrel.
word			infl.	ovrl.	infl.	ovrl.	
Pseudo	dainier	dainier	dainions	3.71	deniais	3.79 lt	taunnie
-verbs				(65%)		(67 %)	
Pseudo	vlâmir	vlâmir	vlâmé	3.68	vlâmmais	3.7 lt	sténon
-verbs				(67%)		(69 %)	

Table 2. Stimuli sample and degree of prime-target orthographic overlap (letters, percentage) for the repetition, the two morphologically related (frequent and non-frequent inflection) and the unrelated conditions for the two types of target (large pseudo-family size PsFam+ verbs, lowpseudo-family size PsFam- verbs) tested in Experiment 1a.

4.2 Experiment 1b

Experiment 1b was identical to experiment 1a, except that targets were not the infinitive forms of French verbs and French-like pseudoverbs, but their 1st/PL inflection. The aim of this manipulation was to modify the relative frequency between prime and target. For a language like French, where infinitive forms tend to have a higher surface frequency than conjugated forms, this means that (conjugated) targets will have a surface form frequency that is lower or equivalent to that of their inflections (see Table 1 for comparative frequencies of the materials used in Experiments 1a and 1b). 32 subjects from the same subject pool participated in this experiment.

4.3 Results. Correct response times (RTs) were averaged across participants after excluding outliers (300 > RTs > 1300ms). The results for word stimuli for experiments 1a and 1b are presented in Table 3. An ANOVA was performed on the remaining data with prime type (repetition, frequent inflection, less frequent inflection, unrelated) and verb category (large pseudo-family size, small pseudo-family size) as within-participant factor. We report only Fs by subjects, since our Latin Square design permits us to

Words	Rep.	Freq.	Non freq.	Unrel.	N	et Primin	g
	(R)	infl.	infl.	(U)		Effects	-
		(F+)	(F-)				
Exp. 1a	RT	RT	RT	RT	U– R	U - F +	U - F-
PsFam+ verbs	602	617	633	634	32*	16*	1
PsFam- verbs	593	597	624	633	40*	36*	9
Exp. 1b	RT	RT	RT	RT	U-R	U-F+	U-F-
PsFam+ verbs	638	663	629	652	14	-11	23
PsFam- verbs	594	618	622	644	50*	26*	22*
Table 3 Prototion times (PT in millissonands) for levies I desiring to							

remove all F2 analyses (Raaijmakers, Schrijnemakers and Gremmen, 1999) which would be very conservative for this type of design.

Table 3. Reaction times (RT in milliseconds) for lexical decisions to targets in the repetition (R), frequent inflection (F+), non-frequent inflection (F-) and unrelated (U) prime conditions for the two categories of verbs, large pseudo-family size (PsFam+) and small pseudo-family size verbs (PsFam-) tested in Experiments 1a and 1b. Net priming effects are given relative to the unrelated prime condition.

4.3.1 Experiment 1a. There was a significant main effect of prime type, F1(3, 366) = 19.86, p<.001. The main effect of pseudo-family size was not significant (F1<1), neither was the interaction between the two main factors, F1(3, 366) = 1.01.

Planned pair-wise comparisons show significant repetition priming for both types of verbs, F1(1, 61) = 12.79, p<.001 for PsFam+ verbs and for PSFamverbs, F1(1, 61) = 33.22, p<.001. Facilitation induced by frequent inflections was significant for large, F1(1, 61) = 5.75, p<.05, as well as for low PSF-size verbs, F1(1, 61) = 33.25, p<.001. Priming induced by non-frequent inflections was not significant, either for PsFam+ verbs, F<1, or for PsFamverbs, F1(1, 61) = 1.61. The two morphological prime conditions did not differ between them for PsFam+ verbs, F1(1, 61) = 3.27, but did for PsFamverbs, where the difference of 27ms between frequent and non-frequent inflections was significant F1(1, 61) = 11.81, p<.001. Repetition did not differ from frequent inflection conditions, either for PsFam+ verbs, F1(1, 61) = 3.29, or for PsFam- verbs, F1 < 1, but they did differ from non-frequent inflexions, both for PsFam+ verbs, F1(1, 61) = 12.99, p<.001, and for PsFamverbs, F1(1, 61) = 25.13, p<.0001. The frequent inflections did not differ from non-frequent ones for PsFam+ verbs, F1(1, 61) = 3.27, but they did for PsFam- verbs, F1(1, 61) = 11.81, p<.001.

4.3.2 Experiment 1b. The same type of analysis was conducted separately for the results of experiment 1b. Again, the main effect of prime type was significant, F1(3, 186) = 6.50, p<.001, and the main effect of pseudo-family size was not significant, F1(1, 62) = 1.99. Contrary to experiment 1a, the interaction between these two factors was significant, F1(3, 186) = 3.58, p < .05. Planned pair-wise comparisons show significant repetition priming for small PsFam size verbs, F1(1, 31) = 28.62, p<.001 but not for large PsFam size verbs, F1(1, 31) = 1.43. Morphological priming due to frequent inflections is significant for small PsFam size verbs, F1(1, 31) = 11.24, p<.001 but not for large ones, F1<1, and priming due to non-frequent inflections follows the same pattern, F1(1, 31) = 4.92, p<.05, and F1(1, 31) =3.49 respectively. Morphological priming between frequent and non-frequent inflections did not differ for PsFam- verbs, F1<1 whereas it did differ for PsFam+ verbs, F1(1, 31) = 6.13, p<.05, which is the opposite situation of the one observed in experiment 1a. The robust repetition priming (50ms) obtained for PsFam- verbs differs significantly from morphological priming, F1(1, 31) = 6.42, p<.05 for frequent and for non-frequent inflections F1(1, 31) = 9.02, p<.001, whereas morphological and repetition conditions do not differ for PsFam+ verbs, either for frequent inflections, F1(1, 31) = 3.45, or for non-frequent ones, F1<1. The frequent inflections differed from nonfrequent ones for PsFam+ verbs, F1(1, 61) = 6.13, p<.05, but they didn't for PsFam- verbs, F1<1.

4.4 Discussion for Experiment 1a and 1b.

The main outcome of this study concerns the role of the pseudo-family size jointly with frequency: under the circumstances of Exp. 1a, only primes that are frequent inflections of the infinitive targets facilitate processing, whereas non-frequent inflections fail to induce any facilitation. The fact that lexical frequencies of the primes influence processing of the targets provides another experimental demonstration that lexical frequency plays a role in morphological processing (*e.g.* for French, Giraudo & Grainger, 2000 with the masked priming technique; Meunier & Segui, 1999 with spoken primes). In other words, in the classic configuration we tested in Exp. 1a, we obtain the classic morphological priming effect induced by frequent inflections, which does not differ from repetition priming. This first result cannot be integrated in any kind of mandatory decomposition approach given that both inflections, frequent and non-frequent ones, are equally decomposable, there is therefore no reason for the frequent ones to prime and the non-frequent ones not to do so.

In experiment 1b, where the relative frequencies between primes and targets are modified comparatively to Exp. 1a, and where the disposition between primes and targets is the opposite of what the literature usually examines, we

observe that the pattern of results changes radically: only small Pseudofamily size verbs, with no antagonists at the word-form level induce repetition and morphological facilitation, which for these verbs is equivalent (arithmetically and statistically) for both frequent and non-frequent inflections. In experiment 1a, repetition priming is equivalent to morphological priming, as in the majority of similar experiments examining repetition and morphological effects, in which the target is the easiest-toactivate member of the paradigm. The fact that in Exp. 1b repetition differs considerably from morphological priming probably suggests that we are not looking at inflection effects through the same window as in the majority of studies. There is also a second reason orientating us towards the interpretation that, under the circumstances of Exp. 1b we observe masked morphological effects through a different window: the fact that frequent inflections of PsFam+ verbs fail to prime, despite having exhibited significant inflectional priming in Exp. 1a. At the same time, we observe that as soon as relative frequencies between primes and targets have been modified, thus broadening our observation window, word-forms lacking antagonists (PsFam- verbs) induce very important repetition priming and significant inflectional priming, equivalent for frequent and for non-frequent inflections.

In conclusion, the experiments presented here provided an experimental demonstration for two important effects: the first one is that not only does lexical frequency of the word-forms influence morphological processing, but relative frequencies between primes and targets also influence inflectional processing. Making the hypothesis that the lexical variable that we call frequency leaves morphological processing unaffected would be equivalent to denying these data, along with other data demonstrating this same thing (*e.g.* Giraudo & Grainger, 2000). The second effect is that of the PsFam size, an exo-lexical variable influencing inflectional processing. This influence is substantiated through inhibition exerted from primes which are pseudo-relatives of targets, *i.e.* word-forms similar in form to the target but morphologically unrelated to it. This inhibition of the pseudo-family size points towards the idea we will develop in the general discussion, namely that morphological processing effects are the sum of various kinds of activation and inhibition.

5. General discussion: Towards a revised model of morphological processing.

The literature clearly points out a number of constraints stemming from experimental data which have to be taken into account by any model of morphological processing: 1) positive priming effects produced by pseudoderived forms suggest that the cognitive system is highly sensitive to the decomposability of linguistic stimuli; (c.f. introduction of the present contribution and section on nonword effects); 2) masked morphological priming effects do not really depend on the semantic relatedness shared by prime-target pairs and differ significantly to semantic priming (c.f. the pseudo-derivation effect); 3) masked priming effects observed between allomorphs (inflected and derived) and their base requires us to consider an upper level of processing containing bases (e.g. Pastizzo & Feldman, 2002); 4) masked morphological priming effects are sensitive to lexical frequency (e.g. Giraudo & Grainger, 2001, as well as Exp. 1a of the present contribution); 5) morphological effects are sensitive to morphological family size : members of the morphological family will act as synagonists (e.g. Schreuder & Baayen, 1997); 6) masked morphological priming effects can be modulated when manipulating the relative frequencies between primes and targets (Exp. 1a and 1b); 7) under the conditions where the target is not the easiest-to-activate member of the paradigm, word-forms (primes) which are formally similar but morphologically unrelated to it, will function as antagonists, thus inhibiting the processing of the target (Exp. 1b).

Points (1) & (2) suggest that in the early stages of identification, each time a decomposable form (a regular word or non-word) is processed, it triggers the activation of its morphemic parts. This activation depends neither on semantics nor on lexicality, given that the pseudo-derived item does not have to be an established lexical unit, nor does it need to have a meaning (as we saw in section 2). We can then logically make the hypothesis that these effects take place at a sublexical level situated before the word level (*i.e.*, the orthographic lexicon). Given that facilitation can be obtained with non-words (composed by two morphemes) as well as with semantically opaque words (e.g. corner), the units coded within this sublexical level do not necessarily have anything to do with morphemes properly¹. The kind of units, thanks to which the pseudo-derivation effect (corn-corner) arises, cannot be considered as morphemes; despite this, these units bring to the fore the high saliency of morphemes across languages, in terms of statistical frequency and productivity. Consequently, we should consider these sub-units as pure surface realizations. We propose to call them morcemes as this label translates well the fact that these units, situated before the word units, are of orthographic nature but capture morpheme regularity in the language. Point (3) implies that some semantically transparent units organize the orthographic lexicon in morphological families. However, morphological variations and

¹ According to Aronoff and Fudeman (2005), morphemes correspond to "the smallest linguistic pieces with a grammatical function. (...). A morpheme may consist of a word, such as *hand*, or a meaningful piece of a word, such as the *-ed* of *looked*, that cannot not be divided into smaller meaningful parts" (p.2)

constraints realised in a given language such as allomorphy, suggest that these units have to be abstract enough. Moreover, because the function of these units consists of organising words in morphological families, they have to correspond, according to our logic, to base-lexemes (*i.e.*, nouns, verbs and adjectives). Point (4) does not need to be described in a very detailed way: it suffices to bear in mind that lexical frequency is a variable relevant for wordforms, not for sublexical units, even if statistical occurrences of bigrams, trigrams or other kinds of sublexical units may indeed affect processing. Given the well documented fact that lexical frequency influences processing, and given that in masked priming protocols two different lexical frequencies intervene, (the prime's and the target's), the manipulation of the frequency ratio between them (point 6), leads to a slight removal of the prism through which we perceive morphological effects : from the classic configuration which displays perceptual saliency between primes and targets (ex. the pseudo-derivation effect) to a situation where we enable ourselves to observe finer effects, such as the role of the pseudo-family size highlighted in point 7. Points (5) and (7) stress the role of the environment of the word-form itself: this environment can be morphologically-friendly, as for the words issued from a large morphological family. Inversely, it can be morphologically hostile, e.g. for members of big pseudo-families, i.e. similar on form but morphologically unrelated, which will act as antagonists and exert inhibition on processing the target. As experiment 1b showed, targets that are word-forms with no pseudo-family (hermits) thus having nothing to compete with at the word-form level, will directly benefit from their inflectional primes, whether they are frequent or non-frequent.

Taken together, these constraints converge towards a hybrid model of morphological processing integrating four levels of coding. As we can see in figure 1, two of them are dedicated to morphology given that morphologically complex words are coded according to two dimensions, their surface form and their internal structure.



Figure 1: The Base-Morceme-Word Model (BMWm)

The first level captures the perceptive regularity and the saliency of morphemes within the language. It contains stems and affixes that have been extracted during word acquisition. Accordingly, during language acquisition, the most salient perceptive units (*i.e.*, recurrent and regular) will be caught and coded by the cognitive system as lexical entries. At this level of coding, morphologically complex words, pseudo-derived words and nonwords whose surface structure can be divided into (at least two) distinct morphemes, are equally processed. As a consequence, this level cannot properly be considered to be a morphological level, but rather as a level containing *morphomes* in the sense defined by Aronoff (1994). Morphomes stand as access units that speed up word identification each time an input stimulus activates one of them. Therefore, there is no need to assume, at this stage, a process of *morphological* decomposition; this would be unnecessary.

Contrary to the first level, the second level deals with the internal structure of words, their formation according to morphological rules. This level contains

base-lexemes, units abstract enough to tolerate orthographic and phonological variations produced by the processes of derivation and inflection. Baselexeme representations are connected to morphologically related word representations and these connections are determined by the degree of semantic transparency between word forms and base-lexemes. Semantically transparent morphologically complex words are connected both with their morphemes and their base-lexeme. Words with a semantically opaque structure, as for example, fauvette 'warbler' (not related anymore to its freestanding stem fauve 'tawny') or with an illusory structure, as for example baguette 'stick' in which bagu- is not a stem and has nothing to do with *bague 'ring'*, are not connected with a base-lexeme. These two types of items are only connected with their surface morphemes situated at the morphome level. Indeed, the model makes the fundamental assumption that base-lexeme representations are created in long-term memory according to a rule that poses family clustering as an organizational principle of the mental lexicon. This rule stipulates that as soon as two words share both form and meaning, a common abstract representation emerges; this representation is then fed by all the incoming forms respecting this principle. In the course of language acquisition and learning family size grows and links are continually being strengthened.

The model is intended to provide an interface framework for both psychological and linguistic phenomena. On the psychological side, the current debate among psycholinguists revolves around the manner in which the lexicon is organized in terms of structural units, and the manner in which these units interact with each other during lexical access. After almost ten years of studies focusing on this issue, in particular through the manipulation of morphemes and pseudo-morphemes within masked priming experiments, two antagonistic approaches, the first one based on a mandatory morphological decomposition mechanism (Taft, 1994; Marslen-Wilson & Tyler, 2007; Rastle & Davis, 2008) and the second one on whole-word access activating intermediate morphemes (Diependaele, Sandra, & Grainger, 2005; Giraudo & Grainger, 2000) still remain.

It is interesting to note that the same antagonism remains in linguistics. On one hand, the morpheme-based approach considers that morphemes are the basic structural units of the lexicon (Halle & Marantz, 1993) and consequently word forms are analyzed as arrangements of morphemes. On the other hand, the defenders of a lexeme-based approach argue that morphology is primarily a set of systematic correspondences between word forms and meanings, and that the source of morphology is the network of paradigmatic relations between words existing in a language (Aronoff, 1994; Bybee, 1988; 2001, Booij, 2002; to appear). From a psycholinguistic perspective, the hybrid model we propose accounts for a large part of psycholinguistic data and can make very precise predictions about morphological priming effects. According to the model, priming effects depend on the kind of relation the prime entertains with the target (formal and/or semantic) and consequently, on the number of excitation springs that target recognition triggers: a) when the prime is semantically transparent and complex M+O+S+ (like in the pairs banker-bank or hatchedhat), its perception gives birth to three springs of excitation (SoE), from morphomes, word forms and base-lexemes; b) when the prime is semantically transparent, complex but not decomposable M+O-S+ (like in the prime-target pair *fell-fall*), it activates two SoEs, from word forms and base-lexemes; c) when the prime is semantically opaque M+O+S- (it concerns complex or pseudo-complex words like *apartment-apart* or *corner-corn*), its recognition triggers two SoEs, from morphomes and word-forms; d) when the prime is not complex and not decomposable M-O-S- (like freeze-free), it gives raise to only one SoE, from word-forms.

The masked priming data collected until now stated the following results: banker-bank=corner-corn > freeze-free about derivation (see Rastle and Davis, 2008 for a review) and hatched-hatch > fell-fall > teach-taught about inflection (Pastizzo & Feldman, 2002; Crepaldi et al., 2010). The predictions derived from the hybrid model we present above provide a more nuanced picture: *banker-bank* (3 SoEs) > *corner-corn* (2 SoEs) > *freeze-free* (1 SoE). Nevertheless, the role of psycho-physical characteristics of the protocol should not be completely discarded. In most masked priming studies, prime exposure duration ranges from 48ms to 60ms. In this case, and with the particular design discussed in section 1, what we observe as priming effects corresponds to a small window of the overall activation, as we demonstrated through the experiments presented here. This characteristic could explain why data revealed a banker-bank effect which was equal to the corner-corn effect. When increasing the SOA, the advantage that morphologically, semantically and orthographically related prime-target pairs have over morphologically (very opaque)² and orthographically related but semantically unrelated pairs, emerges (see Rastle, Davis, Marslen-Wilson, & Tyler, 2000).

² It is interesting to note here that in a significant proportion of psycholinguistic studies, as in the studies cited here, the category of semantically opaque items mixes morphologically complex words, whose structure is opaque as a result of complex etymology but remains relatively accessible synchronically (e.g. *fauvette*), with morphologically simple words whose surface can be segmented into morpheme-like sub-units (e.g. *corn-er*, in the Rastle 2000 study, *chant-ier*, in the Meunier 2005 study), without making any difference between them.

6. Conclusions

Previous models of morphological processing, and in particular those claiming that morphological information has to be represented at a sublexical level of processing, make, in our view, an error in that they confuse the morphemic unit as a subpart of a word with its linguistic function. Starting from the general postulate that cognitive models of information processing are coding external information from basic and primary features (e.g., letter features) to the most complex characteristics (e.g., concepts), the morpheme, a unit perceptively smaller that the word, has been implicitly classified at a lower level on the scale of information complexity. Experimental studies examined morphological processing through various explicit manipulations of morphemes: within non-words or complex words showing their determinant role in reading, between primetarget pairs demonstrating the earliness of morphological processing and the need to represent morphology as a separate level of processing, and finally within simple words and non-words emptying the morpheme of its linguistic functions. A non negligible part of the psycholinguistic literature has, little by little, lost sight of the linguistic function of morphology to focus only on surface information. From a linguistic point of view, morphology is not only reduced to a surface form or a word's subpart. Base-lexemes refer to a semantic field that is common to all their derivations and inflections. There's no need to explicitly state that the French word écolier (which means 'pupil') and scolaire (wich means 'scholar') both derive from the Latin base schola to convince the native speaker of their morphological link. Moreover, this link is not perceived as being only semantic in nature. Even in Hebrew, which doesn't have a linear morphological structure (the consonants of the root are intertwined with the word-pattern phonemes), Velan and Frost (2011) recently demonstrated that native speakers processed Hebrew words with a Semitic structure (with an internal structure) differently to Hebrew words borrowed from Indo-European languages (without any internal structure). Our conclusion is that the above considerations, along with a great deal of experimental data stemming from various techniques, and especially masked priming, strengthen the idea that the readers' morphological representation plays a determinant role in the organization of their mental lexicon. The role

plays a determinant role in the organization of their mental lexicon. The role of perceptive saliency of surface morphemes is certainly very important, yet it constitutes merely "the tip of the iceberg". Bringing out the organisational functions at the interface of form and semantics, which constitute its hidden part, requires us to include variables related to the lexical status of items, as well as paradigmatic relations and factors outside the word-to-be-studied.

References

- Amenta, S., D. Crepaldi. 2012. "Morphological processing as we know it: An analytical review of morphological effects in visual word identification". Frontiers in Language Sciences 3, 232.
- Arduino L.S., C.Burani. 2004. "Neighborhood effects on pseudoword visual processing in a language with shallow orthography". Journal of Psycholinguistic Research 33: 75-95.
- Aronoff, M. 1994. Morphology by itself. Cambridge: MIT Press.
- Aronoff, M., K.Fudeman. 2005. What is morphology? Blackwell Publishing.
- Baayen, R.H., T. Dijkstra, R. Schreuder. 1997. "Singulars and plurals in Dutch:evidence for a parallel dual route model". Journal of Memory and Language 37: 94–117.
- Bertram, R., R. H Baayen, R. Schreuder. 2000. "Effects of family size for complex words". Journal of Memory and Language 42: 390-405.
- Beyersmann, E., J.A. Duñabeitia, M.Carreiras, M. Coltheart, A. Castles. 2013. "Early morphological decomposition of suffixed words: Masked priming evidence with transposed-letter nonword primes". Applied Psycholinguistics 5: 869-892.
- Booij, G. E. 2002. The Morphology of Dutch. Oxford: Oxford University Press.
- Booij, G.E. Word structure. To appear in J. Taylor (ed). The Oxford handbook of the Word. Oxford: Oxford University Press.
- Bowers, J.S., C.J. Davis, D. Hanley. 2005. "Is there a "hat" in "that"? Journal of Memory and Language 52: 131-143.
- Burani, C., F.M. Dovetto, A.M. Thornton, A. Laudanna. 1997. "Accessing and naming affixed pseudo-words" in G. Booij & J.von Marle (eds), Yearbook of Morphology 1996. Dordrecht: Kluer Academic Publishers, 55–72.

- Burani, C., D. Salmaso, A. Caramazza. 1984. "Morphological structure and lexical access". Visible Language 18: 342–352.
- Burani, C., A.M. Thornton. 2003. "The interplay of root, suffix and wholeword frequency in processing derived words" in R.H. Baayen, R. Schreuder (eds), Morphological Structure in Language Processing. Berlin: Mouton de Gruyter, 157–208.
- Bybee, J. 1988. "Morphology as lexical organisation" in M. Hammond and M. Noonan (eds), Theoretical Morphology. Approaches to modern linguistics. San Diego: Academic Press, 119-142.
- Bybee, J. 2001. Phonology and Language Use. Cambridge: Cambridge University Press.
- Caramazza, A., A. Laudanna, C. Romani. 1988. "Lexical access and inflectional morphology". Cognition 28: 297-332.
- Christianson, K., R.L.Johnson, K. Rayner. 2005. "Letter transpositions within and across morphemes". Journal of Experimental Psychology: Learning, Memory and Cognition 31: 1327–1339.
- Coltheart, M., E. Davelaar, J. T. Jonasson, D. Besner. 1977. "Access to the internal lexicon" in S. Dornic (ed), Attention and Performance VI. New York: Academic Press 535-555.
- Coltheart, M., B. Curtis, P. Atkins, M. Haller. 1993. "Models of reading aloud: Dual-route and parallel-distributed-processing approaches". Psychological Review 100: 589-608.
- Coltheart, M. 2001. "Learning to read: The self-teaching hypothesis investigated cross-linguistically. Contemporary Psychology 46: 205-207.
- Colé, P., C. Beauvillain, J. Segui. 1989. "On the representation and processing of prefixed and suffixed derived words: a differential frequency effect". Journal of Memory and Language 28: 1–13.

- Crepaldi, D., K. Rastle, M. Coltheart, L. Nickels. 2010. "Fell primes fall but does bell prime ball ? Masked priming with irregularly inflected primes". Journal of Memory and Language 63: 83–99.
- Davis C. J. 1999. "The Self-organizing lexical acquisition and recognition SOLAR model of visual word recognition". Unpublished Doctoral dissertation, University of New South Wales.
- De Jong, N. H., L. B Feldman, R. Schreuder, M. Pastizzo, R. H. Baayen. 2002. "The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects". Brain and Language 81: 555–567.
- De Jong, N., R. Schreuder, R. H. Baayen. 2000. "The morphological size effect and morphology". Language and Cognitive Processes 15: 329-365.
- Diependaele, K., J. Morris, R. M. Serota, D. Bertrand, J. Grainger. 2013. "Breaking boundaries: Letter transpositions and morphological processing". Language and Cognitive Processes 287: 988-1003.
- Diependaele, K., D. Sandra, J. Grainger. 2005. "Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition". Language and Cognitive Processes 20: 75-114.
- Duñabeitia, J.A., M. Perea, M. Carreiras. 2007. "Do transposed-letter similarity effects occur at a morpheme level? Evidence for morphoorthographic decomposition". Cognition 1053: 691-703.
- Duñabeitia, J.A., M. Perea, M. Carreiras. 2008. "Does darkness lead to happiness? Masked suffix priming effects. Language and Cognitive Processes 23: 1002-1020.

- Ford, M.A., M.H Davis, W.D. Marslen-Wilson. 2010. "Derivational morphology and base morpheme frequency". Journal of Memory and Language 631: 117-130.
- Forster, K. I. 1976. "Accessing the internal lexicon" in R. J. Wales & E. C.T. Walker (eds), New approaches to language mechanisms.Amsterdam: North Holland.
- Forster, K. I., J. C. Forster. 2003. "DMDX: A Windows display program with millisecond accuracy". Behavioral Research Methods: Instruments & Computers 35: 116-124.
- Giraudo, H., J. Grainger. 2000. "Effects of prime word frequency and cumulative root frequency in masked morphological priming". Language and Cognitive Processes 15: 421-444.
- Giraudo, H., J. Grainger. 2001. "Priming complex words: Evidence for supralexical representation of morphology". Psychonomic Bulletin and Review 81: 127-131.
- Giraudo, H., M. Voga, M. 2013. "Prefix units within the mental lexicon" in N. Hathout, F. Montermini, J.Tseng (eds), Morphology in Toulouse. Selected Proceedings of Décembrettes 8. Lincom Studies in Theoretical Linguistics: Lincom Europa, 61-78.
- Grainger, J., P. Colé, J. Segui. 1991. "Masked morphological priming in visual word recognition". Journal of Memory and Language 30: 370-384.
- Grainger, J., A. M.Jacobs. 1996. "Orthographic processing in visual word recognition: A multiple read-out model". Psychological Review 103: 518-565.
- Grainger, J., J. K.O' Regan, A. M.Jacobs, J. Segui. 1992. "On the role of competing word units in visual word recognition: The neighbourhood frequency effect". Perception and Psychophysics, 45: 189-195.

- Halle, M., A. Marantz. 1993. "Distributed Morphology and the pieces of inflection" in K. Hale & J. Keyser, (eds), The View from Building 20. Cambridge: MIT Press, 111-176.
- Haspelmath, M., A. D. Sims. 2010. Understanding morphology. 2nd edition. London: Hodder Education.
- Hay, J.B., R.H.Baayen. 2003. "Phonotactics, Parsing and Productivity". Italian Journal of Linguistics 1: 99-130.
- Humphreys, G. W., L. J. Evett, P.T. Quinlan. 1990. "Orthographic processing in visual word identification". Cognitive Psychology 22: 517-560.
- Jacobs, A.M., J. Grainger. 1994. "Models of visual word recognition -Sampling the state of the art". Journal of Experimental Psychology: Human Perception and Performance 20: 1311-1334.
- Laudanna, A., W. Badecker, A. Caramazza. 1989. "Priming homographic stems". Journal of Memory and Language 28: 531-546.
- Longtin, C. M., J. Segui, P. A. Hallé. 2003. "Morphological priming without morphological relationship". Language and Cognitive Processes 183: 313-334.
- Longtin, M-C., F. Meunier. 2005. "Morphological decomposition in early visual word processing". Journal of Memory and Language 53:1, 26-41.
- Meunier, F., C.-M.Longtin. 2007. "Morphological decomposition and semantic integration in word processing". Journal of Memory and Language 56: 457–471.
- López-Villaseñor, M.L. 2012. "The Effects of Base Frequency and Affix Productivity in Spanish". The Spanish Journal of Psychology 15 (2): 505-512.

- Massol, S., J. Grainger, S. Dufau, P. Holcomb. 2010. "Masked priming from Orthographic Neighbors: An ERP Investigation". Journal of Experimental Psychology: Human, Perception and Performance 361: 162-174.
- McClelland, J. L., D.E.Rumelhart, 1981. "An interactive activation model of context effects in letter perception: Part 1. An account of basic findings". Psychological Review 88: 375-407.
- McCormick, S. F., K. Rastle, M.H. Davis. 2008. "Is there a 'fete' in 'fetish'? Effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition". Journal of Memory and Language 58: 307-326.
- Meunier, F., J. Segui. 1999. "Morphological priming effect: the role of surface frequency". Brain and Language 68: 54–60.
- Morris, J., J. Grainger, P. J. Holcomb. 2013. "Tracking the consequences of morpho-orthographic decomposition using ERPs". Brain Research 1529: 92-104.
- Moscoso del Prado Martín, F., A. Deutch, R. Frost, R. Schreuder, N. De Jong, R.H. Baayen. 2005. "Changing places: a cross-language perspective of the role of morpholgy in the processing of Hebrew and Dutch". Jouranl of Memory and Language 53: 496-512.
- Pastizzo, M.J., L.B.Feldman. 2004. "Morphological processing: a comparison between free and bound stem facilitation". Brain and Language 90: 31–39.

Pinker, S. 1991. "Rules of language". Science 253: 530-535.

- Plag, I. 1999. Morphological productivity: Structural constraints in English derivation. Berlin, New York: Mouton de Gruyter.
- Plag, I. 2004. "Syntactic category information and the semantics of derivational morphological rules", Folia Linguistica 38: 3-4, 193-225.

- Raaijmakers, J.G.W., J.M.C.Schrijnemakers, F. Gremmen, F. 1999. "How to deal with The language-as-fixed-effect fallacy: Common misconceptions and alternative solutions". Journal of Memory and Language 41: 416-426.
- Rastle, K., M. Davis, W. Marslen-Wilson, L. Tyler. 2000. "Morphological and semanticeffects in visual word recognition: a timecourse study". Language and Cognitive Processes 15: 507–537.
- Rastle, K., M.H. Davis. 2008. "Morphological decomposition based on the analysis of orthography". Language and Cognitive Processes 23: 942– 971.
- Rastle, K., M.H. Davis, B. New. 2004. "The broth in mybrother's brothel: morpho-orthographic segmentation in visual word recognition". Psychonomic Bulletin and Review 11: 1090–1098.
- Rueckl, J.G., A. Rimzhim. 2011. "On the interaction of letter transpositions and morphemic boundaries". Language and Cognitive Processes 26: 482–508.
- Sánchez-Gutiérrez, C., K. Rastle. 2013. "Letter transpositions within and across morphemic boundaries: Is there a cross-language difference?" Psychonomic Bulletin & Review 20(5): 988-996.
- Schreuder, R., R. H. Baayen. 1997. "How complex simplex words can be". Journal of Memory and Language 37: 118-139.
- Stanners, R.F., J. J. Neiser, W. P.Hernon, R. Hall. 1979. "Memory representation for morphologically related words". Journal of Verbal learning and Verbal Behavior 18: 399-412.
- Taft, M. 1979. "Lexical access via an orthographic code: The Basic Orthographic Syllabic Structure BOSS". Journal of Verbal Learning and Verbal Behavior 18: 21-39.

- Taft, M. 2004. "Morphological decomposition and the reverse base frequency effect". Quarterly Journal of Experimental Psychology 57A: 745-765.
- Taft, M., K.I. Forster, 1976. "Lexical storage and retrieval of polymorphemic and polysyllabic words". Journal of Verbal Learning and Verbal Behavior 15: 607-620.
- Taft, M., K. I.Forster. 1975. "Lexical storage and retrieval of prefixed words". Journal of Verbal Learning and Verbal Behavior 14: 638-647.
- Voga, M. & H. Giraudo. 2009. "Pseudo-family size influences processing of French inflections : evidence in favor of a supralexical account" in F. Montermini, G. Boyé, J. Tseng (eds), Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux. Somerville, MA: Cascadilla Proceedings Project, 148-155.

INTERNAL LOCALISATION NN_{ADV} REDUPLICATION IN SICILIAN

Giuseppina Todaro, Florence Villoing & Philippe Gréa Università Roma Tre / CLLE-ERSS, CNRS & Université Toulouse 2 Jean Jaurès, Université Paris 8 Saint-Denis, Université Paris Ouest Nanterre

Abstract

This article proposes a classification of reduplicative structures in Sicilian and provides a detailed analysis of one such structure, which forms adverbs through complete reduplication of nominal lexemes or word forms (NN_{adv}). We show that this construction is part of the basic morphology of Sicilian. This subject is of interest for two reasons. First, Sicilian morphology is only rarely addressed in the linguistic literature, and Sicilian itself is always treated only from a dialectological or philological point of view. Moreover, this article proposes a morphological analysis of NN_{adv} reduplication, which is a rare phenomenon in Romance languages, and has consequently never been the object of an in-depth study, particularly not from a lexicalist perspective. While the literature on reduplication cross-linguistically highlights semantic values associated with plurality, verbal aspect, intensity, repetition and expressiveness, this article shows that reduplication can express a locative meaning.

1. Introduction

Reduplication is one of the morphological lexeme-formation processes available in Sicilian, but is only rarely addressed in the literature, and then only from a syntactic perspective (Caracausi 1977, Leone 1995, Sgarioto 2005, Amenta 2010). In this article, we present a classification of reduplicative structures in Sicilian and provide a more detailed formal, categorial and semantic analysis of a rare morphological reduplication rules those forming internal localisation adverbs through nominal reduplication (1):

(1) $[[strata]_N - [strata]_N]_{Adv}^{-1}$ street-street 'in the street'

This study is particularly original in several points:

(i) Sicilian is one of the few Romance languages (along with Sardinian, cf. Floričič 2011), which presents multiple morphologically productive reduplicative rules. Although reduplication exists in many languages (Mel'čuk 1996-97), it is typically considered a basic morphological process in Austronesian languages, ancient African languages of the Niger-Congo family or in Creoles (Kouwenberg 2003), but not in Indo-European languages, where it is rarely found (see overview in Wiltshire & Marantz 2000).

(ii) The reduplication rules studied in Sicilian are considered as being part of constructional morphology, whereas the general tendency in morphological studies of reduplication involves inflexional morphology (Thornton 2009).

From a morphological perspective, the present study differs from recent studies of reduplication in both formal and semantic terms, since:

- morphological reduplication constructions in Sicilian involve only complete reduplication, which is frequently considered trivial and uninteresting, both empirically and theoretically (Katamba 1993, Lieber 1992) compared to partial reduplication, which has drawn much more attention in morphological studies (possibly attributing an affixal status to the reduplicated element, depending on its position in the word (Matthews 1991, Spencer 1991));

– semantic properties of reduplication have drawn less attention in the literature than morpho-phonological (cf. templatic morphology (Marantz 1982)) and prosodic properties (McCarthy & Prince 1990). Since Sicilian NN_{adv} reduplication produces quite original semantic values compared to those typically observed for this type of process (plurality, verbal aspect, intensity, repetition, expressiveness) (Mel'čuk 1996-97, Wiltshire & Marantz 2000).

This paper is organized as follows: we first define **reduplication** as a morphological process, distinguishing it from the syntactic process, which we term **reiteration** (2.1). We use the tests proposed by Gil (2005) to identify the

¹ We use orthographic transcription (that evokes Italian spelling but not for the phonetic peculiarity of Sicilian), since phonetic transcription is not relevant for the present analysis.

boundary between the two levels in which the juxtaposition of two otherwise autonomous forms occurs (morphology and syntax). Next, we propose a theoretical analysis that places reduplication among word-formation rules, although distinct both from derivation and compounding (2.2). Section 3 is devoted to reduplication in Sicilian: we propose a classification of reduplication constructions (3.2) based on the data presented in (3.1). Finally, in section 4, we analyse one of the most common reduplication constructions (NN_{adv}), which forms adverbs through complete reduplication of a noun. We present the categorial (4.1), morpho-syntactic (4.2) and semantic properties (4.3.) of this construction. In particular, we show that this reduplication rule forms dynamic (4.3.1) and static (4.3.2) internal localisation relations.

2. Reduplication

The structures formed by the repetition of the same linguistic form (or part thereof), repeated twice or more, is known by a variety of terms in the literature (reduplication, repetition, reiteration, iteration, doubling). These terms refer to a series of heterogeneous phenomena, both formally (complete vs. partial reduplication) and in terms of the linguistic level involved (phonology, morphology, syntax). In what follows, we try to provide the arguments that characterize this phenomenon on the morphological level. We will keep the term reduplication to account for the morphological processes, involving a constructional (2) or an inflectional (3) phenomenon, that builds lexemes or words through complete (2, 3) or partial (4) reduplication of the base (Anderson 1992, Lieber 1992, Katamba 1993, Booij 2010, Wiltshire&Marantz 2000).

(2)	Italian	
	La bambina mangia un lecca-lecca alla	fragola
	DET little girl eat.3SG DET lick-lick PREP(ART)	strawberry
	'The little girl licked a strawberry lollipop'	

(3)	Sranam saka bag	saka-saka ² bag.PL		
(4)	Ilocano píŋgan plate	piŋpíŋgan plate.PL	dálan road	daldálan ³ road.PL

² Exemple from Aboh, Smith, Zribi-Hertz 2012.

³ Exemple from Gleason 1955.

We distinguish between *reduplication* and *reiteration*, reserving the latter term for syntactic repetition (or *syntactic emphasis, iconic function*, Forza 2011, Aboh, Smith, & Zribi-Hertz (eds) 2012 (introduction), Moravcsik 1978:301), as in (5-6).

- (5) Italian Correva, correva, ma non riusciva a raggiunger-la run.IPFV.3SG run.IPFV.3SG but NEG manage.IPFV.3SG PREP catch-her 'He ran, he ran, but he couldn't catch her'
- (6) Italian

Mi guardava con quei suoi occhi neri neri REFL look.IPFV.3SG PREP DET.PL POSS.PL eyes black.PL black.PL 'He/she looked at me with her/his black-black eyes'

2.1 The external boundaries of reduplication: criteria distinguishing morphological reduplication and syntactic reiteration

In order to differentiate between the morphological process (our *reduplication*) and the syntactic process (our *reiteration, repetition* in Gil 2005), we adopt the criteria proposed by Gil (2005) to establish the boundary between syntactic and morphological phenomena.

Criterion	Repetition	Reduplication	
1. Unit of output	greater than a word	equal to or smaller than a word ⁴	
2. Communicative reinforcement	present or absent	absent	
3. Interpretation	iconic or absent	arbitrary or iconic	
4. Intonational domain of output	within one or more intonation group	within one intonation group	
5. Contiguity of copies	contiguous or disjoint	contiguous	
6. Number of copies	two or more	usually two	

Table 1: Criteria distinguishing reduplication and reiteration (Gil 2005)

⁴ We do not accept the possibility of having a unit smaller than a word as output, since as a morphological process, reduplication cannot produce units larger or smaller than the word level.

In the next sections, we analyze examples from (2-6) above in accordance with five criteria:

- domain (2.2.1)
- interpretation and communicative function (2.2.2)
- phonology (intonational domain) (2.2.3)
- contiguity and written form (2.2.4)
- number of copies (2.2.5).

2.2.1 Domain

The output of reiteration (*repetition* in Gil 2005) should involve units larger than a word, whereas reduplication should involve lexemes and word-forms. In (5), (6), the output is larger than a word, therefore is the result of a syntactic operation (in Gil's 2005 terms). In contrast, in (2-4), the output is found within word boundaries. To test this hypothesis, we can apply the insertion test, which examines whether lexical material can be inserted in the middle of the output string. If the example remains grammatical after insertion, we are dealing with a syntactic unit; if it is ungrammatical, the output is a lexical unit (as illustrated below with one example for each type).

a. Insertion of lexical material into the output in (6)

Italian
 Mi guardava con quei suoi occhi neri proprio neri
 REFL look.3SG PREP DET.PL POSS.PL eyes black.PL so black.PL
 'She looked at me with her black, so black eyes.'

b. Insertion of lexical material into the output in (2)

- (8) Italian
 - * La bambina mangia un lecca e lecca alla fragola DET little girl eat.3SG DET lick and lick PREP(ART) strawberry 'The little girl eats a strawberry lollypop.'

The application of this test demonstrates that (7) involves a syntactic operation, which creates syntactic units, while (8) involves a morphological operation forming a new lexeme which cannot be broken up by syntax. Thus, morphological reduplication is not the iteration of a word but the iteration of a lexical item before it becomes available to syntactic processes, which obviously operate on a higher level of language (Forza 2011).
2.2.2 Interpretation and communicative function

These two criteria are based on the semantics of the output. The intention to generate *communicative reinforcement* or an iconic interpretation (Kouwenberg, S. (ed) 2003) suggests that the process under discussion may be a syntactic operation (for a critical analyses of iconicity see Aboh, Smith & Zribi-Hertz 2012). If the intention and the interpretation are different, then the process involved is morphological reduplication. Examples (5-6), repeated here (an every time) for convenience in (9-10), show communicative strengthening (insistence that underlines the recurrence or long duration of the action (9) or of the particular property (10)) and the interpretation is iconic (repetition of the action in (9) and the amplification of the quality in (10)).

(9) Italian

Correva, correva, ma non riusciva a raggiunger-la run.IPFV.3SG run.IPFV.3SG but NEG manage.IPFV.3SG PREP catch-her 'He ran, he ran, but he couldn't catch her'

(10) Italian

Mi guardava con quei suoi occhi neri neri REFL look.IPFV.3SG PREP DET.PL POSS.PL eyes black.PL black.PL 'He/she looked at me with her/his black-black eyes'

In contrast examples (2-3), repeated here in (11-12), show no communicative strengthening, but a real specific sense (action \rightarrow object (11), plurality (12)).

(11) Italian

La bambina mangia un lecca-lecca alla fragola DET little girl eat.3SG DET lick-lick PREP(ART) strawberry 'The little girl licked a strawberry lollipop'

(12) Sranam saka saka-saka bag bag.PL

2.2.3 The intonational domain criterion

The intonational criterion is based on the assumption that the phonological form of a word forms a single intonational domain, while a syntactic unit may correspond to several intonation domains. This is a criterion that can reinforce the others (although it can be difficult and subjective, in a complex word to distinguish a primary stress, especially in a reduplicate or a compound item). Every Italian speaker would clearly feel the intonational and accentuation difference between the reduplicated nominal in (13) and the imperative form in (14):

- (13) $[[lecca]_V-[lecca]_V]_N \\ lick-lick \\ `lollipop'$
- (14) lecca(!) lecca! lick(!) lick! 'lick (!) lick!'

The sequence in (13) is a reduplicate word (a lexeme), while the sequence in (14) is a sentence, the repetition of an order, a repeated imperative form (a syntactic structure) that contains a longer pause between the two parts.

2.2.4 Graphical contiguity

Elements which are a part of the same lexical unit show a certain degree of contiguity at the phonological level as well as at the graphic level (but we shall see that the written form remains quite arbitrary). If there is no contiguity and the elements are separated, then we are dealing with more than a single lexical item. Contiguity, which is even evident on the written form, exists between the input components in (15), but not in (16), where the two forms are even separated by a comma (in writing) or a long pause (in oral). As it involves the written form, this point is also a rather arbitrary parameter, mentioned only as a strengthening criterion.

(15) Italian

La bambina mangia un lecca-lecca alla fragola DET little girl eat.3SG DET lick-lick PREP(ART) strawberry 'The little girl licked a strawberry lollipop'

(16) Italian

Correva, correva, ma non riusciva a raggiunger-la run.IPFV.3SG run.IPFV.3SG but NEG manage.IPFV.3SG PREP catch-her 'He ran, he ran, but he couldn't catch her'

The example in (15) can have three different representations in written Italian language: *lecca lecca, lecca-lecca, leccalecca* (Thornton-2007). Anyway Italian speakers conceive that as a single word, while in (16) the two elements are certainly separated by a comma and conceived of as two different words.

2.2.5 Number of copies

This criterion is based on the recursivity of the operation. Since reiteration is a recursive process, the ability to repeat the operation several times (and thus to have the base form repeated at least three times) is an argument in favor of the hypothesis that it is a syntactic process. However, a non-recursive operation, that is, if repetition of the operation (having at least three copies of the base form) creates an ungrammatical output, is an argument in favor of the hypothesis that it is morphological reduplication.

(17) Italian

Correva, correva, correva ma non riusciva a raggiunger-la run.IPFV.3SG run.IPFV.3SG run.IPFV.3SG but NEG manage.IPFV.3SG PREP catch-her 'He ran, he ran, but he couldn't catch her'

(18) Italian

*La bambina mangia un lecca-lecca alla fragola DET little girl eat.3SG DET lick-lick PREP(ART) strawberry **The little girl licked a strawberry lick-lick-lick'

The example in (17) indicates that a syntactic operation is involved, while in (18), a prototypical syntactic feature such as recursivity makes the sentence ungrammatical.

2.2 Internal boundaries of morphological reduplication

The aim of this article is not only to show, through the analysis of semantic nuances, that reduplication in Sicilian is a morphological rather than a syntactic process, as assumed in the literature (Caracausi 1977, Leone 1995, Sgarioto 2005, Amenta 2010), but also to show, through the examination of morphological reduplication cross-linguistically, that this phenomenon cannot be classified either as derivation or as compounding. In some analyses which do not view morphological reduplication as a distinct morphological process, this phenomenon has been analysed as a type of derivation (Aronoff 1976, Inkelas&Zoll 2005, Scalise-Bisetto 2008) or a type of compounding (Bauer 2003).

(i) In derivational morphology, every affixal operation is associated with a specific function. Thus, the Italian suffixal rule represented by *mente* (like Fr. *-ment*, Eng. *ly*) always forms adverbs. An affix is a non-autonomous phonological form that is associated with a specific semantic meaning and a lexical category. In contrast, the type of affixes referred to in the discussion of reduplication have neither a specific shape nor their own semantic value. In

other words, the supposed reduplicative affixes are not affixes in the sense that they are not phonological units, while derivational affixes are stable phonological representatives of morphological rules.

(ii) Compounding is, by definition, a process that selects two different lexemes. The analysis of reduplication as a type of compounding can therefore apply only to complete reduplication and cannot be applied to partial reduplication, since one of the components involved is not lexeme. In addition, compounding, by definition, selects two distinct lexemes.

We therefore conclude that neither derivation nor compounding can include reduplication as a subtype. Consequently, we propose the hypothesis that reduplication is an autonomous morphological process, alongside derivation and compounding:

(19) Morphological rules of word formation q|p Derivation Compounding Reduplication

3 Reduplication in Sicilian

Sicilian employs both (syntactic) reiteration and (morphological) reduplication. In this context, we present only reduplicative constructions. Below, we outline our data collection (3.1), present a classification of structures (3.2) and analyze the categorical (3.3) and morpho-phonological (3.4) properties involved. The semantic properties are discussed in section 4.

3.1 The data

The absence of a large-scale corpus limits studies on the morphology of Sicilian. In fact, as it is fundamentally a spoken language, written production is very limited. For example, the only analysis of word formation in Sicilian (Emmi 2011) uses a corpus that dates from the first decades of the 20th century (based on theatrical texts and poems). The data used here are derived from a corpus of units from dictionaries and works in Sicilian Philology, Dialectology and linguistics, and from a field survey of speakers of varied generations (n=20, aged 18-80).

Speakers were interviewed using a questionnaire (based on testing Italian/Sicilian translations) with two objectives in mind: (i) to verify the presence and availability of reduplicative forms; (ii) to analyse the semantic value of these forms. Although the most tested variety in our data was Western Sicilian (Trapani, Palermo) (our examples are in this variety too),

the phenomenon is attested on the entire island and in Southern Italy as well (Rohlfs 1969).

3.2 Classification of reduplicative structures in Sicilian Sicilian has three reduplication rules forming lexical units.

1) Verb reduplication forming adjectives: VV>A (intensification)

(20) [[cala]_V-[cala]_V]_A go down-go down 'goes down easily'
Context: 'stu vinu è cala-cala DET wine be.3SG go down-go down 'it's good wine, it goes down easily, it's sweet, easy to drink'

This kind of morphological rule produces adjectives that can, in some cases, have the syntactic position of a noun.

- (21) [[palla]_V-[palla]_V]_A
 talk-talk
 'person who talks too much / glib talker'
 Context:
- (22) Claudio è **palla-palla** Claudio be.3SG talk-talk 'Claudio speaks too much'
- (23) Claudio è **un palla-palla** Claudio be.3SG DET talk-talk 'Claudio is someone who speaks a lot/too much'

In (22) *palla-palla* occupies the syntactic position of an adjective (after a verb), in (23) the reduplicative form (*palla-palla*) occupies a nominal position (after a determiner).

2) Noun reduplication forming adjectives: NplNpl>A

```
(24) ['mpuddri]<sub>N</sub>-['mpuddri]<sub>N</sub>]<sub>A</sub>
pimple.PL-pimple.PL
'pimply'
Context:
Avi a facci 'mpuddri-'mpuddri
have.3SG DET face pimple.PL-pimple.PL
'His face is pimply'
```

3) Nominal reduplication forming adverbs: NN>Adv (internal localisation)

 (25) [[miccato]_N-[miccato]_N]_{Adv} market-market 'at the market, in the market, moving in the market' Context:

Ti ciccava **miccato-miccato** REFL search.IPFV.ISG market-market 'I looked for you in the market /everywhere in the market'

4 Sicilian Morphological reduplication NN>Adv

 NN_{Adv} reduplication is a morphological rule (cf.2.1.) that forms (i) a lexical unit (e.g., no insertion possible between the nouns, single primary stress, no recursivity) in a different category from its base and (ii) a specific meaning expressing internal location, which does not correspond to the reiteration of the nominal base, as would be expected of a syntactic construction.

4.1 Categorial Properties

The morphological operation can be formalised as NN > Adv, indicating the output is an adverb, as indicated by the following criteria.

- The adverb is uninflected (so, its number does not depend on the syntax of the sentence).

- (26) a. U picciriddru ioca casa-casa DET little boy play.38G house-house 'The little boy plays all over the house'
 - b. I picciriddri iocano casa-casa DET.PL little boy.PL play.3PL house-house 'The little boys play all over the house'

In this case, although the sentence is in the plural, the adverb *casa-casa* ('house-house') retains the same form and the sentence can be ambiguous. Thus, it can mean 'all the little boys are in the same home', but can also have a distributive interpretation referring to different homes, 'every little boy is in his own home' (although this is quite forced and not the first interpretation, it is possible in a clear context).

– The adverb is (usually) used to semantically modify the meaning of a verb, but may also modify an adjective, another adverb or an entire sentence (Beccaria 2004). Thus, *casa-casa* semantically modifies the verb in (26.a) and (26.b), rather than the noun (as would an adjective), by adding localization information on where the action takes place, and by adding a dynamic dimension, which will be analyzed below. Syntactically, the NN-reduplicated adverb is equivalent to a prepositional spatial locative in another language, as in Fr. *à la maison* 'at home' in (27).

(27) Je suis à la maison 'I'm at home'

However, removing the reduplicated form in (26a) leads to ungrammaticality.

(28) ^{*}U picciriddro ioca casa DET little boy play.3SG house ^{*}The boys play house'

Consequently, reduplication falls under the category Adv, since the only way to rescue the grammaticality of (28) is to add a preposition, forming a PP with a similar function to replace the reduplication.

(29) I picciriddri iocano n casa DET.PL little boy.PL play.3PL PREP house 'The little boys play at home'

But, as shown below (sect. 4.3.), the semantic value of reduplication is more than a simple locative Adv (such as in $n \ casa$ 'at home').

4.2 Morpho-phonological properties

 NN_{Adv} reduplication in Sicilian involves complete reduplication: the base noun is fully reduplicated. In most cases, the nominal base is typically a lexeme (26), but may also be a form inflected for plural number (30). The inflected input form can be analyzed here as what Booij (1996) calls *inherent inflection*, rather than *contextual inflection*. Compare (30) and (31).

- (30) U dutturi va casi-casi DET doctor go.3SG house.PL-house.PL 'The doctor goes from house to house'
- (31) I duttura vannu casi-casi DET doctor.PL go.3PL house.PL-house.PL 'The doctors go from house to house'

This comparison shows that the inflection of the reduplicated input is not determined by the syntax of the sentence. Therefore, the inflected input adds a specific semantic value (=from one place to another), sometimes corresponding to *pluralia tantum* (34).

- (32) [[casa]_N-[casa]_N]_{Adv} house-house 'at home, in the house, inside the house'
- (33) [[casi]_N-[casi]_N]_{Adv} house.PL-house.PL 'from house to house'

Context:

Vinni i so cosi firriannu casi-casi sell.3SG DET POSS thing.PL go.PROG house.PL-house.PL 'He sells things going from house to house'

(34) $[[terri]_N-[terri]_N]_{Adv}$ ground.PL-ground.PL

'on the ground, in the country'

Context:

Cecca u so atto terri-terri look.3SG DET POSS cat ground.PL-ground.PL 'He's looking for his cat in the grounds, in the country'

4.3 Semantic properties: internal localisation

From a semantic point of view, NN > Adv reduplication marks a localisation relation between a *landmark* (an anchoring entity) and a *trajector* (an entity to be located) (Talmy 1983, 1985, Langacker 1987). The landmark corresponds to the reduplicated N. The trajectory is generally located within the boundaries of the landmark. We are thus dealing with a relationship of internal localisation (Aurnague 1996, 1997).

To illustrate, the trajector in the following sentences (the bird in (34) and Peter in (35)) is located within the boundaries of the landmark, the tree and

the market respectively.

- (34) L' aceddru è/svulazza arvulu-arvulu DET bird be.3SG/flutter.3SG tree-tree 'The bird is in the tree/ flutters in the tree'
- (35) Petru era miccatu-miccatu Peter be-IPFV.3SG market-market 'Peter was in the market/went round the market'

When the landmark is designed as one-dimensional, reduplication is likely to mark a carrier relationship, similar to that expressed by the Fr. preposition *sur* / Eng. *on* (Vandeloise 1986, chap. XI).

(36) L' aceddru è ramu-ramu DET bird be.3SG branch-branch 'The bird is on the branch/ moves on the branch'

When a one-dimensional landmark is within the scope of a verb of motion, the semantic value of reduplication approximates that of a complex preposition such as Fr. *le long de* / It. *lungo* (it.).

(37) Caminanu binariu-binariu walk.3PL rail-rail'They walked along the railway'

In this case, the localisation relation is no longer internal: the trajectory and the landmark are disjoint. These first observations are sufficient to show the polysemic character of NN > Adv reduplication in Sicilian. Below, we distinguish the different semantic values available to this process.

4.3.1 Dynamic localisation

In the first configuration, the trajector is a mobile entity which moves within the landmark. Following Borillo (1998: 37ff), we define this as a dynamic spatial relationship involving a change of location (rather than a change of place⁵). In (34-36), for example, the bird and Peter are not only

⁵ A movement involving a change of location remains within the boundaries of the landmark (e.g. *Peter runs in the garden*). A change of place crosses the boundary between two different landmarks (e.g. *Peter goes out in the street*).

located relative to the landmark, but also seen as moving within the limits of the tree, the market, the branch or the rails. Under these conditions, the use of a stative predicate is generally prohibited, as in (38).

(38) ²Petru rormi sempi casa-casa Peter sleep.3SG always house-house 'Peter always sleeps all over the house'

This shows how the presence of an animate subject and a reduplicative form bring a dynamic interpretation that conflicts with a stative verb like *ròrmiri* ('to sleep'). The only way to interpret this utterance is to multiply the sleeping activity and attribute different locations to different occurrences.

The dynamic value of the reduplicated locative is confirmed by an additional argument. In Italian, the preposition *per* marks dynamic internal localisation (there is no equivalent in English or French that carries the same meaning). Compare (39.a) with (39.b).

(39) a. Il bambino gioca per casa DET child play.3SG for house 'The child plays in various locations in the house/ all over the house'
b. Il bambino gioca a/in casa

DET child play.3SG at/in house 'The child plays in the house/at home'

In (39.b), the preposition indicates only localisation without movement, as in French/English.

In Sicilian, the dynamic interpretation which involves the children's movement in the location *casa* ('house') arises only with the reduplicative structure *casa-casa*.

(40) U picciriddro ioca casa-casa DET child play.3SG house-house 'The child plays at home, all over the house'

In order to specify only the location of the subject within the landmark in Sicilian, a preposition should be used, as in (41)

(41) U picciriddro ioca n casa DET child play.3SG PREP house 'The child is playing in the house'

4.3.2 Static localisation

The second option involves static localisation. However, there are three interpretations that are likely to accompany the localisation relation, depending on the nature of the landmark, the trajector and the relations between them. These readings are not mutually exclusive and several interpretations may simultaneously arise in the same context.

4.3.2.1 Vague localisation One interpretation provides vague localisation.

- (42) Petru è ciumi-ciumi Peter be.3SG river-river 'Peter is (somewhere) in the river / not far in the river'
- (43) Petru si cuccao voscu-voscu Peter REFL lie down.PST.3SG forest-forest 'Peter is lying down (somewhere) in the forest'

In (42), the landmark can be one-dimensional (along the river). But it differs from (37) by the presence of a stative verb (be, as opposed to *walk* in (37)). The idea of movement is present in (37), but disappears in (42) in favor of a single vague localization: Peter is somewhere by the river. The reduplication in (43) implies that the speaker does not know the exact location of the trajector (Peter), although it remains within the limits of the landmark (the forest).

4.3.2.2 Incongruity

A second interpretation highlights the inappropriate nature of the trajector's location.

(44) Petro si fici na casa muntagna-muntagna Peter REFL do.PST.3SG DET house mountain-mountain 'Peter built his house somewhere on the mountain'

There are two simultaneous interpretations in this example: (i) vague localisation, as we do not know the precise location of the house, and (ii) incongruity, as the location chosen for the house is considered inappropriate. Other examples illustrate the incongruity interpretation.

(45) A pianta è scala-scala
 DET plant be.3SG stair-stair
 'The plant is in the middle of the staircase (in the way)'

In (45), the inappropriateness of the trajector's presence on the stairs may result from an incongruity between the function associated with the staircase (allow passage) and the presence of the plant. However, this interpretation remains in examples where the presence of the trajector is natural and should not pose particular problems (46).

(46) Petro è assittato casa-casa Peter be.3SG sitting house-house 'Peter is sitting in the middle of the house'

Thus, example (46) is only acceptable if Peter is in the way, blocking the passage.

4.3.2.3 Homogenous distribution

A final semantic value related to static localisation involves utterances in which the trajector is an inanimate mass noun (47)-(48) or a plural inanimate count (49). In this case, localisation is interpreted as homogenous distribution of the trajector on the landmark.

- (47) A rrina è casa-casa DET sand be.3SG house-house 'The sand is all over the house'
- (48) U fangu è muru-muru DET mud be.3SG wall-wall 'The mud is (spattered) all over the wall'
- (49) I chiova sunnu muru-muru DET.PL nail.PL be.3PL wall-wall 'The nails are all over the wall'

This third interpretation is explained by the non-delimited character of the mass and plural trajectors, or more precisely, by the fact that they are not intrinsically delimited. In this, they contrast with count nouns, which are conceived of as mobile (4.1), vaguely localized (4.2.1) or even incongruously localized (4.2.2). As count nouns, such trajectors have intrinsic boundaries and thus cannot be homogenously distributed without losing their wholeness (at least in the singular). Example (50) illustrated a

reading akin to *along* with a dynamic sense (provided by the motion verb). When the trajector is a mass or plural noun and the verb is stative, the "along" reading denotes static localization.

- (50) U sangu è corda-corda DET blood be.3SG wire-wire 'The blood is spread along the wire'
- (51) L' aceddri sunnu corda-corda DET bird.PL be.3PL wire-wire The birds are sitting along the wire'

5 Conclusion

This article proposes a classification of reduplicative structures in Sicilian and provided a detailed analysis of one of the most productive of these structures, which forms adverbs through complete reduplication of nominal lexemes or word-forms (NN_{adv}). We have argued for a morphological rather than a syntactic analysis of this construction, and presented the variety of semantic values with which it is associated, by highlighting the relationship of internal location which they share, as well as the polysemic values (dynamic vs. static location). This study therefore opens a typological perspective in the study of morphological reduplication and provides new empirical data to the semantic study of spatiality. Our new examination of reduplicative constructions reveals the specificity of Sicilian, with respect to Italian for example, and has thus contributed to its consideration as a language in its own right. We hope to have opened the door to new research on this language, research which is largely non-existent today, within this type of approach. The collection of data and the establishment of a corpus seem to be a necessary first step in this process.

References

- Aboh, E., Smith, N., Zribi-Hertz, A. (eds). 2012. The morphosyntax of reiteration in creole and non-creole languages, John Benjamins Publishing Company.
- Amenta, L. 2010. "La reduplicazione sintattica in siciliano" *Bollettino del Centro di studi filologici e linguistici siciliani*, 22, 345-358.

- Anderson, S. R. 1992. A-Morphous Morphology, Cambridge, Cambridge University Press.
- Aronoff, M. 1976. *Word Formation in Generative Grammar*. Massachussetts: The MIT Press.
- Aronoff, M. & Fudeman, K. 2004. What is Morphology?, Blackwell Publishing-Fundamentals of Linguistics.
- Aurnague, M. 1996. "Les noms de localisation interne: Tentative de caractérisation sémantique à partir de données du basque et du français." *Cahiers de Lexicologie*, 69.
- Aurnague, M., Vieu L., & Borillo A. 1997. "La représentation formelle des concepts spatiaux dans la langue." in M. Denis (ed.), Langage et cognition spatiale. Paris: Masson.
- Bauer, L. 2003. Introducing linguistic morphology (2nd ed.). Washington, D.C.: Georgetown University Press.
- Beccaria, G. L. 2004. Dizionario di linguistica e di filologia, metrica, retorica, Torino, Einaudi.
- Booij, G. 2010. Construction Morphology. Oxford: Oxford University Press
- Borillo, A. 1998. L'espace et son expression en français, Paris, Ophrys.
- Caracausi, G. 1977. "Ancora sul tipo 'camminare riva riva'." *Bollettino del Centro di Studi Filologici e Linguistici Siciliani, 13*, 383-396.
- Floričič, F. 2011. "On Reduplicated Imperatives in Sardinian", 8th Mediterranean Morphology Meeting, Cagliari (14-17 september 2011)
- Forza, F. (2011) Doubling as a Sign of Morphology: A Typological Perspective, *Journal of Universal Language* 12-2, 7-44.
- Fradin, B. 2003. Nouvelles approches en morphologie, Paris, Puf.
- Gil, D. 2005. "From Repetition to Reduplication in Riau Indonesian." *Studies on Reduplication*, B. Hurch (ed.), Oxford: Oxford University Press. pp. 31-64.

- Gleason, H.A. 1955. *Wordbook in descriptive linguistics*. New York: Holt, Rinehart and Winston (Revised edition: 1967).
- Haspelmath, M. 2002. Understanding Morphology, London, Arnold.
- Inkelas, S., Zoll, C. 2005. *Reduplication: Doubling in Morphology*, with Cheryl Zoll, Cambridge: Cambridge University Press.
- Katamba, F. 1993. Morphology, Palgrave.
- Kleiber G. 2009. "Remarques sur la sémantique du nom espace", *SCOLIA*, n. 24, 9-22.
- Kouwenberg, S. (ed) 2003. *Twice as meaninful. Reduplication in Pidgins, Creoles and other contact languages*, London: Battelerbidge Press.
- Langacker, R. W. 1987. Foundations of Cognitive Grammar: Theoretical Prerequisites (Stanford University Press ed. Vol. 1). Stanford: Stanford University Press.
- Lieber, R. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory*. The University of Chicago Press, Chicago/London.
- Leone, A. 1995. Profilo di sintassi siciliana, *Materiali e ricerche dell'Atlante Linguistico della Sicilia*, *3*, Palermo: CSFLS.
- McCarthy, J. J. & Prince A. 1990. "Prosodic morphology and templatic morphology." *in* M. Eid and J. McCarthy, (eds.), pp. 1–54.
- 1995. "Faithfulness and reduplicative identity." in University of Massachusetts Occasional Papers in *Linguistics 18: Papers in Optimality Theory*. Jill Beckman, Suzanne Urbanczyk and Laura Walsh Dickey (eds). pp. 249–384.
- Marantz, A. 1982. "Re reduplication", *Linguistic Inquiry*, Vol. 13, No. 3 (Summer, 1982), pp. 435-482. MIT Press.
- Mel'čuk, I. 1996-1997. *Cours de morphologie générale*, 5 volumes, Montréal, Presses de l'Université de Montrésal-CNRS Editions.
- Moravcsik, E. A. 1978. "Reduplicative Constructions" in Greenberg, J. H., et

al. Universals of Human Language. Volume 3: *Word Structure*, Stanford University Press.

- Rohlfs, G. 1969. *Grammatica storica dell'italiano e dei suoi dialetti*, vol. II, *Sintassi e formazione delle parole*. Torino: Einaudi.
- Scalise, S. & Bisetto, A. 2008. La struttura delle parole, Il Mulino.
- Sgarioto, L. 2005. "'Caminari riva riva': su un fenomeno di reduplicazione nominale in siciliano", *Quaderni di Lavoro dell'ASIS*, *5*, 36-49, Padova.
- Spencer, A. 1991. Morphological Theory, Oxford & Cambridge, Blackwell.
- Talmy, L. 1983. "How language structures space". In Herbert Pick and Linda Acredolo, eds., *Spatial orientation: Theory, research, and application* 225–282. New York: Plenum Press.
- 1985. "Lexicalization patterns: Semantic structure in lexical forms". In Timothy Shopen, ed., *Language typology and syntactic description*. Vol.3 Grammatical categories and the lexicon 36–149. Cambridge: Cambridge University Press.
- Thornton, A. M. 2009. "Italian verb reduplication between syntax and the lexicon", *Italian Journal of Linguistics/Rivista di linguistica*, 21/1, 2009, pp. 235-261.

Vandeloise C. 1986. L'espace en français, Paris, Éditions du Seuil.

Wiltshire, C. & Marantz, A. 2000. "Reduplication", Formale Prozesse57, VIII, 557-567. in Morphologie / Morphology. An International Handbook on Inflection and Word formation, Geert Booij, Christian Lehman and Joachim Mugdan (eds.) (2000/2004). Berlin: de Gruyter.).

A CONSTRUCTIONIST ACCOUNT OF THE EMERGENCE OF A NEW LEXEME-FORMATION PATTERN: ITALIAN RIGHT-HEADED VERBAL COMPOUNDS

Claudio Iacobini University of Salerno

Abstract

Right-headed verbal compounds (e.g. *teleriscaldare* 'to supply district heating', *termovalorizzare* 'to extract thermal energy from waste', *videochiamare* 'to videocall') are a construction arising from scientific and technical registers spreading into current Italian.

Besides describing the characteristics and the origins of these verbs, the paper focuses on the conditions that may favor the emergence of a new lexemeformation pattern. The theoretical framework of Construction Morphology provides an adequate explanation for the passage from the formation of a new kind of complex lexeme (through reanalysis or analogical extension) to the establishment of a productive lexeme-formation pattern, and it also allows a proper collocation to be given for compounding patterns whose head position differs from the canonical one.

1 Introduction

The emergence in current Italian (as well as in the other Romance languages) of a new and in some respects unexpected lexeme-formation pattern, i.e. verbal compounds such as *biocoltivare* 'to grow according to environmentally friendly methods', *termovalorizzare* 'to extract thermal energy from waste', leads us to consider the conditions that favour the emergence of a lexeme-formation pattern. In particular, it makes us question the factors underlying the establishment of this new pattern with respect to the conditions which bring about the formation of the first instances of new complex lexemes through phenomena of reanalysis or analogical extension. The paper is organized in three parts. In the first part (§2) the specific features of Italian right-headed verbal compounds (henceforth RHVCs) are presented. The second part (§3) deals with the origin of this construction, briefly discussing the formative processes, the favouring conditions and the

enhancing factors that underlie its emergence and use. Part three (§4) takes

Construction Morphology as the theoretical framework in order to analyze the passage from local analogical formation to the productive lexemeformation pattern. The conclusions (§5) focus on the restrictions and the spread in usage of RHVCs.

2 Characteristics of Italian right-headed verbal compounds

Starting from the second half of the twentieth century, and increasingly in the last thirty years, the use of verbs such as those listed in (1) has appeared in the Italian language (cf. Iacobini 2013).

(1) biocoltivare 'to grow according to environmentally friendly methods', crioconservare 'to cryo-preserve', elitrasportare 'to transport by helicopter', fotosegnalare 'to photo-signal', geolocalizzare 'to geolocalize', radiotrasmettere 'to radiobroadcast', teletrasportare 'to teleport', televendere 'to telesale', termoregolare 'to thermoregulate', termovalorizzare 'to extract thermal energy from waste', videochiamare 'to videocall', videoregistrare 'to videorecord'.

It is estimated that there are about two hundred such verbs. Apart from the indepth corpus-based analysis on the French language by Namer (2012), these verbs have not yet been thoroughly investigated, even though they can be documented in other Romance languages (cf. some Catalan, French and Spanish examples in 2).

 (2) Cat. crioconservar, helitransportar, teleprogramar, teleportar.
 Fr. aérofreiner, hydromasser, photosculpter, thermoréagir.
 Sp. bioestimular, criopreservar, geolocalizar, teletrasportar, turboalimentar.

As is evident from the examples in (2), these verbs differ in some respects from those inherited by Romance languages from Late Latin (e.g. Lat. *manūtenēre* 'to have tangible evidence or personal knowledge of, to know for certain' > Sp. *mantener*, Fr. *maintenir*, It. *mantenere* 'to maintain, to support'), or formed during the Middle Ages and Renaissance periods; see in (3) some examples taken from the hundreds of verbal compounds collected by Klingebiel (1989).

(3) M.Fr. *saupoudrer* 'to sprinkle (with salt) from *sel* 'salt' and *poudrer* 'to powder'.

Fr. *chantourner* 'to meander; to whipsaw' from *chant* 'edge' and *tourner* 'to turn'.

Gase. *bouque-barrà* 'to remain silent' from *boca* 'mouth' and *barrar* 'to squeeze, to tighten'.

Romance medieval verbal compounds were formed with patrimonial lexical elements of common usage, mostly connected to domains of experience of everyday life.

In the sixteenth century, the productivity of this pattern and the number of these verbs started to decrease. Nowadays, they occur to a varying extent in all modern Romance languages, especially in Catalan (e.g. *captrencar* 'to strike, hit in the head') and in Occitan varieties (e.g. *caplevar* 'to tumble, seesaw, raise the hand or head'), cf., among others, Gràcia & Fullana (1999) and Padrosa Trias (2007). The formation of these compounds is still considered to be productive (albeit declining) in Catalan only, while in the other Romance languages the presence of these compounds (reduced in the current usage to a handful of verbal items) is considered a legacy of previous phases (cf. Klingebiel 1989 for a diachronic account of their diffusion in Western Romance languages).

The "new" verbal compounds display characteristics that set them apart from the "old" compounds, while also differing from the typical and most productive lexeme-formation patterns of current Italian. Contemporary Italian RHVCs can be considered as an exceptional and in some respects unexpected lexeme-formation pattern because of the order of constituents, the syntactic category of the compound, the initial constituent, and the distribution of compounds.

2.1 Order of constituents

The order of constituents is Non-Head/Head, whereas the canonical position of the head in Italian productive endocentric compounds is on the left (e.g. *buono pasto* 'luncheon voucher', *pesce spada* 'sword fish', *treno merci* 'freight train', *vagone letto* 'sleeping car'), cf. Scalise & Fábregas (2010).

2.2 Syntactic category

Productive compound formation rules usually generate nouns; adjectives may also be the outcome of compound rules, although to a somewhat smaller degree. What is more relevant for our discussion is that in the Italian language there is no other way to form a verb through compounding. Even the combination of two verbs (e.g. *usa e getta* lit. use and throw 'disposable') or the reduplication of the same verb (e.g. *fuggi-fuggi* lit. run away-run away 'stampede') result in a noun or an adjective, cf. Thornton (2007); (2008); Masini & Thornton (2008).

Italian complex-verb constructions, such as verb-particle constructions (e.g. *andare via* 'to go away', *mettere dentro* 'to put in', *tirare fuori* 'to pull out') and light-verb constructions (e.g. *dare spazio* 'to give space'), must be kept apart from our discussion, since they differ from RHVCs in both the nature and the position of constituents and the formative patterns: both verb-particle and light-verb constructions originate from phrasal structures and are commonly used in everyday speech; moreover the non-verbal element is right-sided (cf. Iacobini & Masini 2006; Iacobini 2009; Iacobini in press b).

2.3 Initial constituent

The elements normally employed as the initial constituent in RHVCs are not free lexical elements, but combining forms used in compound words to coin technical terms, cf. *termo-* (from Gr. *thermo-*, cf. *thermós* 'heat') in *termometro* 'thermometer', *termovalorizzare* 'to extract thermal energy from waste'. The elements that can be used both as initial constituents in RHVCs and as free lexical forms (e.g. *foto* 'photo, photography' in *fotosegnalare* 'to photo-signal', *tele* 'television' in *televendere* 'to telesale') result from the shortening of neo-classical nominal compounds (*photography, television*), as is shown by their meaning (cf. the different semantics of the "etymological" elements *foto-* from Gr. *phōto-*, cf. *phôs, phōtós* 'light', and *tele-* from Gr. *tēle-*, cf. *têlē* 'far, far off, distant') and by their combinatorial properties, i.e. the possibility to be used as left constituents in right-headed compounds (for a classification of combining forms based on a detailed analysis of a representative corpus of the Italian language, see Iacobini 2004 and Iacobini & Giuliani 2010).

As far as the relation between the two constituents of the compound is concerned, the initial constituent is typically interpreted as an adjunct of the verbal head (e.g. *televendere* 'to telesale') mainly expressing an instrumental meaning, in a few cases it can play an argument role (e.g. *termoregolare* to thermoregulate 'to regulate body temperature') with a patient semantics; it can never play the role of the subject.

2.4 The distribution of compounds

RHVCs are fewer in number and less frequently used than the corresponding nominal and adjectival compounds. In both lexicographic and corpus sources, it is hard to find verbs without corresponding nominal or adjectival derivatives, and most nominal compounds have no corresponding verbs. For example, in the Gradit dictionary only eleven of the approximately two hundred words beginning with *video* are verbs, while there are no verbs corresponding to nouns and adjectives like *biodepurazione* 'biodepuration',

bioproduzione 'organic production', *idroassorbente* 'water-absorbing'. As is generally true for neologisms, RHVCs are mostly used in participial and infinitive forms, present tense, and only occasionally in other verb forms.

3 Origin of Italian right-headed verbal compounds

All RHVCs are very recent, most of them having been coined in the last two or three decades. In this section we show how the origin of this construction can be explained with the convergence of the results of two different processes: back-formation §3.1, and conversion §3.2 (on the relations between back-formation and conversion, cf. Nagano 2007), two non-canonical processes (cf. Corbett 2010) with regard to directionality (cf. Marchand 1963; 1964; Iacobini 2000), transparency and diagrammaticity (cf. Dressler 1987). In §3.3 we illustrate how the spread in usage of nominal and adjectival neo-classical compounds paves the way for the emergence of verbal compounds, and we argue that the presence of inherited patrimonial verbal compounds makes it easier for a new compounding pattern to emerge. Discursive motivation of text cohesion and name-worthiness are mentioned in §3.4 as important factors favouring the coinage and use of RHVCs.

3.1 Back-formation

The starting point for back-formation lies in nominal compounds formed by an initial combining form and a suffixed deverbal noun or adjective (on backformation in Italian, cf. Rainer 2004; D'Achille 2005). The steps that lead to the formation of a verb like *telecomunicare* can be schematically represented as in (4). As the right constituent of the compound in (4a) is a suffixed deverbal noun, it licenses the reinterpretation of the nominal compound in (4b) as derived from a complex verb (4c).

According to Rainer (2004), the back-formation hypothesis for verbs like *telecomunicare* is justified by: i. the combinatorial properties of the initial constituent (a combining form), which mainly combines with nominal or adjectival bases; ii. the distributional implications with respect to nominal compounds (cf. § 2.4); iii. the date of first attestation: nominals strongly tend to precede verbs (see table 1).

telecomposizione	(1979)	telecomporre	(1991)
telecomunicazione	(1911)	telecomunicare	(1955)
telediffusione	(1965)	telediffondere	(1979)
telemisurazione	(1961)	telemisurare	(1990)
telepilotaggio	(1961)	telepilotare	(1987)
teleregolazione	(1979)	teleregolare	(1987)
telericevente	(1965)	telericevere	(1983)

 Table 1: First attestation of words starting with tele- according to Disc dictionary (from Rainer 2004)

The size and regularity of the phenomenon leads Rainer (2004: 497) to believe that in cases such as those reported in Table 1 we are not concerned with the sporadic reversal of a lexeme-formation rule, but rather with "a conventionalized back-formation process, which in turn became a rule [my translation]." One outcome of the institutionalization of a back-formation rule is the extension of the possibility to employ initial combining forms in combination with verbs, thus paving the way for the establishment of a compound formation rule (cf. Shimamura 1983 for an analysis of English compound verbs as the result of backformation, and Kiparsky 1982 who claims that compound verbs should be generated directly by a compound formation rule).

3.2 Conversion

Conversion is based on nominal compounds formed by an initial combining form and a noun. The steps in the formation of a verb according to the conversion hypothesis are reported in (5).

(5) $foto-+copia \rightarrow fotocopia \rightarrow fotocopiare$

The hypothesis of conversion is highly plausible for verbs (e.g. *videoclippare* 'to videoclip', *videomessaggiare* 'to videomessage') in which the corresponding nominal compound is a current word (*videoclip* 'videoclip', *videomessaggio* 'videomessage'), while the suffixed one (°videoclippatore, °videomessaggiatore) is not.

For verbs like *fotocopiare* 'to photocopy', both conversion from the compound noun *fotocopia* 'photocopy' and back-formation from the suffixed noun *fotocopiatrice* 'photocopier, duplicating machine' are plausible formative processes. In cases like this a criterion supporting one or the other hypothesis may be the date of its first attestation. In the case of *fotocopiare*, we can remark from the data reported in Table 2 that the suffixed nominals (possible sources of back-formation) are recorded significantly later than the

verb, and this in turn takes place later than in the compounded noun *fotocopia*.

fotocopia	1917; comp. of foto- and copia, cf. Fr. photocopie 1880
fotocopiare	1952; der. of fotocopia with -are, cf. Fr. photocopier 1907
fotocopiabile	1995; der. of fotocopiare with -bile
fotocopiatore	1983; der. of fotocopiare with -tore
fotocopiatrice	1973; der. of fotocopiare with -trice
fotocopiatura	1980; der. of fotocopiare with -tura

Table 2: Etymon and first attestations of *fotocopia* and related words, according to Gradit dictionary

However, the as yet controversial and often difficult to apply criterion of a word's date of first attestation becomes less and less reliable for the verbs coined in recent years. Indeed, data extracted both from dictionaries and neologism corpora agree in indicating a clear tendency: the date of the first attestation of verbal compounds in the last thirty years tends to coincide with that of the co-radical nominals.

We may therefore assume that the concurrence of the outcome of backformation and conversion laid the foundations for the emergence of a new kind of complex verb that, thanks also to their quantity and diffusion in use, can be interpreted as the result of a compounding schema in which an initial combining form is combined with a verb. This kind of compound can be the base for a suffixation process as is exemplified in (6).

(6) $video + comunicare \rightarrow videocomunicare \rightarrow videocomunicazione$

3.3 Premises for the emergence of right-headed verbal compounds

The most important premise for the emergence of RHVCs is the spread in usage of nominal and adjectival neo-classical compounds, however the presence of patrimonial verbal compounds is a not a trivial factor.

3.3.1 Diffusion in use of nominal neo-classical compounds

The phenomenon that has played a major role in the emergence of RHVCs is certainly the extension in current usage of right-headed nominal and adjectival compounds formed by a combining form and a word (e.g. *cicloraduno* 'cycling rally', *idromassaggio* 'hydromassage', *videochiamata* 'videocall', cf. Iacobini & Thornton 1992; Iacobini 2004; Radimský 2006; Iacobini in press a).

An important difference between Italian native and neo-classical compounds regards derivability: neo-classical compounds may constitute the base for derivation (e.g. *telefono* \rightarrow *telefonista*, *telefonico*, *telefonare*; *psicologia* \rightarrow *psicologismo*, *psicologico*, *psicologizare*), while native compounds usually cannot. The compounds formed by an initial combining form and a word have even more opportunities to be derived, and with a greater variety of suffixes (e.g. *biodegradabile* \rightarrow *biodegradabilità*; *cicloamatore* \rightarrow *cicloamatorismo*, *cicloamatoriale*). They also differ from those which are genuinely Italian due to their greater adaptabilità, *cardio-stimolatore*, *crio-conservazione*, *eco-conservatorismo*).

3.3.2 Diffusion in use of foreign right-headed compounds

Foreign influence on current Italian lexeme-formation must be seen above all as an impulse for the spread of compounding, and secondly, for the spread of new compound types, especially right-headed ones (cf. Iacobini in press a). The presence of foreign compounds in the contemporary Italian lexicon is particularly evident in the great number of non-adapted English compounds. Their dissemination was preceded by the absorption of foreign compounds in the form of calques (e.g. scuolabus from school bus, pubbliche relazioni from public relations). Right-headed nominal compounds borrowed from English (and to a lesser extent from other Germanic languages) have a structure which is similar to that of neo-classical compounds, especially the ones formed by an initial English modifier and an Italian head (e.g. baby-piscina 'baby pool', computer assistito 'computer-assisted', internet-caffè 'internet café', net-azienda 'net-company', sexy-scandalo 'sex scandal', websondaggio 'web poll'). Their presence in the current language favoured the formation of a new type of Italian right-headed nominal compounds with native lexemes (e.g. acquascivolo 'water slide', agopuntura 'acupuncture', cartamodello 'paper pattern', calciomercato 'transfer market', from calcio 'football').

3.3.3 Presence of verbal compounds of common usage

The long presence in the Italian lexicon of verbal compounds of common usage inherited from Late Latin or formed in Italian during the Middle Ages, such as *capovolgere* 'to overturn', *manomettere* 'to tamper', *mantenere* 'to maintain', together with their suffixed derivatives (cf. *capovolgimento* 'reversal, overturning', *manomissione* 'tampering', *mantenimento* 'maintaining, maintenance'), has surely facilitated the identification of a verbal compound inside suffixed nominal compounds, and has induced the emergence of new right-headed verbal compounds.

3.4 Enhancing factors

Besides the above listed favoring conditions, the question on the reasons that motivate the need to use such compound verbs remains open. We can identify two main factors: textual co-reference and name-worthiness.

The spread in usage of RHVCs may depend on discursive motivations of text cohesion: hence on the speaker's need to use a co-radical verb in order to express the process of common activities like *video-calling* or *video-recording*.

Name-worthiness (see Mithun 1984: 848 and the similar concept of "naming strategy" in Booij 2009) is the need to coin compound verbs as names of recognizable activities. The motivation that triggers this lexicalization process is that compounds have a lexical status lacking in their syntactic counterparts.

"Compounding is done for a reason. Some entity, quality, or activity is recognized sufficiently often to be considered name-worthy in its own right; thus Eng. *bus money* or *lunch money* are more likely nominal compounds than *sock money* or *screwdriver money*. [...]. The same is true of verbal compounds, which are coined as names of recognizable activities." (Mithun 1984: 848)

4 From analogical formation to productive lexeme-formation patterns

Apart from their lexicological novelty, the new Italian right-headed complex verbs provide an interesting example of the emergence of a compound pattern in a language with very few verbal compounds and in which productive compounds are typically nominal and left-headed.

It is interesting to note that Classical Latin was not as rich in compound verbs as Sanskrit and Greek, and that Latin compound verbs like *manūtenēre* from which Romance languages developed their own formation in the Medieval and Renaissance periods emerged from similar conditions to those determining contemporary right-headed Italian verbal compounds. For example, Lat. *manūmittěre* was derived through back-formation from the nominal compound *manū missus* 'emancipated' (cf. Klingebiel 1989: 11-41), and the diffusion of the verbal compounding pattern in Late Latin was supported by calques from (the "technical" terminology of) Ecclesiastic Greek (e.g. Lat. *genuflectěre* from Gr. *gonuklineîn*), by native compounds with a right constituent of nominal or adjectival origin (e.g. Lat. *agricultor, altitōnans*), and by participial forms which preceded the original related verbs (e.g. Lat *belligerāns* \rightarrow *belligerāre*).

Moreover, even when (during the Medieval and the Renaissance periods) Noun + Verb became an established and productive compounding pattern, the multiple sources of this construction were still visible: in fact, according to Klingebiel (1989: 122), back-formation from nominal or adjectival bases and other analogical processes accounted for a large part of the complex verbs in each Western Romance variety.

In the previous sections we have shed light on the conditions that led to the emergence of right-headed verbs with an initial combining-form both as regards the relationships of the these new verbal formations with words already existing in the lexicon and with respect to the convergence of the outcomes of the processes of back-formation and conversion.

It is beyond the scope of this paper to try the answer the basic question of whether a categorical distinction has to be drawn between lexeme-formation rules and pattern imitation (cf. Rainer 2013 for an insightful account of the use of the notion of analogy in lexeme-formation theories and descriptions, and the different views about the distinction between the productive application of lexeme-formation rules and analogy).

Here we limit ourselves to pointing out that the passage from the first instances of a new morphological structure based on analogical extension to the establishment of a productive lexeme-formation pattern is hard to explain by an Item-and Process model of morphology, according to which the morphological structure is discrete. Instead, this passage can be profitably interpreted in a theoretical framework which recognizes that morphological structure is intrinsically graded and emerges from the statistical regularities that characterize the forms and meanings of words (cf. Hay & Baayen 2005).

The basic tenets of Construction Morphology (cf. Booij 2010), namely: a) the claim that both individual complex words and abstract generalizing schemas are part of the lexicon; b) the hierarchical organization of the lexicon with layers of subgeneralizations linked through default inheritance; and, above all, c) the paradigmatic relations between (sets of) complex words, form a theoretical framework that makes it possible to explain not only the systematic restrictions (which determine availability of coinage), but also the emergence of new lexeme-formation processes.

The framework of Construction Morphology is based on the idea that morphology is about the pairing of form and meaning in complex words, at various levels of abstractions, from individual complex words to abstract morphological schemas. The native speaker's competence in creating new compounds and derived words is based on abstractions over sets of existing complex words and the words that are paradigmatically related to them (new formation schemas may be also constructed on the basis of paradigmatic relationships among words sharing their stem). The morphological structure assigned to a word is a projection of paradigmatic relationships onto the syntagmatic axis of word structure. Constructions are organized into hierarchies with generalizations on different levels of abstraction (there are intermediate schemas between the individual words and the more abstract lexeme-formation schemas, which express generalizations about subsets of complex words of a certain type, e.g. constructional idioms, partially lexically specified productive patterns with both variable and lexically fixed positions, cf. Pitt & Katz 2000). Moreover constructional schemas can be unified into complex schemas (template conflation) that express the co-occurrence of certain types of lexeme-formation.

The notion of template conflation has been employed by Booij (2007; 2009) to justify the productive Dutch lexeme-formation pattern of the type [NV-suff]_N exemplified in (7) as derived by suffixation from a Noun + Verb compound (a relation that closely resembles that between Italian complex words such as such *termoregolazione* and *termoregolare*).

 (7) brand-bluss-er 'fire extinguisher' rokkenn-naai-ster 'skirts sewer (f.)' hand-oplegg-ing 'hands imposition' bijen-houd-erij 'bee keeping'

The idea of a hierarchical lexicon, with intermediate levels of generalization allows Italian right-headed nominal complex words (8) to be analyzed as compounds without conflicting with the generalization that most Italian compounds are left-headed (cf. Booij 2009).

(8) *audioguida* 'audioguide', *autolinea* 'bus route', autoservizio 'car service', *bioalimento* 'bio food', *cicloamatore* 'cycling enthusiast', *ecomuseo* 'eco museum', *eurozona* 'eurozone', *fotobiografia* 'photobiography', *monouso* 'disposable', *narcotrafficante* 'drug dealer', *paleoindustriale* 'paleoindustrial', *pseudonotizia* 'pseudo news', *psicofarmaco* 'psychotropic drug', *televendita* 'TV sale', *videoconferenza* 'videoconference'.

By specifying that the initial constituent of this class of right-headed compounds is restricted to combining forms (i.e. members of a restricted and definable set of formative elements), this compound pattern can be expressed by assuming a constructional idiom $[CF [x]_V]_V$ which expresses that combining forms can only occur as part of a complex word, and do not determine the syntactic category of the compound of which they are part. Moreover, this class of compounds will be specified as right-headed, and therefore, the constructional idiom will not be linked to the more general node for left-headed compounds in the lexicon of Italian.

On the same grounds, the compound structure that may be assigned to verbs like *teleriscaldare, termovalorizzare, videochiamare* is a consequence both of the instantiation relation and of the part-of relation which are based on paradigmatic relationships between words in the lexicon. The knowledge of existing compounds (both as autonomous words and embedded in more complex constructions) is a precondition for the language user to develop new formative patterns. According to Booij (2009: 205), "people acquire the morphological system of a language, that is, the abstract morphological schemas, on the basis of their knowledge of a set of words that instantiate these patterns. Once they have come across a sufficient number of words of a certain type, they can infer an abstract schema, and will be able to expand the relevant class of words".

5 Conclusions

Although verbs such as *teleriscaldare, termovalorizzare, videochiamare* are still limited in number (and some new formations might still possibly be interpreted as the result of back-formation or conversion processes), they represent quite an important novelty in the morphological system of the Italian language.

We must acknowledge that, despite their recent acceptance, they are strongly limited since they can only employ combining forms as first constituents, and because they are mostly used in scientific and technical registers. For example verbs such as *piattilavare* lit. 'to dish wash' or *dischigirare* lit. 'to disc turn' are completely unacceptable despite current nouns such as *lavapiatti* 'dishwasher', *giradischi* 'record player'.

We cannot predict what the possible evolution may be or what limits there may be in the spreading of the usage of RHVCs. The main aim of this paper was to show how a new lexeme-formation pattern can emerge.

However, the results of Namer's (2012) detailed corpus-based analysis of French RHVCs show that these verbs are not occasional bizarre formations confined to scientific terminology. On the contrary, Namer (2012) clearly shows that, compared to co-radical suffixed compound nouns and adjectives, RHVCs appear to have been adopted by a wider audience and are used in the context of less formal exchanges.

Construction Morphology provides an adequate theoretical framework on how a new lexeme formation pattern can emerge, and it also makes it possible to give a proper collocation for a compound pattern whose head position does not conform to the generalization that most Italian compounds are left-headed.

References

- Booij G. 2007. "Construction morphology and the lexicon" *in* F. Montermini,G. Boyé, N. Hathout (eds), *Selected proceedings of the 5th Décembrettes: Morphology in Toulouse*. Somerville: Cascadilla Press, 34-44.
- 2009. "Compounding and Construction Morphology" in R. Lieber, P. Štekauer (eds), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 201-216.
- 2010. Construction Morphology. Oxford: Oxford University Press.
- Corbett G. 2010. "Canonical derivational morphology". *Word Structure* 2: 141–155.
- D'Achille P. 2005. "Le retroformazioni in italiano" in C. Giovanardi (ed), Lessico e formazione delle parole. Studi offerti a Maurizio Dardano per il suo 70° compleanno. Firenze: Cesati, 75-102.
- DISC 1997. Disc. Dizionario italiano Sabatini-Coletti. Firenze: Giunti.
- Dressler W. U. 1987. "Word Formation as Part of Natural Morphology" in W.U. Dressler, W. Mayerthaler, O. Panagl, W.U. Wurzel (eds), *Leitmotifs* in Natural Morphology. Amsterdam: John Benjamins, 99-126.
- Gràcia L., O. Fullana 1999. "On Catalan verbal compounds". *Probus* 11: 239-261.
- Gradit 1999. *Grande dizionario italiano dell'uso*, ideato e diretto da T. De Mauro. Torino: Utet, 6 voll.
- Grossmann M., F. Rainer 2004 (eds). *La formazione delle parole in italiano*. Tübingen: Niemeyer.
- Hay J. B., R. H. Baayen 2005. "Shifting paradigms: gradient structure in morphology". *Trends in Cognitive Sciences* 9: 342-348.
- Iacobini C. 2000. "Base and direction of derivation" in G. Booij, Ch. Lehmann, J. Mugdan (eds), Morphology. An International Handbook on Inflection and Word Formation, vol. 1. Berlin: De Gruyter, 865-876.

- 2004. "Composizione con elementi neoclassici" in M. Grossmann, F. Rainer (eds), *La formazione delle parole in italiano*. Tübingen: Niemeyer, 69-95.
- 2009. "The role of dialects in the emergence of Italian phrasal verbs". Morphology 19-1: 15-44.
- 2013. "Il tipo "videoregistrare": da retroformazione all'emergere di un nuovo processo compositivo fondato su di un vecchio modello romanzo" *in* J. Born, W. Pöckl (eds), "Wenn die Ränder ins Zentrum drängen... " Außenseiter in der Wortbildung(sforschung). Berlin; Frank & Timme, 189-212
- in press a. "Foreign word-formation in Italian" in O. Müller, I. Ohnheiser,
 S. Olsen, F. Rainer (eds), Word-Formation. An Internation Handbook of the Languages of Europe. Berlin: De Gruyter.
- in press b. "Particle-Verbs in Romance" in O. Müller, I. Ohnheiser, S. Olsen, F. Rainer (eds), Word-Formation. An Internation Handbook of the Languages of Europe. Berlin: De Gruyter.
- Iacobini C., A. Giuliani 2010. "A multidimensional approach to the classification of neoclassical combining forms", *Italian Journal of Linguistics* 22: 287-316.
- Iacobini C., F. Masini 2006. "The emergence of verb-particle constructions in Italian: locative and actional meanings". *Morphology* 16-2: 155-188.
- Iacobini C., A.M. Thornton 1992. "Tendenze nella formazione delle parole nell'italiano del ventesimo secolo" in B. Moretti (ed), *Linee di tendenza dell'italiano contemporaneo*. Roma: Bulzoni, 25-55.
- Kiparsky P. 1982. "Lexical morphology and phonology" *in* I.-S. Yang (ed), *Linguistics in the Morning Calm.* Seoul: Hanshin, 3-91.
- Klingebiel K. 1989. *Noun* + *verb compounding in Western Romance*. Berkeley: University of California Press.

- Marchand H. 1963. "On Content as a Criterion of Derivational Relationship with Back-Derived Words". *Indogermanische Forschungen* 68: 170-75.
- Marchand H. 1964. "A Set of Criteria of Derivational Relationship between Words Unmarked by Derivational Morphemes". *Indogermanische Forschungen* 69: 10-19.
- Masini F., A.M. Thornton 2008. "Italian VeV lexical constructions" in G. Booij, A. Ralli, S. Scalise (eds), Morphology and Dialectology. On-line Proceedings of the 6th Mediterranean Morphology Meeting (MMM6). Patras: University of Patras, 148-189.
- Mithun M. 1984. "The evolution of noun incorporation". *Language* 60: 847-894.
- Nagano A. 2007. "Marchand's analysis of back-formation revisited: back-formation as a type of conversion". *Acta Linguistica Hungarica* 54: 33-72.
- Namer F. 2012 "Nominalisation et composition en français : d'où viennent les verbes composés?". *Lexique* 20: 173-206.
- Padrosa Trias S. 2007. "Catalan Verbal Compounds and the Syntax-Morphology Competition" in F. Montermini, G. Boyé, N. Hathout (eds), Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse. Somerville, MA: Cascadilla, 91-107.
- Pitt D., J. J. Katz 2000. "Compositional idioms". Language 76: 409-432.
- Radimský J. 2006. *Les composés italiens actuels*. Paris: Cellule de recherche en linguistique.
- Rainer F. 2004. "Retroformazione" in M. Grossmann, F. Rainer (eds), La formazione delle parole in italiano. Tübingen: Niemeyer, 493-497.
- Rainer F. 2013. "Formación de palabras y analogías: aspectos diacrónicos" inI. Pujol Payet (ed), Formación de Palabras y Diacronía. A Coruña,Universidade da Coruña (Anexos Revista de Lexicografía, 19), 141-172.

- Scalise S., A. Fábregas 2010. "The head in compounding" in S. Scalise, I. Vogel (eds), Cross-Disciplinary Issues in Compounding. Amsterdam: John Benjamins, 109-126.
- Shimamura R. 1983. "Backformation of English Compound Verbs" in J. F. Richardson, M. Marks, A. Chukerman (eds), *Papers from the parasession* on the interplay of phonology, morphology and syntax. Chicago: Chicago Linguistic Society, 271-282.
- Thornton A. M. 2008. "Italian Verb-Verb reduplicative Action Nouns". *Lingue e linguaggio* 7: 209-232.
- Thornton A. M. 2010. "Il tipo *fuggifuggi*" in M. Iliescu, H. Siller-Runggaldier, P. Danler (eds), *Actes du XXVe CILPR* (Congrès International de Linguistique et de Philologie Romanes), Innsbruck, 3 – 8 septembre 2007. Berlin: De Gruyter, 7-527–7-536.

A CONSTRUCTION-BASED ACCOUNT OF THE RELATION BETWEEN CONVERSION AND INFLECTIONAL CLASS IN MODERN GREEK

Nikos Koutsoukos University of Patras

Abstract

The paper discusses the relation between conversion and inflectional class of the output in Modern Greek within a Construction Morphology framework. Denominal conversion in Modern Greek offers intriguing data for the relevant discussion which are rather unknown hitherto. It is argued that conversion is not related to the default inflectional class of the output verb, but rather with an inflectional class which uniquely characterizes this type of formations. The schematic representation can account for the formal and semantic properties of both the input and the output, among which the inflectional properties of the output is the most significant one.

1. Introduction¹

Construction Morphology (Booij, 2010) is based on the idea that *constructions*, i.e. form-meaning (or function) pairs, are the basic units of the description and analysis of the linguistic phenomena. Construction Morphology (CM) has been developed in a wide variety of morphological phenomena, but it is not a priori clear how this approach can account for phenomena related to the interaction between inflection and derivation.

It has been generally argued that conversion in languages with rich inflection may manifest itself by virtue of a change of the inflectional properties of the output without any change in its phonological or morphological structure. In most cases, converted formations are -in principle- inflected according to the default inflectional class. A classic example of this phenomenon is conversion in Italian (Dressler, 2003). The same is also confirmed with

¹ This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund. I gratefully acknowledge the funding support.

respect to French, where verbs derived from adjectives may be class 1 (infinitive *-er*, past participle *-é*), e.g. *allonger* 'to extend, or class 2 (infinitive *-ir*, past participle *-i*), ex. *aplatir* 'to flatten'. Class 1 is default, whereas class 2 is not.²

The present paper discusses the relation between conversion and inflectional classes in Modern Greek (MGr) and puts the question of whether the notion of construction can be useful in the analysis of the relevant phenomena. The examination of the data entails first taking a position in the difficult question of the grammatical nature of conversion. I start with this problem by examining the different proposals regarding the representation of the process and take a position in favour of a paradigmatic account of conversion.

Conversion in MGr has not been thoroughly discussed yet. I focus on denominal conversion and show that it is mainly productive in compound formations with bound stems and less productive in simple bases. I show that converted formations have many interesting formal and semantic properties and may display interesting semantic correlations between the different parts of their members. Among the formal properties, the most dramatic one is the inflectional class of the output, since the output of the process is not inflected according to the default inflectional class.

Last but not least, I show that CM can account for the formal and semantic properties of the conversion members and the process *per se* can be represented as a set of paradigmatically related schemata. In this respect the assignment of the inflectional properties of the output finds a natural account.

2. Conversion in Modern Greek

2.1 The problem of conversion

Conversion can be defined as the process which changes either the category or (some of) the inherent properties of lexical items without a concomitant change in their form (Booij, 2002), e.g. $[Google]_N > [[google]_N]_V$.³ Conversion

 $^{^2}$ As the reviewer of the paper has noted, in both languages, French and Italian, A>V converts fall into two inflectional classes: some converts show the default conjugation (theme vowel /a/ in Italian, /e/ in French), others belong to a non-default conjugation (theme vowel /i/ in both languages).

³ Clark & Clark (1979) distinguish between denominal verbs, such as *land* and *to land*, and innovative denominal verbs, that is, formations which have a shifting sense and denotation -one that depends on the time, place and circumstances of use. In the latter group, they classify formations which are not well-established but the listener can figure out the meaning of the verb on the basis of the verb itself, the linguistic context and other mutual knowledge. They also argue against a derivational account of the first category in English, since as they claim, in many instances verbs do not

is a classic example of form-meaning asymmetry in morphology and thus it has given rise to a number of theoretical issues including, among others, the questions of (a) whether conversion should be treated as a derivational process and (b) what is the best way to account for the relation between form and meaning in conversion. I will briefly present the theoretical proposals that have been put forward regarding conversion and then I will present the analysis of the MGr data.

The asymmetry between form and meaning in conversion pairs has been discussed in Lieber's (1980, 1981, 2004) seminal work. Lieber was the first to draw a distinctive line between affixational processes and conversion on the basis of the criterion of directionality. In any sort of affixational process, the addition of an affix to a base signals the derivation of a new item and -at the same time- the directionality of the process with respect to formal as well as semantic compositionality. In conversion, however, there is no addition of a discrete, intrinsically meaningful element and, thus, the directionality of the process cannot be determined *a priori*. On the basis of this difference, Lieber argues that conversion cannot be regarded as an affixational process, and, going a step further, that conversion cannot be regarded as a grammatical process, at all. Instead, conversion should be expressed as a redundancy relation in the permanent lexicon. Her definition of conversion reads as follows 1980: 198): 'Conversion would be defined as a relation R such that lexical terminals X and Y satisfy R if and only if they differ only with respect to their category class membership'

As the definition itself implies, this account does not entail a formal representation of conversion. Instead, the creation of a converted item can be ascribed to a copying process in the lexicon. Lieber's RH can be illustrated by the following example of conversion pair in MG (from Ralli, 1988: 147):

(1) Lexical entry 1

Odig(os) 'driver' Lexical category: N^4 $R = [Odig(os)]_N \leftrightarrow [Odig(o)]_v$ Lexical entry 2 Odig(o) 'drive' Lexical category: V

have a corresponding noun, they display semantic idiosyncrasies and they do not contain the meaning of the verb.

⁴ Glossing and abbreviations follow the *Leipzig Glossing Rules* (available at: http://www.eva.mpg.de/lingua/resources/glossing-rules.php). The following abbreviation has been used for the analysis of the data in this paper: N= noun, V= verb, INFL= inflectional marker, CM= compound marker, M=masculine, FEM=feminine, PL=plural, ASP=aspect, IC=inflectional class.

Each member of the conversion pair in (1) has a separate lexical entry, specified individually as to its lexical class and category membership, and a relation R relates the two members of the pair. On this account, neither member of the conversion pair should be considered as basic.

Marchand (1969: 356) describes conversion as follows: 'By derivation by a zero morpheme I understand the use of a word as a determinant in a syntagma whose determinatum is not expressed in phonic form but understood to be present in content, thanks to an association with other syntagmas where the element of the content has its counterpart on the plane of phonic expression'. Marchand compares English derivatives employing the suffix -ize, such as legalize, nationalize and sterilize, with verbs such as clean, dirty and tidy, and observes that the syntactic and semantic properties of the formations are the same in both groups; a verb is derived from an adjective and has the meaning 'render sth <adjective>'. However, in the first group the content element is expressed by the overt morpheme -ize, while in the second group the content element has no counterpart in the phonic expression. Marchand therefore claims that the derivational morpheme is zero-marked in the second group.⁵ The asymmetry between formal and semantic structure in conversion pairs is accounted for by the postulation of a zero affix which changes the category of the base, and the heuristic principle for the analysis of conversion pairs is the comparison of these formations with other derivational pairs displaying the same properties.

An alternative approach to conversion can be found in Don's (1993) work. His model essentially consists of two parts: a *Lexicon* which accounts for the 'paradigmatic' mismatches found across the inventory of morphemes within a specific language, and an *Engine* which takes the form of *Finite State Transducer* (FST). This FST performs the mapping between the formal and semantic level of representation and accounts for the 'syntagmatic mismatches'.

According to Don (1993), conversion constitutes a case of syntagmatic mismatch, since there is an affix at the morphosyntactic level, but there is no phonological material expressing the content of this affix. In Don's (1993: 99-100) analysis, the English converted noun $[walk]_N$ has the following representation:

⁵ The basic criterion for the recognition of zero derivational relations has been the existence of appropriate analogues involving overt morphological marking of the same derivational function, the so-called 'overt analogue criterion' (see Sanders, 1988).


The representation in (2) shares many features with the traditional affix-based models: for example, at the morphosyntactic level, the affix is considered as the head of the converted word. However, it differs from them, in that it assumes a second level of abstraction for the morphosyntactic properties of the affixes (represented in capital letters in (2). According to Don, a basic advantage of assuming two levels of representation lies in the fact that it allows for the existence of a more complex form at the morphosyntactic level, but a simpler form at the morphophonological level, and thus gives a more straightforward explanation for the asymmetry found in cases of conversion.

The main advantage of Don's analysis is that he convincingly argues for the directionality of conversion. Crucial evidence for the directionality of a derivational process can be adduced by the examination of the lexical-class properties of the output. According to the criterion of uniformity of the output-class, if the outputs of conversion always fall into the same lexical class, conversion should be considered as a type of affixational process.

Unlike classic constructive (either affix-based or rule-based) models which assume that individual forms are derived in isolation from other forms in a grammatical system, paradigmatic models assume that derivation can be conceived of as a set of paradigmatic relations. In this view, conversion can be interpreted as the correlation between members of word sets which have the same degree of morphological complexity, but differ with respect to their meaning or their morphosyntactic properties. This correlation is established by the speakers of a language on the basis of the linguistic evidence available to them, and can be considered as the *locus of interpretation* of the properties of the converted elements. An elaborated paradigmatic account of conversion can be found in Booij (1997).

Booij examines the relation between conversion and gender assignment in Dutch and shows that, although in many cases the gender value of a complex noun in Dutch is determined by one of its constituents, there are also several cases in which the gender value cannot be predicted this way. Gender assignment correlates with the formal complexity of the verbal base. In cases such as *raad* 'to advise', in which a simple verb is converted into a noun, the

gender value is [common], while when the same verbal base is prefixed, such as *beraad* 'to deliberate', the conversion results in nouns with the gender value [neuter]. In this respect, the gender value of the noun cannot be interpreted as the contribution of a particular morphotactic unit of the morphological structure. Instead, the gender value of the converted nouns is predicted only by making use of information about the corresponding verb.

Based on the assumption that the relation between the two items is the *locus of interpretation* of the properties of the converted noun, the systematic difference in gender can be considered as part of this relation. A representation of this paradigmatic relation reads as follows:

- $(3) \quad a. \quad {<}V_i{>} \ \approx {<}[V_i]_{N \ [\text{COMMON GENDER}]}{>}$
 - b. <[prefix-V_i] $> \approx <$ [prefix-V_i]_{N [NEUTER GENDER]}>

Words are form-meaning pairs and the symbols $\langle \rangle$ demarcate the whole construction. The schemas in (3) represent the correlation between conversion pairs -i.e. pairs of words that have the same phonological make-up, but differ with respect to their meaning and morphosyntactic properties.

Thus far, I have presented some models for the representation of the grammatical properties of conversion. Marchand's zero-affixation approach solves the problem of the asymmetry between form and meaning by introducing zero morphemes. However, this is a rather disputable solution in the analysis of conversion. On the basis of data from Greek, Ralli (1988, 2005) argues that we can assume zero morphemes in inflection but not so often in derivation. Similarly, among others, Booij (2002) and Lieber (1980) claim that the positing of zero affixation as a derivational process raises a number of problems. In many languages, such as English, the bases that form input to conversion processes do not fall into a uniform lexical class, and a proliferation of zero morphemes for the different categories of the bases is thus an unavoidable consequence of such a move. Moreover, there is an inherent difficulty in defining the formal and semantic properties of zero morphemes, since there is no independent evidence for their combinability properties, their position in the formation (prefix or suffix), or their selectional properties. In most cases, the sole rationale that would motivate this choice would be to force the morphology to fit the theoretical position that every morphological structure approaches the ideal of one-to-one correspondence between form and meaning.

In what follows, I compare these models and argue that an analysis based on paradigmatic relations is to be preferred. Conversion pairs display a clear directionality in terms of semantic compositionality and semantics along with the formal properties of structures can serve as a valuable test for the formal account of conversion pairs. A paradigmatic account of the relation between the members of the conversion pairs can solve the problems of the zero derivational analysis. However, it can be likened to the Relisting Hypothesis in that it relates existing words with a redundancy rule and it does represent the morphological creativity. In the following section, I discuss conversion pairs in MGr and argue that paradigmatic relations can prove useful in the discussion of conversion, but one should go beyond the relation between existing words.

2.2 Conversion in Modern Greek

Thus far, I have addressed the question of the grammatical nature of conversion. I have also discussed the possible alternatives regarding the formal representation of the process. In this section, I look in more detail at the MGr data. I aim to show that conversion should be analysed as a set of paradigmatic related constructions established in the lexicon and to form the basis for the discussion of the relation between conversion and inflectional class of the verb.

Conversion in MGr changes nominal bases, either simple bases (4a) or compound formations (4b) into verbs:

(4)	a.	$[[pygmax]_{N}-os]^{6} >$	[[pygmax] _v -o]
		STEM-INFL	STEM-INFL
		'boxer'	'perform the activity of a boxer'
	b.	[[arthr-o-graf] _N -os] > STEM-CM-STEM- INFL 'columnist'	[[arthr-o-graf] _v -o] STEM-CM-STEM-INFL 'perform the activity of a columnist'

In the data above we notice that there is a change in the category of the base without a concomitant change in the formal make-up. Moreover, it should be underlined that inflection does not participate in the derivational process and cannot change the category of the base.

Van Marle (1985: 161) has argued that 'conversion on the basis of a complex starting-point is by far the most exceptional: "normally" conversion takes the simplex words of a word-class as its starting point', while Aronoff (1980: 747 ft 2) claims that 'it is well known that the rule is restricted to monomorphemic nouns". The Greek data do not verify this claim, since

⁶ Data and examples are given in the citation form, i.e. first singular present for verbs and nominative singular for nouns.

denominal conversion is a process mainly productive in compound formations.

Conversion in MGr does not apply to all types of compounds, but it is predominantly productive in one specific pattern, i.e. compounds with a bound stem⁷ as their second constituent. Consider the following data:

Noun	Verb
$[[arthr-o-gráf]_{N}-os]_{N}^{8}$	$[[[arthr-o-graf]_N]_V-ó]_V$
STEM-CM-BOUND STEM(N)-INFL	STEM-CM-BOUND STEM-INFL
'columnist'	'perform the activity of a columnist'
$[[gloss-o-lóg]_{N}-os]_{N}$	$[[[gloss-o-log]_N]_V-\acute{O}]_V$
STEM-CM-BOUND STEM(N)-INFL	STEM-CM-BOUND STEM-INFL
'linguist'	'perform the activity of a linguist'
$[[\text{theoritik-o-lóg}]_{N} - \text{os}]_{N}$	[[[theoritik-o-log] _N] _V -ó] _V
STEM-CM-BOUND STEM(N)-INFL	STEM-CM-BOUND STEM-INFL
'theoretician'	'perform the activity of a theoretician'

 Table 1: Converted compounds in Modern Greek

In table (1) verbs have the same morphological make-up as their corresponding nouns, but they differ from them with respect to their semantic compositionality.

Let us now examine the formal and semantic properties of these formations.⁹ Ralli (2008, 2013) argues that conversion in compound formations with bound stems displays a clear direction of derivation; the nominal formations are to be considered as the input to conversion, whereas the verbal formations are to be seen as the result of the process.

As shown by Ralli, robust evidence comes from the historical development of these formations: as a matter of fact, the nominal formations always precede the corresponding verbal formations during the history of Greek. It should be mentioned however, that nowadays there are some verbs without a corresponding noun. For example, the verb *pliktrologo* 'to type' does not have a corresponding noun *pliktrologos* 'typist'.

⁷ The interested reader may read Ralli (2008, 2013) for a detailed discussion of the grammatical characteristics of compound formations with bound stems in Greek.

⁸ According to the rules of MGr orthography, the stress mark is obligatory. In the data under discussion I indicate the stress mark since it is of particular relevance to our analysis.

⁹ The analysis proposed for the compounds is equally applicable to simple verbs. However, since compound formations are more productive, in the rest of the paper I will focus on these formations.

Corroboration for the directionality of the process can also be found in the accentual pattern of the verbal stems. In every verbal formation there is a movement of the stress towards the last syllable of the verb, e.g. $[glossológos]_N$ 'linguist' vs $[glossologó]_V$ 'perform the activity of a linguist'. Since this stress pattern is very systematic in conversion pairs, one may assume that the derivational process is expressed by a change in suprasegmental information, i.e. the stress pattern.¹⁰ Both criteria prove useful in determining the direction of derivation in conversion pairs with bound stems.¹¹

A notable property of these converted verbs is the semantic compositionality of the formations.¹² As shown in table (1), the verb always contains the meaning of the noun, not vice versa. For example, the verbal formation [glossologó]_V has the interpretation 'perform the activity of a [glossológos]_N 'linguist'. Therefore, we can assume that the meaning of the verb is defined on the basis of the meaning of the corresponding noun. Some verbal formations display a kind of semantic idiosyncrasy and they do not fit into this pattern. For example, the verb *kinimatografó* 'to film' is not connected to the corresponding noun *kinimatográfos* 'cimena'. Moreover, the semantic relation between the members of the conversion pairs with simple bases cannot be defined without avoiding circularity in the interpretation, e.g. *odigos* 'driver' versus *odigo* 'drive'. The verb does not contain the meaning of the noun. Thus, this relation cannot be considered as criterial for cases of conversion pairs with simple bases.

The semantics of the examples in Table (1) is also interesting from a broader perspective, since the meaning 'to work as N' is not freely available in many other languages. Take for example the English nouns *linguist*, *policeman* and *professor* which cannot form corresponding verb by conversion **to linguist*, **to policeman*, **to professor*.

Denominal conversion in MGr is not without restrictions. It is a commonplace in morphological theory that derivational processes may be sensitive to semantic restrictions imposed by the base. Huning (2009) argues

 ¹⁰ A similar change in the stress pattern can be observed in a limited number of conversion pairs in English (see, principally, Kiparsky, 1982; Marchand, 1969).
 ¹¹ Conversion pairs with simple bases display the same formal and semantic

¹¹ Conversion pairs with simple bases display the same formal and semantic characteristics as compound formations with bound stems: there is an asymmetry between form and meaning, since verbs on the right-hand column display a change in the category of the base without an overt morphological marking and verbs in these conversion pairs follow the stress pattern of the verbal compounds with bound stems (stress on the final syllable) (cf. Koutsoukos, 2013). ¹² Semantic compositionality has been considered as a questionable criterion for the

¹² Semantic compositionality has been considered as a questionable criterion for the directionality of conversion (see the discussion in Tribout, 2010).

that word-formation processes often develop *semantic niches*, i.e. groups of words (subsets of a morphological category) kept together by formal and semantic criteria. Conversion in MGr displays a similar kind of semantic fragmentation; that is, conversion is not applied blindly to all nominal formations. The following examples are illustrative:

Nominal formation	Verbal formation
[+animate agent]	
arthrográf-os	arthrograf-ó
'columnist'	'perform the activity of a columnist'
glossológ-os	glossolog-ó
'linguist'	'perform the activity of a linguist'
Nominal formation	Verbal formation
[-animate agent]	
tomográf-os	*tomograf-í 3sg ¹³
'device for	'produce an image of the inside of the human
tomography'	body or a solid object using X-rays or ultrasound'
logográf-os	*logograf-í 3sg 'perform recordings'
'device for	
recordings'	

Table 2: Selectional restrictions on conversion

The data in table (2) imply that nominal compound formations with the semantic feature [-animate agent] cannot be subject to conversion. All these properties will be discussed in section 4.

3. Conversion and inflectional classes in Modern Greek

3.1 Conversion and typological features of languages

It has been claimed that there is a direct connection between the appearance of conversion and the type of inflectional system of the language; that is, conversion is usually assumed as a common phenomenon only in languages with weak (or restricted) inflectional systems. As an illustrative example, it has been claimed that in English the loss of inflectional markers gave rise to derivation by means of zero morpheme. Jespersen (from Marchand, 1969: 363) claimed that 'as a great many native nouns and verbs had...come to be identical in form..., as the same things happened with numerous originally French words'.

¹³ I use 3sG person instead of the standard citation form, i.e. 1sG, since I refer to devices.

Marchand casts doubt on this claim and argues that conversion existed in English, when English was still a more amply inflected language and inflectional differences were more in evidence. As he puts it: 'Derivation by a zero-morpheme is neither specifically English nor does it start, as Jespersen's presentation would make it appear, when most endings had disappeared'.

On synchronic grounds, although it cannot be denied that conversion is very productive in languages with restricted inflection, this is not necessarily the case, since conversion can be productive also in languages with strong (or rich) inflection.

Manova (2011) provides an analysis of conversion in Slavic languages with special focus on Bulgarian, Russian, and Serbo-Croatian and she claims that conversion can form a cline from the most to the least prototypical cases. According to Manova conversion in Slavic languages may apply to words which already display an inflectional marking in that case the inflectional marker is substituted or deleted. In other words, the derivational process of conversion can change the category (or some of the morphosyntactic properties of the base), *even after* the inflectional marking. However, this should not be considered as prototypical conversion. The prototypical case of conversion does not involve any intervention of inflection. The following examples should be considered as less prototypical cases of conversion:

(5) Russian: učitel' 'teacher' > učitel'-it' 'work as a teacher (colloq.)'

In the data above we observe that conversion results in change in the wordclass and, thus, resembles very much the English examples. However, the addition in the inflectional slot of the Russian derivative renders this type of conversion less prototypical than the English one. Even less prototypical should be considered data from Russian which do not display a change in the category of the input (Manova, 2011: 61):

(6) Russian: $[[matematik]-a]_N$ 'mathematics'> $[matematik]_N$ 'mathematician'

According to Manova, both the input and the output of the conversion process presented in (6) have the same category and are semantically related. However, they differ with respect to their inflectional paradigms. For such cases, where in a derivational conversion there is no word-class change, Manova uses the term non-prototypical conversion.

With respect to MGr, as mentioned earlier, denominal conversion is productive only within a limited domain of constructions, i.e. compounds with bound stems. Conversion is not very productive in simple bases. The critical question is what happens with the relation between conversion and inflection in languages with strong inflection. In other words, should we claim that inflection has derivational properties? In these cases, we should not consider inflectional endings as derivational elements. As claimed by Marchand (1969: 363), 'stems are immediate elements for the speaker who is aware of the syntagmatic character of an inflected form'. Derivational processes apply to stems (before the addition of the inflectional material) and may have the potential to determine the inflectional properties of the output. For example, Aronoff (1994) has shown that sometimes the assignment of an abstract morphological property will be the only morphological effect of a derivational process; the clearest and most dramatic among the abstract morphological properties that may be assigned by a derivational process is inflectional class (Aronoff, 1994: 127). Inflectional marking takes place at a later stage in order to express the morpholoyntactic properties of the words.

3.2 Conversion and inflectional classes in Modern Greek

It has been argued that conversion may be connected with specific inflectional properties of the output. Thornton (2004: 503) argues that converted formations are inflected according to the default inflectional class. The formal correlate between conversion and inflectional properties of the output can be more evident in languages with rich inflection.

A classic example comes from Italian. In Italian almost each word belongs to an inflectional class expressed by the endings in (nearly) all paradigm slots (Gardani, 2009: 97). Nouns formed by conversion are inflected according to two maximally productive subclasses (Dressler, 2003):

(7)	(a) degrad-are 'degrade'	il degrad-o 'degradation'	i degrad-i (Italian)
	STEM(V)-INFL	STEM(N).M-INFL	STEM(N).M-INFL.PL
	(b) revoc-are 'revoke'	la revoc-a 'revocation'	le revoch-e (Italian)
	STEM(V)-INFL	STEM(N).FEM-INFL	STEM(N).M-INFL.PL

In order to examine the relation between conversion and inflectional class of the output the first task is to outline the verbal system in MGr.

The verbal inflectional system of MGr is organized around two major inflectional classes. The key feature for the classification of the verbal classes is the pattern of allomorphy followed by the verbs (cf. Ralli 1988, 2005). Verb stems belonging to the second inflectional class (IC2) display a pattern $X(a) \sim X(i/e)$ (X represents part of the stem and the vowel in parenthesis is the stem-final vowel), whereas the absence of this pattern characterizes verb formations belonging to the first inflectional class (IC1):

(8)	(a) lyn-o 'to solve'	~	ely-s-a (IC1)
	STEM-INFL		STEM-ASP-INFL
	(b) agapa-o 'to love'	~	agapi-s-a (IC2)
	STEM-INFL		STEM-ASP-INFL

Inflectional class 1 is a very productive inflectional pattern and should be considered as the default inflectional class in MGr. Inflectional class 2 can be further divided into two sub-classes (Ralli, 1988, 2005):

(9) IC 2a: Xa ~ Xi, e.g. agapa-o 'to love' ~ agapi-s-a
 IC 2b: X ~ Xe, e.g. diair-o 'to divide' ~ diaire-s-a

These subclasses display major differences which can be easily detected in the paradigms of the present and the aorist:

	IC2a: X(a) ~ Xi		IC2b: X ~Xe	
	Present	Aorist	Present	Aorist
1sg	agap(á)-o	agápi-s- a	diair-ó	diaíre-s-a
2sg	agapá-s	agápi-s-es	diair-eís	diaíre-s-es
3sg	agapá (-ei)	agápi-s-e	diar-eí	diaíre-s-e
1pl	agapá-me~ agap-oúme	agapí-s-ame	diair-oúme	diairé-s- ame
2pl	agapá-te	agapí-s-ate	diair-eíte	diairé-s-ate
3pl	agapá-ne~ agap-oúne	agápi-s-an	diair-oún	diaíre-s-an

 Table 3: Inflectional Class 2

Let us now turn to the examination of the converted stems. The examination of the data shows that conversion and inflectional class of the output interact. Attention should be drawn, however, to the kind of interaction involved, since converted verbs in MGr are not inflected according to the default inflectional class.

Converted verbs are associated with some inflectional properties which *uniquely characterize* this type of formation. Converted verbs, both simple bases and compound formations, display a pattern of allomorphy which can be considered as a synthesis of the sub-classes of inflectional class 2. This can be shown in the analysis of the inflectional paradigms of the compound formations:

	Present,	Imperfect	Aorist	Future
	Future			[+perfective]
	[perfective]			
1	glossolog-ó	glossolog-oús-a	glossológi-s-a	glossologí-s-o (θa)
SG	(0 a)			
2	glossolog-eís	glossolog-oús-es	glossológi-s-es	glossologí-s-eis
SG	(0 a)			(0 a)
3	glossolog-eí	glossolog-oús-e	glossológi-s-e	glossologí-s-ei
SG	(0 a)			(0 a)
1	glossolog-oúme	glossolog-oús-ame	glossologí-s-ame	glossologí-s-oume
PL	(0 a)			(0 a)
2	glossolog-eíte	glossolog-oús-ate	glossologí-s-ate	glossologí-s-ete
PL	(0 a)			(0 a)
3	glossolog-oún	glossolog-oús-an	glossológi-s-an	glossologí-s-oun
PL	(0 a)			(θa)
	I	C2b	I	C2a

 Table 4: Inflectional patterns of converted stems

As the data in table (4) suggest, in Present, Future [-perfective] and Imperfect tense converted stems display the pattern of allomorphy which corresponds to IC2b, whereas in Aorist and Future [+perfective] tense they display the pattern of allomorphy which corresponds to IC2a. In other words, converted verbs (both simple bases and compound formations) display a pattern of allomorphy which can be considered as a *synthesis* of the two major patterns of allomorphy; this kind of phenomenon has been generally described as *heteroclisis* (cf. Maiden, 2009).

The pattern of converted verbs can be schematically represented in the following:

Inflectional class	Pattern of allomorphy
Verbs: Inflectional class 2a	Xa ~ Xi
Verbs: Inflectional class 2b	$X \sim Xe$
Converted verbs: Inflectional class 2c	X ~ Xi

Table 5: Allomorphy patterns in Modern Greek

On this view, I claim that conversion in MGr behaves similarly to other languages with rich inflection and, thus, the hypothesis that conversion and inflection interact is reinforced. However, it is not necessarily the case that conversion should be connected to the default inflectional class of the output. Conversion in MGr is connected to an inflectional class which is less productive. It should be mentioned that a number of simple bases also follow the same inflectional pattern. For example, the verb *tilefono* 'to call', in Present, Future [-perfective] and Imperfect tense display the pattern of allomorphy which corresponds to IC2b, whereas in Aorist and Future [+perfective] tense it displays the pattern of allomorphy which corresponds to IC2a. The basic difference between simple bases and converted verbs is that the latter are systematically connected to this pattern, whereas simple bases are not.

As a side remark, it should be mentioned that although IC2b is not very productive in MGr, it keeps (or reinforces) its productivity in converted formations from nominal bases. Productivity should be regarded as a phenomenon with a *gradient character* (cf. Bauer, 2001). Lieber and Baayen (1993) have argued that an affix which is not terribly productive may in fact gather strength in some well-defined subset of formations and reemerge as highly productive there.

A possible counterargument to this analysis would suggest that some verbs which synchronically belong to the class of converted verbs and descend from Ancient Greek may keep some inflectional properties of their Ancient Greek antecedent. For example, the verb *philosopho* 'a person who talks about philosophy', which synchronically belongs to the class of converted verbs, can also be found in Classical Ancient Greek. In this view, one may well assume that the pattern of allomorphy which corresponds to the Aorist and Future [+perfective] tense is not the pattern of the IC2a, but a relic from Ancient Greek verbal paradigms. Contrary to this claim, it should be mentioned that conversion is not restricted only to formations coming from Ancient Greek, since we find converted verbs which are not attested in Ancient Greek. For example, the verb *glossológos* is a newly coined formation which was not attested in Ancient Greek as compound.

The question that comes next is how best to represent the inflectional properties of the output verb. One may well assume that the inflectional properties of the verb can be considered as inherent lexical information of the converted formations. However, this is a questionable assumption since converted verbs do not correspond to a uniform inflectional pattern of the system. In the next section, I will present a fully fledged analysis of the relation between conversion and the properties of the output within a CM framework.

4. Construction Morphology and the relation between conversion and inflectional classes

In the model of grammar adopted here, words and abstract schemas of word formation are conceived of as a triple of Phonological Structure (PS), Syntactic Structure (SS), and Conceptual Structure (CS), in the sense of Jackendoff (2002, 2013). In Construction Morphology constructions at the word level are multidimensional formal units which codify phonological, semantic and morphological properties of the words (Booij, 2010, 2013).

Construction Morphology (CM) is based on the assumption that constructions, i.e. form-meaning (or function) pairs, are the basic units of the description and analysis of the linguistic phenomena. CM assumes that each word is a pairing of form and meaning. The form of a word in its turn comprises two dimensions, its phonological form, and its morphosyntactic properties.¹⁴ The structure of each component is generated independently and words form the 'correspondence' between the different levels of representations (Booij, 2010: 7).

CM provides a fully articulated model for the organization of the morphological component and the analysis of word-formation phenomena. The key idea behind the model proposed in CM is that there are *systematic form-meaning correspondences* between lexical items and these systematic relationships between sets of words form the starting point for the morphological analysis. In this respect, CM is similar to other constructionist models which assume that the grammatical creativity of language users can be accounted for by a network of relations between different lexical items (cf. Goldberg, 2006). However, contrary to other models which also assume paradigmatic relations (for example, Bybee, 2001) in CM patterns which represent the morphological creativity of the language user coexist with the individual formations within the same (morphological) component.

Morphological patterns are abstracted away from the comparison between members of lexical pairs. These patterns 'express predictable properties of existing words, indicate how new ones can be coined and give structure to the lexicon since complex words do not form an unstructured list but are grouped into subsets' (Booij, 2010: 4).

As Jackendoff (2010: 587-588) puts it, in parallel architectures, the interface relation between different components cannot be a sequenced derivation, since structures in different components often stand in a many-to-many relation. Rather, the interface components must be treated as *constraints* (possibly violable), which establish (or license) well-formed links among different kinds of structure.

Similarly, in CM every complex word is linked to output-oriented schemas which codify the phonological, morphological and semantic restrictions of

¹⁴ CM can be likened *Head-driven Phrase Structure Grammar* (HPSG) framework which assumes that the basic units of the linguistic analysis are signs which allow for parallel representation of phonological, syntactic, semantic and other information (Riehemann, 2001: 8).

word formation. The formal operation of word formation is *unification*. Through the interpretation of the variables, the schema turns into a lexical entry. If word formation does not meet the constraints, unification will fail. We now turn to the question of the representation of conversion pairs. In section 2.1 I argued that conversion can be conceived of as a paradigmatic relation between two lexical items, and that the formal and semantic properties of the converted item can be considered as part of this relation. A concrete example of this account in MGr would read as follows:

(10) < [[glossológ]_N-os]_N 'linguist' > \approx

 $< [[[glossolog]_N]_V - \acute{o}]_V$ 'perform the activity of a [glossológos]'>

The symbols $\langle \rangle$ indicate the form-meaning pair, while the paradigmatic relation is formalized with \approx . The schema in (10) represents the relation between the noun and the verb of a conversion pair which already exists. However, this representation faces some problems. First, it does not capture the fact that the same pattern can act as the basis for the coining of new words and second it does not show that nominal and verbal formations share some structural and semantic generalizations that can be projected onto the syntagmatic axis. Last but not least, it does not show that there are some semantic and formal correlations between parts of the members of this conversion pair. In what follows, I propose a representation which solves these problems.

Let us now examine what kind of information should be represented on the schemas. First, we need to show the change of the category, i.e. that nouns are converted into verbs. Second, the stress properties of the verb should be analysed as an output-constraint in the abstract schema. Last but not least, the schema should capture the semantic restrictions and the semantic correlations between the members of a word.

At this point, one may raise the question of whether the final stress of the verbal compound is a property of the constructional schema or a lexically specified property of the verb. Important evidence can be adduced from the examination of specific minimal pairs. Take, for example, the verbal formation *arthrografó* 'to perform the activity of a columnist', which has final stress, whereas the verb gráfo 'to write' has penultimate stress. These minimal pairs show that the final stress should be analysed as a phonological feature of the constructional schema.

The critical question of directionality remains to be answered. Booij (1997) uses the formal mechanism of *indices*, in order to represent the formal and semantic correspondences between the members of the conversion pair. The use of indices gives a straightforward account of the fact that there is some kind of correspondence between two words which have the same

phonological make-up and display a kind of semantic dependency. Seemingly, the use of indices opens the way to express the notion of directionality in conversion as the correspondence between certain formal and semantic aspects of the two lexical items involved.

In this view, each member of the conversion pair should be treated as a construction which links all these aspects. Accordingly, the representation in (10) should be reformulated as follows:

The schema in (11) represents conversion in MGr as a relation between constructional schemas. In order for a speaker to acquire the verbs in this set of structures, a specific element A (noun) should meet the formal and semantic requirements of the input structure and a specific element B (verb) of the output structure, with the same interpretation of the variables. Through interpretation of the variables -i.e. unification, the output structure turns into a lexical entry. If the two elements do not meet the relevant formal and semantic requirements, unification will fail.

The advantages of this representation are the following: this relation not only represents the formal and semantic properties of existing pairs, but also serves as the model for the formation of new words inasmuch as the noun meets the formal and semantic requirements of the input structure and the verb of the output structure, with the same interpretation of the variables.

But the question that now arises is: *what is the main theoretical proposal of Construction Morphology with respect to the relation between inflection and derivation (or conversion in this specific case)*? Construction Morphology, as developed by Booij (2010), offers the framework for the discussion of the interaction of two processes, since it has a strong lexicalist perspective, that is, it considers both inflection and derivation as morphological processes which are placed in the lexicon. Moreover, it proposes that the formal apparatus of schema is adequate for analysing a number of morphological phenomena.

As shown earlier, conversion in MGr is linked to specific inflectional properties of the output. The output of conversion is characterized by an inflectional class which is the synthesis of two inflectional classes, i.e. IC2a

¹⁵ I do not take a position as to the difficult question of the status and the position of the compound marker in Greek compounds. See Ralli (2013) for the relevant discussion.

and IC2b. This property can be codified on the schema of the verb as an output constraint:

 $\begin{array}{l} (12) \ [[[Xi-o]-[graf]_N]_V - \acute{o}]_{V[IC:2a/b][final stress]} & `Perform the activity of \ [[X_i-o]-[graf]_N - os]_{N[+ANIMATE AGENT]} & `one_who_writes Xi' \end{array}$

The inflectional property stems from the relation between the input and the output. It should be mentioned that the representation in (12) represents only the abstract information about the inflectional properties of the output and it does not show the distribution of the different inflectional paradigms in the entire conjugation of the verb.

As a corollary, the relation between conversion and inflectional properties of the output receives a straightforward account and the schema in (11) can be revised as follows:

5. Conclusion

The present paper had as a starting point the hypothesis that inflection and conversion interact, especially in languages with rich inflection, and examined this hypothesis with respect to the Greek data. The examination of this issue raised a number of important problems such as the position one takes on the grammatical status of conversion and the grammatical features of conversion in MGr.

The grammatical nature of conversion has been a matter of discussion in the literature. I argued that a paradigmatic account of conversion has in principle many advantages over the zero affixation analysis and the relisting hypothesis. However, it does not solve the problem of the representation of the formal and semantic properties of the output and it does not show the interrelations between the formal and semantic parts of the input and the output since it mainly represents existing pairs of conversion. Thus, I argue that it is not sufficient for the MGr data.

Denominal conversion in MGr has not been examined in detail. I argued that this type of conversion is mainly productive in compound formations with bound stems and very limited in simple bases. This fact falsifies the claim that conversion is mostly productive in simple bases. I also presented the formal and semantic properties of the output verb. Among others, I showed that the output verb has a specific stress pattern and its meaning contains the meaning of the corresponding noun. Moreover, I showed that the formation of the verb is subject to semantic restrictions linked to the feature [animate] of the input.

Among the formal properties of the output, the most dramatic one is the assignment of inflectional properties which are characteristic of this type of formations. The inflectional class of the output can be considered as a synthesis of two different inflectional classes in MGr, i.e. inflectional class 2a and 2b. Neither of these classes is the default one in MGr. In this respect, converted formations in MGr validate the general hypothesis that conversion may be related to the inflectional properties of the output, but they do not seem to validate the claim that the output of the process is inflected according to the default inflectional class.

The main challenge of the present paper was to present a construction-based account of the relation between denominal conversion and inflectional classes in MGr. I argued that conversion in MGr can be represented as paradigmatically related schemata which represent already existing formations and can form the basis for the coining of new formations. This representation has a number of advantages. First, it can adequately account for the asymmetry between form and meaning without making use of additional machinery, such as zero affixation, which has no independent justification and which is hard to define in terms of formal characteristics. Second, it represents the stress pattern of the output and the semantic correlations between the input and the output formation, such as (a) the semantic restrictions on the process and (b) the fact that the verb contains the meaning of the noun. Last but not least, it represents the inflectional properties of the output as an output constraint which needs to be satisfied in order for a formation to turn into an existing formation. In this respect, it shows that the schema can account for the interaction of the two processes in an adequate way.

References

Aronoff, M. 1980. "Contextuals". Language 56(4): 744-758.

- Aronoff, M. 1994. *Morphology by itself. Stems and inflectional classes* (Linguistic Inquiry monographs). MIT Press.
- Bauer, L. 2001. *Morphological productivity*. UK: Cambridge University Press.

- Booij, G. 1997. "Autonomous morphology and paradigmatic relations" in G. Booij & J. van Marle (eds), *Yearbook of Morphology 1996*. Dordrecht: Kluwer, 35-53.
- 2002. The morphology of Dutch. Oxford: Oxford University Press.
- 2010. Construction morphology. Oxford: Oxford University Press.
- 2013. "Morphology in Construction Grammar" in T. Hoffmann & G. Trousdale (eds), *Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, 255-273.
- Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Clark E. V. and H. H. Clark 1979. "When nouns surface as verbs". *Language* 55(4): 767-811.
- Don, J. 1993. *Morphological Conversion*. Ph.D. Dissertation. Utrecht: LEd.
- Dressler, W. 2003. "Degrees of grammatical productivity in inflectional morphology". *Italian journal of linguistics-Rivista di linguistica 15*(1): 31-62.
- Gardani, F. 2009. Dynamics of Morphological Productivity. A synchronic analysis and diachronic explanation of the productivity of nominal inflection classes from Archaic Latin to Old Italian in terms of Natural Morphology. Ph.D. Dissertation. University of Vienna, Vienna.
- Goldberg, A. 2006. *Constructions at work: the nature of generalization in language*. USA: Oxford University Press.
- Hüning, M. 2009. "Semantic niches and analogy in word formation. Evidence from contrastive linguistics." *Languages in contrast 9* (2), 183-201.

- Jackendoff, R. 2002. Foundations of language: Brain, Meaning, Grammar, Evolution. USA: Oxford University Press.
- 2010. "The parallel architecture and its place in cognitive science" in B. Heine & H. Narrog (eds), *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press, 583-605.
- 2013. "Constructions in the parallel architecture" in T. Hoffmann & G. Trousdale (eds), Oxford Handbook of Construction Grammar. Oxford: Oxford University Press, 70-92.
- Kiparsky, P. 1982. "Lexical Phonology and Morphology" in S. Yang (ed.), *Linguistics in the morning calm.* Seoul, Korea: Hanshin Publishing Company, 3-91.
- Koutsoukos, N. 2013. A constructionist view of complex interactions between inflection and derivation: the case of SMG and Griko.Ph.D. Dissertation. University of Patras, Patras.
- Lieber, R. 1980. *On the organization of the lexicon*. Ph.D. Dissertation. Massachusetts Institute of Technology.
- 1981. "Morphological conversion within a restrictive theory of the lexicon" in M. Moortgat, H. Van der Hulst & T. Hoekstra (eds), *The scope of lexical rules*. Dordrecht: Foris Publications, 161-200.
- 2004. Morphology and lexical semantics. Cambridge, UK: Cambridge University Press.
- Lieber, R. and H. Baayen 1993. "Verbal prefixes in Dutch: a study in lexical conceptual structure" in G. Booij and J. van Marle (eds.), *Yearbook of morphology 1993*. Dordrecht: Kluwer, 51-78.
- Maiden, M. 2009. "Where does heteroclisis come from? Evidence from Romanian dialects". *Morphology* 19: 59-86.

- Manova, S. 2011. Understanding morphological rules: with special emphasis on conversion and subtraction in Bulgarian, Russian and Serbo-Croatian (Studies in Morphology 1). Dordrecht: Springer.
- Marchand, H. 1969. *The categories and types of present-day English word formation. A synchronic-diachronic approach* [2nd edition]. München: C.H. Beck'sche Verlagsbuchhandlung.
- Marle, J. van 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.
- Ralli, A. 1988. *Eléments de la morphologie du grec moderne: la structure du verbe*. Ph.D. Dissertation. Université de Montréal, Montreal.
- 2005. Μορφολογία [Morphology]. Athens: Patakis Publications.
- 2008. "Greek Deverbal Compounds with bound stems". Southern Journal of Linguistics 29 (1/2): 150-173.
- 2013. Compounding in Modern Greek (Studies in Morphology 2).
 Dordrecht: Springer.
- Riehemann, S. 2001. *A constructional approach to idioms and word formation*. Ph.D. Dissertation. Stanford University.
- Sanders, G. 1988. "Zero derivation and the overt analogue criterion" in M. Hammond and M. Noonan (eds.), *Theoretical morphology: approaches in modern linguistics*. San Diego, California: Academic Press, Inc, 155-178.
- Thornton, A.M. 2004. "Conversione. Introduzione" [Conversion. Introduction] in M. Grossmann and F. Rainer (eds), La formazione delle parole in italiano. Tübingen: Niemeyer, 501-533.

Tribout, D. 2010. *Les conversions de nom à verbe et de verbe à nom en français*. Ph.D. Dissertation. Université Paris 7.

ON MULTIFUNCTIONAL LEXEMES IN FRENCH

Daniela Marzo

Ludwig-Maximilians-Universität München (Germany)

Abstract

This article critically discusses the existence of *multifunctional lexemes* in the French lexicon alongside the well-established phenomena of *conversion* and *categorial distortion*. The objects of analysis are homonymous adjective-noun doublets of different semantic groups: (1) quality as well as (2) colour terms, (3) language names, (4) ethnic doublets and (5) profession terms. While at first sight their formal and morphosyntactic properties qualify these doublets as multifunctional lexemes in the sense of Bauer & Valera (2005), a more fine-grained analysis of their prototypical adjectival and nominal characteristics shows, on the contrary, that multifunctionality is close to non-existent in the French adjective-noun domain (and the lexicon in general). Recent classifications of certain semantic subgroups as multifunctional lexemes are thus challenged (*cf.* Kiefer 2011).

1 Research questions and aims

The main purpose of this article is to critically discuss the relevance of the phenomenon of *multifunctionality* for the French lexicon. While there is broad agreement on the existence of related phenomena such as *conversion* and *categorial distortion*,¹ the hypothesis of multifunctionality is relatively new to the study of Romance languages (*cf. e.g.* Kiefer 2011 for French) and has originally been developed mainly on the basis of non-inflecting languages (*cf.* Manova & Dressler 2005: 71; for a brief overview *cf.* Vogel 1996: 224-

¹ In the adjective-noun domain: for conversion, *cf. e.g.* Thornton (2004: 508-515) for Italian; Schwarze (2012) and Corbin (1987: 479) for French; Rainer (1993: 74, 678-683) for Spanish; Lüdtke (2005: 45, 125-126) for all Romance languages; for *categorial distortion* or *mismatch*, *cf.* Kerleroux (*e.g.* 1996: 132) and Lauwers (2008) for French; in the verb-noun domain: for conversion, *cf. e.g.* Soares Rodrigues (2009) for Portuguese; Tribout (2012) for French; Marzo (2013) for Italian; for *categorial distortion cf.* Kerleroux (*e.g.* 1996: 101-108) for the French and Marzo & Umbreit (2013) for the French and Italian nominalised infinitive.

236). The discussion will focus on French² adjective-noun doublets such as in (1) to (5), that constitute particularly plausible candidates for multifunctionality thanks to the homonymy of the adjectival and nominal word forms (for definitions of the notions of conversion, categorial distortion and multifunctionality, cf. section 2).³

(1)	a.	calme A SG MASC/FEM 'calm'	-	calme N SG MASC 'calm'
	b.	inapte A SG MASC/FEM 'inapt'	_	inapte ⁴ N SG MASC 'inaptitude'
(2)		jaune A SG MASC/FEM 'yellow'	_	jaune N SG MASC 'colour yellow'
(3)		italien A SG MASC 'in Italian langua	– nge'	italien N SG MASC 'Italian language'
(4)		italien A SG MASC 'Italian'	_	Italien N SG MASC 'male Italian person'
(5)		fermier A SG MASC 'country-related'	_	fermier N SG MASC 'farmer'

A hitherto open question is, however, how multifunctionality can in these cases be distinguished from conversion and categorial distortion and, in particular, what role inflectional markers play for the distinction of the three categories.

² The discussion is limited to French for reasons of space, but the argumentation can be transferred quite straightforwardly to other Romance languages, as they resemble each other profoundly with respect to the properties of adjectives and nouns.

³ Homonymous adjective-noun pairs are very numerous in French (cf. Noailly 1999: 14). Schwarze (2012: 154), e.g., counts 2242 pairs in Petit Robert.

⁴ In expressions such as (6), cf. section 2.2.

To this aim, the differences between these categories will first be discussed from the perspective of the existing traditional definitions (*cf.* section 2). Before the French adjective-noun domain – exemplified by (1) to (5) – will be analysed with regard to the pertinence of the three categories (*cf.* 3.3), it will be discussed which formal and functional characteristics a multifunctional lexeme must have, by definition, in the Romance adjective-noun domain. In this context a closer look at the notions of *prototypical adjectives* and *nouns* (*cf.* 3.1) as well as *lexical category overlap* (*cf.* 3.2) will be necessary. It will be shown that *gender fixation* is a particularly valuable criterion for the distinction of multifunctionality, conversion and categorial distortion in Romance languages. Section 4 will summarise the reflections of this paper and give an outlook on further research.

2 Conversion, categorial distortion and multifunctionality: different phenomena

2.1 Conversion

The relation of the adjective to the noun in most of the doublets in (1) to (5) is traditionally labelled as a conversion relation (*cf.* the references in footnote 1) in the sense that one lexical item that is specified for a specific part of speech is – without the addition of any word-formation affix – converted into another lexical item bearing another part of speech value (for the definition *cf. e.g.* Bauer 2005: 18-19). From this perspective, the French noun *italien* 'Italian language' in (3), for example, has been formed on the basis of its adjectival homonym *italien* 'in Italian language' by conversion from the category *adjective* to the category *noun*.

While the output of a conversion process in Romance languages is – in contrast to categorial distortion (*cf.* section 1.2) and multifunctionality (*cf.* section 1.3) – always an independent word, i.e. a new entry in the lexicon, it is not always clear whether the input to the conversion process is a *root*, a *stem* or a *word form* (*cf.* the levels of categorisation in Lehmann 2008: 548). In contrast to the verb-noun domain, where there is, in many cases, a formal hint (*e.g.* incorporated inflectional markers or the theme vowel)⁵ that points to

⁵ While in weakly-inflecting languages such as English the distinction of the three levels is – from the perspective of the form – not self-evident (but *cf.* Arad 2003 for a semantic distinction), in Romance languages the formal aspect of the sign can be attributed, at least in some cases of the verb-noun domain, to one of Lehmann's levels of categorisation (Lehmann 2008: 548): The French nominalised infinitive *pouvoir* 'power', *e.g.*, still contains the infinitival inflectional marker –*oir* of its verbal base *pouvoir* 'to be able to' and therefore constitutes an instance of *word form conversion* (*cf.* also Gévaudan 2007: 122). *Stem conversion* traditionally refers *e.g.* to feminine

the level of categorisation the conversion basis belongs to, the adjective-noun domain is deprived of such hints, as not only the root level, but also the stem and the word form level are perfectly homonymous thanks to homonymous inflectional affixes (*cf. e.g.* Goes 1999: 140-141 for French; Thornton 2004: 508 for Italian). Most classifications of conversion within the Romance adjective-noun domain seem, however, not to bother with a differentiation between the root and the stem level (*cf.* Gévaudan 2007: 121; Lüdtke 2005: 127-128; but *cf.* Lehmann 2008: 554 for Spanish) and sometimes not even between the latter two and the word form level (*cf.* Thornton 2004: 508-515; Rainer 1993: 676-683).

2.2 Categorial distortion

According to Kerleroux (especially 1996: 137-139) the examples in (1) to (5) are not all instances of conversion, but form actually two different groups, i.e. (morphological) conversion such as the noun French calme 'calm' in (1a) and (syntactic) categorial distortion, such as the French nominal form *inapte* 'inapt' in (1b). In her opinion (cf. Kerleroux 1996: 137-139), an analysis of the linguistic elements that can successfully be used directly after adjectives in nominal position (such as those in (6) and (7)) allows for the distinction of categorial distortion and conversion. For instance, the prepositional phrase aux travaux manuels in (6a) is an argument of the French adjective inapte. Its argumental status strongly speaks in favour of the rather adjectival character of French *inapte*, irrespective of the nominal syntactic position of the latter. Accordingly, French *inapte* cannot take adjectives as modifiers (cf. total in (6b)) and be followed by modifying prepositional phrases (cf. de première catégorie in (6c)) even if used in a nominal position where we might expect such elements. Moreover, it cannot be directly followed by a relative clause such as in (6d). These restrictions all speak against *inapte* being a full noun which is why Kerleroux labels nominal uses of adjectives such as *inapte* as categorial distortion. In contrast, cases such as French calme as in (7) are full

nouns such as It. *sosta* 'stop, rest' that result from conversion of the infinitival stem containing the verbal theme vowel -a-(cf. the stem *sosta-* of the infinitive *sostare* 'to stop, to rest'). After conversion, the vowel -a- is reanalysed as grammatical marker of the typically feminine nominal class in -a (cf. Thornton 2004: 517-518). In the case of *root conversion*, in contrast, a root (in Romance languages typically consonantic) is converted from one category to the other. According to Thornton (2004: 516-517), the Italian verbal root *acquist-* (*e.g.* in *acquistare* 'to buy') is converted to the nominal root *acquist-* that is then – in the absence of an obligatory vocalic ending – integrated into the most productive and typically masculine nominal class in -o (for a more detailed and critical discussion of the input forms for conversion in Romance languages).

nouns insofar as they can be modified by adjectives (*cf. impressionant* in (7a)) as well as prepositional adjuncts (*cf. de statue* in (7b)) and be the antecedent of relative clauses (*cf. qui désarme ses adversaires* in (7c)).

- (6) *cf.* Kerleroux (1996: 136)
 - a. Ce type est d'un inapte aux travaux manuels ! this guy is of an inapt [to the work manual] ARG PP 'this guy is inapt to manual work'
 - b. *Ce type est d'un inapte total. this guy is of an inapt total A 'this guy is totally inapt'
 - *Ce type est d'un inapte de première catégorie.
 this guy is of an inapt [of first category] MOD PP 'this guy is a very inapt person'
 - d. *Ce type est d'un inapte qui décourage ses professeurs. this guy is of an inapt that discourages his teachers 'this guy is that inapt that his teachers are discouraged'
- (7) *cf.* Kerleroux (1996: 132)
 - a. Il est d'un calme très impressionnant. he is of a calm [very impressive] AP 'he is impressively calm'
 - b. Il est d'un calme de statue. he is of a calm [of statue] MOD PP 'he is calm like a statue'
 - c. Il est d'un calme qui désarme ses adversaires. he is of a calm [that disarms his adversaries] MOD REL CLAUSE 'he has a calm demeanour that disarms his adversaries'

Summing up, we can say that the output of categorial distortion is, in contrast to the output of a conversion process (cf. section 2.1), not a new and

independent entry in the lexicon,⁶ but a lexical item specified for one category used in a syntactic position that is generally reserved to other categories. It follows from this that the input to categorial distortion always is a word form, while conversion may take roots or stems as input forms.

2.3 Multifunctionality

Whereas the distinction of both conversion and categorial distortion is by now well-established in Romance linguistics, the hypothesis of multifunctionality of lexical items is still new to the study of Romance languages and has so far been advanced mainly for ethnic adjective-noun doublets (*e.g.* in Kiefer 2011: 115) such as French *italien* (A) 'Italian' – *Italien* (N) 'male Italian person' in (4). According to Kiefer (2011: 115) and Bauer & Valera (2005) ethnic-adjective noun doublets qualify for multifunctionality rather than conversion or categorial distortion crosslinguistically, because they meet the following definition:

[...] any member of the set of words which can function as if being part of word-class *w* can also be used, under appropriate semantic circumstances, as if it were a part of word-class $x \ (x \neq w)$ [...] (*cf.* Bauer & Valera 2005: 10-11)

While at first sight this definition resembles quite closely to those of conversion (cf. section 2.1) and of categorial distortion (cf. section 2.2) insofar as different word-classes as well as contextual appropriateness play a major role, the characteristic that distinguishes multifunctionality is, upon a closer look, the application of the necessary conditions (Mi), (Mii) and (Miii) implied in the above-mentioned definition:

⁶ Unless, of course, the outcome is lexicalised in a second step (*cf. e.g.* most of the modern French nominalised infinitives; Kerleroux 1996 and Marzo & Umbreit 2013). Lauwers critically discusses some of Kerleroux's examples and states, *e.g.*, that French *calme* (N) 'calm' in (1b) also is the product of categorial distortion, but one of the rare cases that have been lexicalised as a noun (*cf.* Lauwers 2008: 172). The main argument in favour of such an analysis is the gradual lexicalisation of the adjectives as a noun that can be observed in their different degrees of nouniness across doublets (*cf.* the analysis in Lauwers 2008: 148-159). He adds, however, that "from a synchronic point of view, this legitimates an analysis by conversion, as also a separate lexicographic treatment (i.e. two different entries)" (Lauwers 2008: 172). His example for conversion in the strict sense are colour terms of type (2) *jaune* (A) 'yellow' – *yellow* (N) 'colour yellow' (*cf.* Lauwers 2008: 172), because they have prototypical nominal properties from the beginning of their formation.

(M) i. semantic stability

ii. generality

iii. prototypicality

(Mi), the semantic stability criterion, ensures that the semantic relation across doublets is always the same. This condition is generally met by ethnic adjective-noun doublets such as italien (A) 'Italian' - Italien (N) 'male Italian person' in (4), because the relation between the ethnic adjective and the ethnic noun can always be described as something like "A = that concerns the N, the culture of N". According to (Mii), the generality criterion, every lexical item X used as an instance of category A can also be used as an instance of category B. Again, this is in general true for ethnic adjective-noun doublets such as in (4), as there is, at first sight, a homonymous adjective for every ethnic noun and a homonymous noun for every ethnic adjective.⁷ As for (Miii), the prototypicality criterion, it means that a lexical item X can indifferently be used in nominal as well as in adjectival contexts, because its use as an instance of both categories A and B is prototypical in both contexts. Ethnic adjectives and nouns usually meet this condition, too, as they are prototypical instances of their respective categories insofar as they have all morphosyntactic characteristics of both categories. While it is true that some of these properties can also be found with single instances of conversion⁸ and categorial distortion,⁹ the distinctive characteristic of multifunctionality is that (Mi) to (Miii) are necessary conditions and only sufficient in combination. In contrast, the only necessary condition for conversion is the criterion of category change¹⁰ (cf. section 2.1 and (C) below) that typically goes along with the acquisition of the prototypical characteristics of the target category. The only necessary condition of categorial distortion is, in turn, category mismatch between the category A of a word form and the category B that is required by a given syntactic context (cf. section 2.2 and (D) below) - a mismatch that is most prominently manifest in the morphosyntactic

⁷ Still, in Romance languages there seem to be exceptions, such as *hébraïque* 'Hebrew' in French that is not used as a noun (many thanks to Franz Rainer for this hint during a discussion at *Universals and Typology in Word-Formation II* in Košice). ⁸ *E.g.*, (Miii) in the case of *marcher* (V) 'to walk' – *marche* (N) 'walk', where both the verb and the noun are perfectly representative and prototypical instances for their respective categories.

 $^{{}^9} E.g.$, (Mii) in the case of the Italian nominalised infinitive; in contrast to French, where the nominal use of infinitives is no longer productive, each Italian infinitive can, in principle, be used as a noun (*cf. e.g.* the comparison of French and Italian in *e.g.* Marzo & Umbreit 2013).

¹⁰ But cf. Thornton (2004: 505-508) on intracategorial conversion.

behaviour of the "distorted" item that does not adapt to the characteristics of the target category.

- (C) category change (with acquisition of the prototypical morphosyntactic behaviour of the target category)
- (D) category mismatch (without acquisition of the prototypical morphosyntactic behaviour of the target category)

3 Multifunctionality in the French lexicon?

As exposed in section 2 and defended by Kiefer (2011), French ethnic adjective-noun doublets such as in (4) may be interpreted as one single multifunctional lexeme rather than categorial distortion or conversion. In order to tell whether other French adjective-noun doublets are actually one multifunctional lexeme, a lexeme that is only distorted into another category, or two independent lexemes (cf. 3.3), we first need to specify the prototypical characteristics of the categories *adjective* and *noun* in French (cf. 3.1). From the discussion of the prototypical properties of adjectives and nouns in Romance languages (such as French) it will be clear that multifunctionality can only occur where otherwise distinct and distinguishable categories *overlap* (cf. 3.2). In this respect, fixed gender will turn out to play a major role for the distinction of multifunctionality from related phenomena.

3.1 A note on prototypical adjectives and nouns in French

In Romance languages, nouns and adjectives share important characteristics: From the semantic perspective, there is a large zone of continuity between the two categories (*cf.* Goes 1999: 170) which is paralleled by the fact that both adjectives and nouns display the same grammatical phenomena, i.e. gender and number. In addition, they generally use homophonous inflectional morphemes to express these categories (*e.g.* Thornton 2004: 508 for Italian; Goes 1999: 140-141 for French). As a consequence of especially the latter characteristic, adjective-noun doublets are, in general, perfectly homonymous and therefore a priori more susceptible to be instances of multifunctionality than other word pairs. The only formal difference between nouns and adjectives is that nouns typically display fixed gender (*e.g.* Winther 1996: 44) and thus inherent inflection in the sense of Booij (1993), while adjectives vary in gender contextually (*e.g.* Thornton 2004: 508 for Italian; Goes 1999: 141 for French) and therefore show, in Booij's terms, contextual inflection.

However, there are two exceptions to these prototypical characteristics of adjectives and nouns: First, some adjectives are formally invariable such as the French colour term *marron* 'brown' in (9), or only inflect for plural and

not for gender such as French *jaune* 'yellow' in (10) (in contrast to (8) *vert* 'green' that inflects for both gender and number).

(8)	a. le DET SG MASC 'the green pull	pull N SG MASC lover'	vert A SG MASC
	b. la DET SG FEM 'the green skir	jupe N SG FEM t'	verte A SG FEM
	c. les DET PL MASC 'the green pull	pulls N PL MASC lovers'	verts A PL MASC
	d. les DET PL FEM 'the green skir	jupes N PL FEM ts'	vertes A PL FEM
(9)	a. le DET SG MASC 'the brown pul	pull N SG MASC llover'	marron A SG MASC
	b. la DET SG FEM 'the brown ski	jupe N SG FEM rt'	marron A SG FEM
	c. les DET PL MASC 'the brown pul	pulls N PL MASC llovers'	marron A PL MASC
	d. les DET PL FEM 'the brown ski	jupes N PL FEM rts'	marron A PL FEM

(10)	a. le DET SG MASC 'the yellow p	pull N SG MASC ullover'	jaune A SG MASC
	b. la DET SG FEM 'the yellow sł	jupe N SG FEM kirt'	jaune A SG FEM
	c. les DET PL MASC 'the yellow p	pulls N PL MASC ullovers'	jaunes A PL MASC
	d. les DET PL FEM 'the yellow sł	jupes N PL FEM kirts'	jaunes A PL FEM

Second, some nouns vary in gender according to the sex of the referent, such as the nominal use of *Italien* 'Italian' in (4) (*cf.* also (11)) and, in general, many nouns designing human professions such as in (12) (*cf.* also (5)):

- (11) a. l' Italien DET SG MASC N SG MASC 'a (male) Italian'
 - b. l' Italienne DET SG FEM N SG FEM 'a (female) Italian'
- (12) a. le boulanger DET SG MASC N SG MASC 'a (male) baker'
 - b. la boulangère DET SG FEM N SG FEM 'a (female) baker'

Although such examples seem, at first glance, to be exceptions to the prototypical behaviour of adjectives and nouns in French, they are still prototypical members of their respective category. In French, adjectives that only inflect for number, but not for gender, seem even to be majoritarian in comparison to fully inflected adjectives (cf. Goes 1999: 60). What is even more important, however, is that they can nevertheless be used with feminine and masculine nouns, just as *vert* in (8). The only difference is that instances

such as *jaune* in (10) just do not show gender formally. Consequently, from the perspective of their quantitative importance and their compatibility with either gender, instances such as *jaune* in (10a) and (10b) are not less prototypical for the category adjective than *vert* in (8a) and (8b). Similarly, nouns that vary in gender according to the sex of the referent are by no means untypical nouns. The question whether they actually inflect for gender (and, by the way, for number) or show fixed gender (and number) is a question of lexical semantics and does not affect the degree of prototypicality of their use as nouns in syntactic positions reserved to nouns.

3.2 Category overlap, multifunctionality and gender fixation

As sketched in section 2.3, according to the prototypicality criterion (Miii) a multifunctional lexical item X can indifferently be used in nominal as well as in adjectival contexts, because it has characteristics that perfectly fit both contexts. In other words, if a lexeme X has the prototypical inflectional characteristics of category A and these are not shared by category B, it is not a prototypical instance of category B and hence not a multifunctional lexical item. Consequently, if X can without any exception function both as a prototypical instance of category A and of category B, the inflectional differences that typically hold between the two categories must somehow be neutralised. This can, by definition, only be the case in the area in which categories overlap. Since Romance adjectives always agree in gender with the nominal they modify (even if they do not show it formally as in the case of French *marron* in (9)), the only logically possible candidates for an overlap between the categories are adjective-noun doublets that can be used in both masculine and feminine gender in contexts reserved to adjectives as well as in contexts reserved to nouns. As in Romance languages the only nouns that are used in both genders are nouns referring to (fe)male entities,¹¹ only adjectivenoun doublets whose nominal uses refer to (fe)male entities are potential candidates for multifunctionality. In contrast, adjective-noun doublets whose nominal variant display fixed gender are no candidates for multifunctionality. Gender fixation is thus a major criterion for the distinction of multifunctionality, conversion and categorial distortion in French (and other Romance languages).

3.3 Multifunctionality outside the ethnic domain?

Against the background outlined in sections 3.1 and 3.2 the question arises whether there are other adjective-noun doublets than ethnic doublets in French that qualify for multifunctionality. In the following sections we will

¹¹ Unless specified otherwise in their lexical meaning, *cf.* French *le médecin* (DET MASC + N MASC) 'female or male doctor' (*cf.* also (15)).

therefore discuss whether instances such as quality terms in (1), colour terms in (2) and language names in (3) as well as doublets whose nouns refer to professions in (5) also qualify for multifunctionality if tested against the necessary conditions listed in (M).

3.3.1 The semantic stability condition (Mi)

As in the case of the ethnic adjective-noun doublets in (4), condition (Mi) is also respected in the case of quality terms in (1), colour terms in (2) and language names in (3) as well as doublets whose nouns refer to professions such as in (5).¹² The semantic relations that hold between the respective adjectives and the nouns can be characterised in the following ways:

The relation between the nominal use and the adjectival use of abstract quality doublets such as *calme* (1a) and *inapte* (1b) can always be paraphrased as "A = having the quality designated by N". The relation between colour term doublets such as in (2) *jaune* (*cf.* also (8), (9) and (10)), in turn, can be described as "A = having the colour designated by N." As for language names such as *italien* in (3) (Mi) is met thanks to the stable relation "A = in language designated by N". Profession nouns such as *fermier* 'farmer' in (5) and *boulanger* 'baker' in (12), too, are at first sight good candidates for multifunctionality for the reason that the semantic relation between the nominal and the adjectival use is constantly "A = property related to N".

3.3.2 The generality condition (Mii)

As stated in section 2.3, multifunctionality is characterised by the fact that, within a semantic group (cf. (Mi) and section 3.3.1) every lexical item X that is used as an instance of category A can also be used as an instance of category B. In contrast to the semantic stability criterion (Mi), the generality condition (Mii) is not systematically met by all doublets in (1) to (3) and (5). This is, for instance, the case with the quality doublets in (1): Even if presumably every adjective of type (1b) *inapte* can be used in a nominal position such as illustrated in (6a), the reverse situation is not true, as not every abstract quality noun can also be used in positions that are usually

¹² These doublets exemplify well-defined semantic groups that are discussed as such in the literature (*cf. e.g.* Goes 1999: 163-170 for French profession terms; Kiefer 2011: 116 for French colour terms; Lauwers 2008 and Beauseroy & Knittel 2007 for different types of French quality doublets; Schwarze 2012 for all these types; Villalba 2009 and Rainer 1993: 681-683 for Spanish abstract nominalisations; Rainer 1993: 678-681 for Spanish nouns designing humans – including professions – and 1989 for Italian quality nouns).

reserved to adjectives, as for instance *inaptitude* (N) 'inaptitude' in (13a) versus (13b):

- (13) a. Il est d'une inaptitude très impressionnante. he is of an inaptitude N very impressive 'he is impressively inapt'
 - b. *Il est très inaptitude. he is very inaptitude N 'he is very inapt'

In addition, full abstract nouns such (1a) *calme* are limited to a few lexicalised instances (*cf.* Lauwers 2008: 161), which also speaks against their being multifunctional lexemes.¹³ From this perspective abstract quality terms are not instances of multifunctionality, because they do not respect the necessary generality condition (Mii).

As for language terms of type (3), too, we can state that there are exceptions to a general utilisation in both adjectival and nominal contexts (*cf.* (14b) in contrast to (14a)), as French *hébraïque* 'in Hebrew language' has no homonymous nominal language term. Similarly, the French noun *quéchua/quichua* 'the language Quechua' has no adjectival counterpart (*cf.* (14c)), which is why the adjectival concept has to be expressed by the help of a prepositional phrase (*cf.* (14d)).

(14)	a. hébreu	_	hébreu
	A MASC		N MASC
	'in Hebrew language'		'Hebrew language'
	b. hébraïque	_	*hébraïque
	A MASC/FEM		N MASC
	'in Hebrew language'		'Hebrew language'

¹³ While it can be observed that in French and other Romance languages every adjective can somehow be used as a noun, it is interesting to see that the domain of abstract quality terms is less productive than other semantic groups with respect to the nominal use. The most productive group is the one in which the adjectival meaning undergoes a sort of reification in that its nominal use refers to the totality of what has the quality expressed by the adjective. While the adjective *beau* means 'beautiful', the noun in *le beau* refers to all beautiful things in that its meaning can be paraphrased as '(the totality of) what is beautiful' (*cf.* Lauwers 2008 and more recently Gauger 2011), while *beau* in structures such as *il est d'un beau* 'he is of a beauty (=he is beautiful)' is synonymous to the suffixed abstract quality noun *beauté* 'beauty'.

c. *quéchua/quichua –			-	quéchua/quichua
	A MASC/FEM			N MASC
	'in Quechua?	,		'the language Quechua
d	. un	texte	en	quéchua/quichua
	DET MASC	N MASC	PREP	N MASC
'a text in Quechua'				

Profession terms such as in (5), too, are no multifunctional lexemes and rather instances of conversion from the perspective of (Mii). Not every profession term has a homonymous adjective as can be seen in (15). Although *médecin* can be used in attributive position as in (15a), it is not an adjective as can be seen if we try to use it in predicative position (cf. (15b)):

(15)	a. une	femme	médecin			
	а	woman N FEM	doctor N MASC			
	'a female doctor'					
b. *Paul est très médecin. ¹⁴						
	Paul is very doctor A MASC.					
	'Paul is very doctorlike'					

Similarly, one might want to sustain that in the case of colour term doublets such as *jaune* in (2) the generality condition is violated by the existence of adjectives such as *bleuâtre* 'bluish' in (16), that have no nominal correspondence:

(16)	bleuâtre	—	*bleuâtre
	A SG MASC/FEM		N SG MASC
	'bluish'		'the colour bluish'

However, as instances such as *bleuâtre* 'bluish' in (15) are no proper colour terms insofar as the French suffix $-\hat{a}tre$ (just as, by the way, its English counterpart -ish) is used to express the fact that the designated colour shade is only a marginal member of a given colour, this argument is not a very strong argument for the violation of the generality condition.

Summing up what has been said so far about the doublets in (1), (2), (3) and (5) as candidates for the phenomenon of multifunctionality, we can say that, from the perspective of the generality condition (Mii), only colour term

¹⁴ The same construction can, of course, be used in other senses as *e.g.* 'N likes/likes wearing N', such as in *Marie est très jupe* (literally: 'Marie is very skirt').

doublets such as *jaune* in (2) clearly qualify for multifunctionality. While quality terms as in (1) and profession terms as in (5) are at once ruled out by the existence of numerous (suffixed) quality nouns and profession nouns that cannot be used as adjectives, the exceptions to a general use in both categories seems, at least at first sight, to be rather marginal in the area of language names such as in (3). Further research should investigate the relative importance of these exceptions for the notion of multifunctionality.

3.3.3 The prototypicality condition (Miii)

As stated in section 2.2, Kerleroux (especially 1996: 137-139) considers the relation between the adjectival and the nominal use of quality terms such as in (1a) calme 'calm' as a conversion relation, whereas she classifies instances such as (1b) inapte 'inapt' as categorial distortion, the main difference between the two being that instances of categorial distortion do not adapt to the prototypical characteristics required by syntactic positions reserved to nouns (cf. (6) versus (7)), but bring along properties of the category adjective (also cf. (D) in section 2.3). From the perspective of multifunctionality, it is precisely the fact that instances such as (1b) inapte cannot be used as perfectly prototypical instances of both the categories adjective and noun that they do not qualify for multifunctionality. This means that condition (Miii) is not met insofar as even in a context reserved to nouns (1b) inapte behaves like an adjective and not like a prototypical noun with respect to its morphosyntactic characteristics. In opposition to (1b) *inapte*, instances such as (1a) *calme* display all morphosyntactic properties of both the categories adjective and the categories noun. It has already been shown in (7) that (1a) calme behaves just like every other prototypical noun would do in the same context. Similarly, *calme* is also a prototypical adjective, as it is modified by adverbs, inflects for number and can be used with nouns of both masculine and feminine gender as in (17).

(17)	a. une nuit a a night N SG FEM o 'a quite calm night'		assez quite AI	assez quite ADV		calme calm A SG FEM	
	b. un matin a morning N SG MA 'a very calm morning'			ASC	très very AD	V	calme calm A SG MASC
	c. quelo some 'som	ques e e quite ca	nuits nights M alm night	N PL FEM ts'	assez quite AI	DV	calmes calm A PL FEM

d. quelques matins	très	calmes	
some mornings N PL MASC	very ADV	calm A PL MASC	
'some very calm mornings'			

However, French *calme* in (1a) is nevertheless different from the ethnic doublets in (4), even if it behaves, just like them, as a prototypical adjective in its adjectival and as a prototypical noun in its nominal use. Unlike in the case of ethnic doublets, the nominal use of quality doublets is confined to masculine gender only which speaks, as has been shown in section 3.2, strongly against multifunctionality.

The picture is quite similar in the case of colour doublets such as (2) *jaune*: They cannot only be used as prototypical adjectives (*cf. e.g.* gender and number agreement in (8), (9) and (10) and modification by adverbs in (18a)), but also as prototypical nouns, such as in (18b), where *marron* is pluralised and modified by an adjective. This pullover is a little too brown

- (18) a. ce pull est un peu trop marron this pullover N PL MASC is a little too ADV brown A SG MASC 'this pullover is a little too brown'
 - b. les marrons clairs que Picasso utilise the browns N PL MASC light A PL MASC that Picasso uses 'the different types of light brown that Picasso uses'

But again, in their nominal form, these colour terms cannot be used in both genders, but always display fixed masculine gender, while the adjectival uses agree in gender with the noun they refer to even if gender agreement is not necessarily marked formally as in the case of invariable adjectives (*cf. e.g.* (9)). As the grammatical characteristics of the nominal and the adjectival uses are thus not the same, colour terms are not instances of multifunctionality in the strict sense of the term (*cf.* section 3.2). Most importantly, Kiefer (2011: 116) comes to the same conclusion with respect to colour terms, but from the perspective of conversion. He shows that there are two different types of colour term conversion, one that systematically forms nouns on the basis of basic colour adjectives such as *jaune* 'yellow' in (2), the other that forms adjectives from nouns that have a characteristic colour such as the conversion of *marron* (N) 'chestnut' to *marron* (A) 'brown'.

In comparison to quality and colour terms, language name doublets such as (3) *italien* are particular insofar as not all adjectival uses behave like prototypical adjectives. For instance, while *italien* and *hébraïque* can be used with feminine as well as masculine nouns and *italien* even uses inflectional gender markers (*cf.* (19a) and (19b) in contrast to (19c) and (19d)), the
adjective *hébreu* is limited to the context of masculine nouns ((*cf.* 19e) in contrast to (19f)). (Miii), the prototypicality criterion, is thus not observed by all adjectival language terms, a factor speaking against them being multifunctional items.

(19)a. un texte italien N MASC A MASC 'a text in Italian' b. une lettre italienne N FEM A FEM 'a letter in Italian' c. un texte hébraïque N MASC A MASC 'a text in Hebrew' d. une lettre hébraïque N FEM A FEM 'a letter in Hebrew' e. un texte hébreu N MASC A MASC 'a text in Hebrew' f. une lettre *hébreu / *hébreue N FEM A FEM 'a letter in Hebrew'

This analysis is corroborated by the fact that the nominal use is again confined to one fixed gender (*cf.* above and section 3.2).

On the contrary, profession terms such as in (5) *fermier* refer (just as the ethnic doublets) to human beings and are therefore susceptible to be employed in both genders according to the sex of the referent. Whereas this is actually the case in (12) *boulanger* (N MASC) – *boulangère* (N FEM) and (5) *fermier* (N MASC) – *fermière* (N FEM), it is a known phenomenon that not all profession terms vary obligatorily in gender according to the sex of the referent. French *le médecin* (DET MASC + N MASC) 'doctor', for example, designates female as well as male doctors, though *la médecin* (DET FEM + N FEM) in the sense of 'female doctor' can nowadays also be found (*cf. Le Petit Robert* 2014). Together with the fact that *médecin* cannot be used as a

full adjective (*cf.* 3.3.2 on the generality condition (Mii)), this constitutes a serious challenge for the multifunctionality of profession terms.

3.3.4 Intermediate summary

Table 1 sums up the patterns observed for (1) to (5) with respect to the question whether they might be multifunctional lexemes rather than instances of conversion and categorial distortion.

Semantic stability (Mi)	Generality (Mii)	Prototypicality (Miii)
yes	no	yes
yes	no	no
yes	yes	yes
yes	no	yes (with very few exceptions)
yes	yes (with very few exceptions)	yes
yes	no	yes
	Semantic stability (Mi) yes yes yes yes yes	Semantic stability (Mi)Generality (Mii)yesnoyesnoyesyesyesyesyesnoyesnoyesnoyesnoyesnoyesnoyesnoyesnoyesno

 Table 1: Multifunctionality in French?

In addition to the French ethnic adjective-noun doublets classified by Kiefer (2011) as multifunctional lexemes, colour terms, too, seem – at least at first sight – to respect all necessary conditions for multifunctionality (cf. table 1, (2)). However, as has been shown in section 3.2, the only doublets in which the categories adjective and noun potentially overlap with respect to their inflectional behaviour are doublets whose nominal forms can be used in both genders. It is from this perspective that colour terms cannot be seen as multifunctional lexical units (cf. sections 3.3.2 and 3.3.3).

4 Concluding Remarks and Outlook

The main purpose of this article was to critically discuss the relevance of the phenomenon of *multifunctionality* for the French lexicon. As was shown in section 3, no multifunctional lexemes could be found among the adjective-noun doublets analysed in this paper, that is (1) quality as well as (2) colour terms, (3) language names and (5) profession terms. Although it is clear that these doublets do not constitute the only semantic subgroups in the French adjective-noun domain, we can safely assume that there are - if any - not many other groups that qualify for multifunctionality for one simple reason: Full morphosyntactic overlap between the categories *noun* and *adjective* (a prerequisite for indifferent usage of an item as an adjective and as a noun)

only exists in the case of nouns that vary in gender contextually just as their adjectival counterparts do. While nouns with variable gender still are prototypical nouns (cf. section 3.1), by far the large majority of French nouns have fixed gender (cf. 3.2 and 3.3.3). From this perspective multifunctionality must by definition be a marginal phenomenon in the adjective-noun domain. This assumption can safely be extended to other categories: As has been pointed out in section 2.1, the verb-noun domain is characterised by the fact that in many cases there are formal hints that speak in favour of a conversion relation between word pairs. For instance, Tribout (2010; 2012) shows that French noun to verb and verb to noun conversion is always stem-based. From the perspective of a truly multifunctional lexeme its employment in category A should not include formal elements (such *e.g.* a stem) that are prototypical only for category B (cf. (Miii)). Besides, the inflectional categories and affixes of nouns and verbs do not overlap. Whereas this is also the main reason for the adjective-verb domain being a bad candidate for multifunctionality, it has to be noticed that non-suffixed adjective-verb pairs of the type calme (A) 'calm' - calmer (V) 'calm down' are, in addition, not subject to a general productive rule (cf. (Mii)).

Yet, there still are domains that are worth being researched for multifunctionality, but have to be left for future research. A much discussed candidate is, for example, the past participle that has been observed to be at the cross-roads between verbal, adjectival and nominal forms in French as well as other (Romance) languages (*cf. e.g.* Laurent 1999 for Latin and Romance languages in general; Evrard 2002 and Rivière 1990 for French; Remberger 2012 mostly for Latin and Italian).

Future research should also take a closer look at ethnic adjective-noun doublets. Whereas in this paper Kiefer's (2011) classification of French ethnic doublets as multifunctional lexemes has been taken for granted, Franz Rainer's example *hébraique* (A) 'Hebrew' that cannot be used as an ethnic noun suggests that there might be other examples challenging the generality condition (Mii). Last but not least, other work on ethnic adjective-noun doublets shows that some contain formal elements that point in the direction of conversion rather than multifunctionality (*cf. e.g.* Roché 2008).

References

Arad M. 2003. "Locality constraints on the interpretation of roots: the case of

Hebrew denominal verbs". *Natural Language & Linguistic Theory* 21: 737-778.

- Bauer L. 2005. "Conversion and the notion of lexical category" in L. Bauer,S. Valera (eds.), Approaches to Conversion/Zero-Derivation. Münster,New York, München, Berlin: Waxmann, 18-30.
- Beauseroy D., M.-L. Knittel. 2007. "Nombre et détermination: Le cas des noms de qualité". *Rivista di Linguistica* 19/2: 231-262.
- Booij G. 1993. "Against split morphology" in G. Booij, J. van Marle (eds.), Yearbook of Morphology 1993. Dordrecht: Kluwer, 27-50.
- Corbin D. 1987. *Morphologie dérivationnelle et structuration du lexique*, vol. 1. Tübingen: Niemeyer.
- Evrard I. 2002. "Le temps, c'est de l'agent! Être + Participe passé: Structure prédicative et référence aspecto-temporelle". *Revue de linguistique romane* 66: 245-260.
- Gauger H.-M. 2011. "Enfin, il vit son quotidien'. Neutrale Adjektivsubstantivierungen im Vorrücken" in S. Dessi Schmid, U. Detges, P. Gévaudan, W. Mihatsch, R. Waltereit (eds.), Rahmen des Sprechens. Beiträge zu Valenztheorie, Varietätenlinguistik, Kreolistik, Kognitiver und Historischer Semantik. Peter Koch zum 60. Geburtstag. Tübingen: Narr, 191-203.
- Gévaudan P. 2007. Typologie des lexikalischen Wandels. Bedeutungswandel, Wortbildung und Entlehnung am Beispiel der romanischen Sprachen. Tübingen: Stauffenburg.
- Goes J. 1999. L'adjectif entre nom et verbe. Paris, Bruxelles: Duculot.
- Kerleroux F. 1996. *La coupure invisible*. *Etude de syntaxe et de morphologie*. Villeneuve-d'Ascq: Presses Universitaires du Septentrion.
- Kiefer F. 2011. "La dérivation impropre". Cahiers de lexicologie 98: 113-121.
- Laurent R. 1999. *Past Participles from Latin to Romance*. Berkeley, Los Angeles, London: University of California Press.

- Lauwers P. 2008. "The nominalization of adjectives in French: From morphological conversion to categorial mismatch". *Folia Linguistica* 42/1: 135-176.
- Lehmann C. 2008. "Roots, stems and word classes". *Studies in Language* 32/3: 546-567.

Le Petit Robert en ligne (2014). Paris: Le Robert.

- Lüdtke J. 2005. *Romanische Wortbildung*. *Inhaltlich diachronisch synchronisch*. Tübingen: Stauffenburg.
- Manova S., W. Dressler. 2005: "The morphological technique of conversion in the inflecting-fusional type" in L. Bauer, S. Valera (eds.), Approaches to Conversion/Zero-Derivation. Münster, New York, München, Berlin: Waxmann, 66-101.
- Marzo D. 2013. "Italian verb to noun conversion: the case of nouns in -a deriving from verbs of the 2nd and 3rd conjugation". *Linguística* 8, 69-87.
- accepted for publication. "Quelques observations sur l'hypothèse de la sousspécification catégorielle: le cas de la conversion dans les langues romanes" in A. Lemaréchal, P. Koch, P. Swiggers (eds.), Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Section 1: Linguistique générale/linguistique romane. Nancy: ATILF.
- Marzo D., B. Umbreit 2013. "La conversion entre le lexique et la syntaxe", in E. Casanova Herrero, C. Calvo Rigual (eds.), Actes del 26é Congrés de Lingüística i Filologia Romàniques (València, 6-11 de setembre de 2010), vol. 3. Berlin: Walter de Gruyter, 565-576.

Noailly M. 1999. L'adjectif en français. Paris: Ophrys.

- Rainer F. 1989. I nomi di qualità nell'italiano contemporaneo. Wien: Braumüller.
- 1993. Spanische Wortbildungslehre. Tübingen: Niemeyer.

- Remberger E. 2012. "Participles and nominal aspect" in S. Gaglia, M.-O. Hinzelin (eds), *Inflection and word formation in Romance languages*. Amsterdam/Philadelphia: Benjamins, 271-294.
- Rivière N. 1990. "Le participe passé est-il verbe ou adjectif?". *Travaux de linguistique et de littérature* 28: 131-169.
- Roché M. 2008. "Structuration du lexique et principes d'économie: le cas des ethniques" in J. Durand, B. Habert & B. Laks (eds.), Congrès Mondial de Linguistique Française - CMLF'08. Paris: Institut de Linguistique Française, 1571-1584.
- Schwarze C. 2012. "Le vert, les jeunes, l'important: aspects sémantiques de la conversion adjectif-nom". *Scolia* 26: 153-170.
- Soares Rodrigues A. 2009. "Portuguese converted deverbal nouns: Constraints on their bases". *Word Structure* 2: 69-107.
- Thornton A.-M. 2004. "Conversione" in M. Grossmann, F. Rainer (eds.), La formazione delle parole in italiano. Tübingen: Niemeyer, 500-553.
- Tribout D. 2010. *Les conversions de nom à verbe et de verbe à nom en français*. Ph.D. Dissertation. Université Denis Diderot (Paris VII), Paris.
- 2012. "Verbal stem space and verb to noun conversion in French". Word Structure 5/1: 109-128.
- Villalba X. 2009. "Definite adjective nominalizations in Spanish" in M.T. Espinal, M. Leonetti, L. McNally (eds.), Proceedings of the IV Nereus International Workshop "Definiteness and DP Structure in Romance Languages". Konstanz: Fachbereich Sprachwissenschaft Universität Konstanz, 137-151.
- Vogel P. 1996. Zur Konversion und verwandten Erscheinungen im Deutschen und in anderen Sprachen. Berlin: Walter de Gruyter.
- Winther A. 1996. "Un point de morpho-syntaxe : la formation des adjectifs substantivés en français". *L'information grammaticale* 68: 42-46.

GENDER AND INFLECTION CLASS AS EVIDENCED BY THE LANGUAGE CONTACT FACTOR: A CASE STUDY ON TWO MODERN GREEK CONTACT INDUCED VARIETIES¹

Dimitra Melissaropoulou University of Patras

Abstract

This paper provides a comparative analysis of nominal loanword integration in two different contact induced varieties of Greek (i.e. Grico and Cappadocian) in order to offer further insights into the notion of gender (from a morpho-semantic viewpoint, i.e. gender assignment) and its relation to the notion of inflection class. By providing an analysis of the general mechanisms (e.g. natural gender, formal correspondences, semantic equivalences, analogy) which account for the integration of loanwords in the examined systems, it is shown that notwithstanding the divergence, grammatical gender splits into its two major primitives, the semantic one relating to sex and animacy and the structural one, i.e. as an inflectional classifier -in correlation with the notion of inflection class- in the organization of nominal classification types, offering further support to the claim that gender is not a purely morphological or a purely semantic category, but a combination of the two. The realization of those two facets, of one, or none of them, is subject to parametric variation depending, especially in contact induced varieties, on the interplay between the grammatical properties of all the involved systems.

1. Introduction

Loanwords, as the most commonly attested language contact phenomenon, have attracted the attention of linguistic research in many different perspectives, touching upon different linguistic subfields. Some of the major questions that are tackled in the study of lexical borrowing involve, among others, the nature of loanwords, the borrowability of different spheres of the

¹ An extended version of this paper is being published in *Open Journal of Modern Linguistics*, 3, 4. (December 2013). The author would like to thank the editors and an anonymous reviewer for comments and feedback.

vocabulary or of different grammatical categories, their adaptation strategies, and their role in the organization of grammar.

In this vein, this paper aims to a comparative analysis of nominal loanword integration in two different contact induced systems of Greek in order to offer further insights the notions of gender from a morpho-semantic viewpoint (i.e. gender assignment) rather than a syntactic one (i.e. gender agreement) and inflection class.

It is true that integration of loanwords in the standard variety has been treated and led to relevant publications (cf. Anastasiadi-Symeonidi 1994; Christophidou 2003 for S(tandard) M(odern) G(reek)). However, the situation is totally different when it comes to its dialectal variation since this is the first attempt to make a comparative analysis of the integration of loanwords in varieties which are in contact with both typologically and genetically divergent linguistic systems.

Our data set involves on the one hand, Cappadocian in contact with the agglutinative Altaic Turkish, while, on the other hand, Grico, in contact with the semi-fusional analytic Indo-European Italian. Dialectal data are extracted from the available written sources (among others Tommasi 1996; Stomeo 1996; Karanastasis 1997; Rohlfs 1977; Filieri 2001; Dawkins 1916; Mavrochalyvidis 1990; Janse forthcoming; Sasse 1992 etc.) and the oral corpora of the Laboratory of Modern Greek dialects at the University of Patras.

The paper is structured as follows: in section 2 basic premises and assumptions on the notions of gender, inflection class, and loanword integration mechanisms are summarized. In section 3, a sketchy description of the sociolinguistic background is offered and all the relevant data are presented accompanied by generalizations on the attested phenomena. In section 4, discussion, specific claims and proposals are put forward in order to account for the commonalities and the particularities of the role and the realization of gender and inflection class in the two divergent language contact situations.

2. Premises

The notion of inflection class has been studied in depth and several approaches (among others Carstairs 1987; Dressler 1987; Carstairs-Mc Carthy 1994; Ralli, 2000, 2006; Corbett 2005, 2007; 2008) have been proposed within different frameworks in order to account for it as a classifier of nouns into different groups based on varied criteria. On the other hand, notwithstanding the respectable relevant literature (among others Corbett 1991, 2005; Corbett and Fraser 2000; Dahl 2000a,b), grammatical gender is still to some extent obscure, especially if one takes into account that, on the

one hand, it complicates morphological production, while, on the other hand, there are languages that do perfectly without it.

One important aspect of the realization of grammatical gender in loanword elements concerns the interaction between grammatical gender and the notion of inflection class. In languages with rich morphology, the notions of gender and inflection are acknowledged to be strongly related (among others Corbett 1991; Aronoff 1994; Dressler & Thorton 1996; Ralli 2000, 2002 etc.). However, grammatical gender cannot be thought of as being identical with a specific inflection class type though there is a frequent correlation between the two categories. It is often the case that from the phonological shape of a word and its gender inflection class can usually be deduced.

There have been proposed totally opposite theses -usually on the basis of a specific linguistic system- on which of the two notions dominates the other². Aronoff (1994: 74) claimed that the gender to class dominance is the 'normal' direction while the opposite, the class to gender dominance, the 'inverted' one. However, a universal principle cannot be established and this relationship admittedly varies cross-linguistically.

With respect to Greek, Ralli (2000, 2002, 2003) following a generative tradition, considers gender as a lexical feature and although she does not underestimate the role of semantics for the assignment of a specific grammatical gender value on the basis of animacy, argues that the role of morphology (related to the processes of inflection, derivation and compounding) is more important in grammatical gender assignment. Although both gender and inflection class provide a type of classification for nouns, they do not coincide, not in all different cases at least. Christophidou (2003: 114) on the other hand, within a natural morphology framework, argues that in Greek there is a mono-directional relationship between gender and inflection class, in the sense that inflection class could be described on the basis of gender³.

Although subject to parametric variation depending on the involved systems, the main mechanisms governing loanword integration are considered to be the following (cf. Ibrahim 1973; Poplack, Pousada & Sankoff 1982; Corbett 1991; Thornton 2001; Winford 2010):

² Unsurprisingly, totally different accounts for the direction of dominance (gender \rightarrow inflection class or vice versa) have been offered for the same language (e.g. Russian cf. Corbett 1991; Aronoff 1994) as well.

³ In SMG gender is argued to have priority over inflection class, since all loans or neologisms are assigned grammatical gender whether inflected or not.

a. The natural gender (sex) of the referent

b. The formal (phonological-structural) shape of the word

c. Analogy to the recipient language suffix

d. Analogy to the recipient language semantic equivalent (semantic analogy)

However, apart from factors reflecting the dynamics-characteristics of the recipient system, Anastasiadi-Symeonidi (1994: 189-190), proposed that when a loan element comes from a gendered donor language, its value may influence the value it will be assigned in the recipient language, while Stolz (2009) advocates that the source language as well may employ special strategies such as the preference for a default gender (see also Kilarski 2003) or for a special gender-noun class.

Let us now examine the dialectal data after a sketchy description of the dialects' sociolinguistic background.

3. Data

3.1 Grico

3.1.1 Sociolinguistic background

The dialectal variety of Grico is spoken in Southern Italy, in the area of Puglia, Salento, widely known as *Grecia Salentina* (cf. Karanastasis 1984). The dialectal enclave of Grico is situated at the heart of Salentino peninsula and consists of nine communities. The sociolinguistic status of this Greekspeaking enclave varied during centuries. Till 80's Grico was in danger of extinction. The last decades, it experiences some revitalization efforts (cf. Caratzas 1958; Profili 1999 α , β), having as a starting point its official recognition as a minority language (1999).

Being spoken for great many centuries in an Italian area (see Minas 1994, 2004; Fanciullo 2001; Manolessou 2005 and references therein for the different opinions with respect to Grico origin, i.e. Ancient Greek vs. Byzantine Greek), Grico was in long term contact with Italian, not only in its standard form (the language of school and media), but in the local Romance varieties as well, (*dialetti salentini*), used in every day speech (street conversations, local commerce), a situation that inevitably limited the sphere of its usage to family situations (cf. Profili 1985; Katsoyannou 1996, 1999). Following Profili (1999 α), speakers of Grico do not advocate a Greek identity. They are Italian citizens and their national identity is Italian. The dialectal varieties constitute for them a link that brings them closer to their Greek neighbors from a viewpoint of mentality and culture, but no other bond is implied in anyway.

3.1.2 Grammatical gender assignment in Grico loanwords

As already mentioned in the previous sections, Grico variety is a threegendered system. It distinguishes between masculine, feminine and neuter nouns. More specifically, Grico distinguishes between masculine nouns in -a, -i, and -o, feminine nouns in -a, and neuter nouns in -o, -i, and -a, as shown in the examples under (1), (2), and (3) respectively.

(1) Masculine nouns in -a, -i, and -o
-a: mina.MASC 'month'
-i: tfuri.MASC 'master'
-o: milo.MASC 'mill'

Grico masculine inflectional markers are reminiscent of but not identical with the corresponding Standard Modern Greek inflectional affixes (*-as e.g. minas* 'month', *-is* e.g. *ciris* 'master' and *-os* e.g. *milos* 'mill' respectively). This is mainly due to final *-s* dropping resulting from the preference of Italiot systems for open (CV) syllables.

(2) Feminine nouns in *-a* -a: jineka.FEM 'woman'

Feminine nouns seem to be confined basically to one group of nouns those in -a, as opposed to SMG and other dialectal varieties where two classes of feminines are distinguished, those in -a (e.g. *jineka* 'woman' and those in -i (e.g. *limni* 'lake'). In Grico variety the vast majority of the former feminine nouns in -i are transferred to the -a group without the reverse tendency being seriously at play.

- (3) Neuter nouns in -o, -i, and -a
 - -o: fsilo.NEU 'wood'
 - -i: gala.NEU 'milk'
 - -a: krovatti.NEU 'bed'

Adaptation of nominal loan elements seems to show a preference to specific gender-inflection class values. More specifically:

a. nominal loan elements ending in -a (from loan feminine forms in -a) are generally assigned a feminine grammatical gender value due to their correspondence with the productive feminine -a declension in the Grico system⁴. E.g.:

⁴ As already mentioned by Newton (1963: 22), the retention of feminines in -a in both Italian and Grico systems facilitates their transference.

(4)	⁵akula.FEM	<	acula.FEM	Salentino
	'eagle'		'eagle'	
	avina.FEM	<	vena.FEM	Italian/Salentino
	'vein'		'vein'	
	t∫ista. FEM	<	cista.FEM	Salentino
	'basket'		'basket'	

a. Nominal loan elements ending in -*i* (mainly from loan masculine forms in -*e* and few from -*i*), for the account of which other mechanisms may also be involved (e.g. suffix addition, *pilaci*.NEU < *pila*.FEM+-*aci* 'must tank' etc.), are generally assigned the neuter grammatical gender value and become members of the -*i* subgroup of nouns. E.g.:

(5)	kapetali.NEU	<	capitale.MASC	Salentino
	ʻpillow' paisi NEU	<	'pillow' paise MASC	Salentino
	'country'		'country'	Sulentino
	picciuni.NEU	<	pecciune.MASC	Salentino
	'dove'		'dove'	
	sapali.NEU	<	sapale.MASC	Salentino
	'hedge'		'hedge'	

b. Nominal loan elements ending in -o (from loan masculine forms in -u or -o) are generally assigned the masculine grammatical gender value. E.g.:

(6)	fiuro.MASC	<	fiuru.MASC	Salentino
	flower		nower	Calantin a
	fundo.MASC 'fond'	<	'fond'	Salentino
	guito.MASC	<	uitu.MASC & gomito.MASC	Salentino Italian
	'elbow'		'elbow'	
	gualano.MASC 'peasant'	<	calanu.MASC 'peasant'	Salentino

⁵ As regards the conventions for transliteration employed throughout the paper, dialectal data are transliterated in broad phonetic transcription, while the corresponding forms in the source systems are exemplified as they appear usually in the sources, using the Latin alphabet.

What can be seen is that from the total of different seven inflectional classes of Grico, nominal loans are adjusted entering three specific ones, one masculine, one feminine and one neuter. Moreover, this preference is not accidental at all. Masculine nouns in -o and feminine nouns in -a correspond to two of the most productive inflection classes both for Standard Italian and Salentino inflectional systems. Relative examples can be seen under (7) below:

(7) Italian productive nominal declensions



The choice of neuter in -i inflection class can be accounted for on the basis of the following: a. it is the most productive Grico inflection class, b. its inflectional marker does not coincide with the markers of the other inflection classes as neuter in -o and -a would do (-o and -a are found correspondingly in masculine and feminine nouns as well) and c. it is phonologically very close to /e/, which characterizes another productive declension in Italian (masculine-feminine nouns in -e (plural in -i) e.g. il paese.MASC 'country'). Crucially, a contrastive look at the Grico vs. Romance nominal subgroups reveals that there are formal (structural and phonological) correspondences between the two groups of systems that cannot but have contributed to the adaptation of nominal loanwords in the specific inflection classes and their assignment of a specific grammatical gender value. Thus, dia-morphemic structural and morphological schemata, in this particular case what we would call dia-classes (cf. Melissaropoulou 2014), are proven to influence morphological adaptation and grammatical gender assignment into the Grico system.

⁶ Few feminine nouns in *-o* can be traced in both Standard Italian and Salentino dialect. E.g. la *mano*.FEM 'hand', *la radio*.FEM 'radio'.

⁷ In Salentino dialect, the mid vowels and /e/ and /o/ are raised into /i/ and /u/ respectively when found in final position (cf. Maiden & Parry 1997).

According to Anastasiadi-Symeonidi (1994: 189-190), when a loan element comes from a gendered donor language, its value influences the value it will be assigned in the recipient language unless other conditions are in operation. In our case study we would add that the formal correspondences as realized through a specific inflectional marker, which bears a specific grammatical gender value, influence integration (both the grammatical gender assignment and the inflection class membership) of loanwords in the recipient system.

Moreover, apart from the formal shape of the word, which seems to play a very crucial role for the vast majority of loanword elements and is highly ranked, there are some other mechanisms involved in grammatical gender assignment of loanwords. These are as follows:

a. The natural gender of the referent. The phonological and structural correspondences can be biased and a different grammatical gender value can be assigned when human nouns or more generally animate nouns⁸ are involved, since in this case nouns have to bear the grammatical gender value that matches their sex (masculine when the referent is male and feminine when female). For example nouns in *-i* are assigned the neuter grammatical gender value when non human and the masculine grammatical gender value when human males. You can see the examples under (8) below:

(8)	paisi.NEU.	< paise.MASC	Salentino
	country	country	
	vucceri.MASC 'butcher'	< ucceri.MASC 'butcher'	Salentino
	sarturi.MASC 'tailor'	< sartore.MASC 'tailor'	
	speciali.MASC 'pharmacist'	< speciale.ADJ 'particular'	Italian/Salentino

b. Analogy to the recipient system suffix. The status of suffixes as heads that are marked for a specific gender value and attach to a specific inflectional marker plays also an important role in integration of loanword elements, offering further support to the claim that gender is a lexical feature (cf. Spencer 1999, Ralli 2003) that actively participates in word-formation processes. You can see the examples below:

⁸ Some domesticated animals bear the grammatical gender value that matches their sex as well.

(9)	a. vardedhhi.NEU < varda.FEM	Salentino
	NOTE: the suffix <i>-eddhi</i> in the recipient system b grammatical gender value.	ears the neuter
	b. kasciuna.MASC < cascia.FEM 'big box' 'box'	Salentino

NOTE: the suffix *-una* in the recipient system bears the masculine grammatical gender value.

However, these formations are not abundant and it is often the case that both simple and derived loan forms are found in the recipient system.

c. Although marginally, analogy to the recipient system semantic equivalent⁹. In few cases the nominal loan does not bear the grammatical gender value that would be expected given the above mentioned mechanisms, but it acquires the grammatical gender value of its semantic equivalent in the recipient system. E.g.:

(10)	fikato.NEU 'liver'	<	fegato.MASC 'liver'	Salentino	
	NOTE: the Grico	semanti	c equivalent sikoti is neuter.		
	faratzo.MASC 'bulb'	<	farazza.FEM 'bulb'	Salentino	
	NOTE: the Grico semantic equivalent volvos is masculine				
	spirlingoi.MASC	<	perlangoi.F	Salentino	
	NOTE: the Grico	semanti	cequivalent <i>melisofao</i> is masculi	ine	

Generalizing on the mechanisms governing integration of loanwords in Grico, we would say that natural gender is ranked in the highest position even though the formal shape of the word determines integration for the vast majority of nominal loanwords, since the latter can be biased and a different grammatical gender value can be assigned when human nouns are involved. Analogy to the recipient system suffix and semantic analogy are operative in a very small number of loanwords thus are not thought of as prevailing but

⁹ Semantic analogy or concept association in Corbett's (1991) terms.

rather as additional mechanisms. Thus, grammatical gender assignment is thought to be predictable only in those cases where natural gender is involved. In all the other cases formal (phonological-structural) correspondences prove to be the most powerful mechanism governing integration of loanwords. From the total of seven different inflectional subgroups in Grico, nominal loans are adjusted entering three specific ones, revealing that formal correspondences between the involved systems contributed to the integration of nominal loanwords in the specific inflection classes and their assignment of a specific grammatical gender value. Let us now turn to Cappadocian.

3.2 Cappadocian

3.2.1 Sociolinguistic background

Cappadocian came under the Turkish influence during the late byzantine period for the first time in the 11th century after the Seljuk invasion and subsequently in the 14th century after the conquest of Asia Minor by the Ottoman Turks. It was spoken till 1923 (i.e. till the exchange of populations that followed the treaty of Lausanne in the former Asia Minor (today's central Turkey) in an area that covered 32 communities approximately. The dialect is subdivided into two basic groups, North and South Cappadocian (cf. Dawkins 1916) and an intermediate one, namely Central Cappadocian (cf. Janse forthcoming) showing intra-dialectal divergence¹⁰. Today it is spoken by descendants of Cappadocian refugees (second and third-generation refugees) in several parts of Northern Greece (Kavala, Alexandroupoli, Kilkis, Thessaloniki, Karditsa, Volos, Larisa).

Cappadocian is often used in the literature as a prototypical example of 'heavy borrowing' in terms of Thomason & Kaufman's borrowing scale, referring to 'overwhelming long-term cultural pressure' (Thomason & Kaufman 1988:50). It should be noted that although Cappadocian is originally a Greek variety and its basic morphological structure is fusional, it displays some agglutinative patterns due to language contact with Turkish (cf. Dawkins 1916; Janse, 2004, 2009, forthcoming).

3.2.2 Grammatical gender assignment in Cappadocian loanwords

The situation in Cappadocian seems to be quite differentiated compared to that in Grico. In this case, the dominant language, Turkish, is both genetically

¹⁰ The division of Cappadocian into zones is not clear cut since for example Northeast Cappadocian system is in some aspects similar to that of Axó's which belongs to Central Cappadocian.

and typologically divergent, namely it is a non Indo-European, Altaic, agglutinative, genderless language.

As already acknowledged in the relevant literature (among others Dawkins 1916; Janse 2004, forthcoming), Cappadocian holds a prominent position compared to all other Modern Greek dialects and SMG, since it is characterized by the following innovations: a. the distinction between animate and inanimate nouns in North and Central Cappadocian (to some extent), b. the progressive loss of gender distinctions, especially in South Cappadocian (cf. Dawkins 1916; Janse 2004, forthcoming), and c. the emergence of a generalized agglutinative declension, innovations that are relevant for the purposes of this paper.

Our presentation of integration of loanwords in Cappadocian follows the geographical subdivision into North, Central, and South Cappadocian in order to be able to capture the intra-dialectal divergence and account for it in terms of mirroring the gradualness of linguistic change towards a specific direction: the establishment of a genderless system.

Crucially, in Cappadocian the original categorization of nouns into different subgroups, i.e. inflection classes, based on their different inflectional endings in combination with their different grammatical gender values, as shown in (11) below, is retained to some extent only in the North Cappadocian zone (and much less to the central Cappadocian zone). The original subgrouping of Cappadocian inflection can be seen from (11) to (13) below:

(11)	Masculine nouns in -os, -is, and -as			
	- os :	aθropos	'man'	
	-is:	kleftis	'thief'	
	-as:	papas	'priest'	

(12)	Femin		
	-a:	neka	'woman'
	-i:	nif(i)	'bride'

(13)	Neuter nouns in <i>-i</i> , <i>-a</i> and <i>-o</i> :			
	-i:	fti	'ear'	
	-a:	konizma	'icon'	
	-0 :	metapo	'forehead	

More specifically, in the admittedly 'less corrupted' North Cappadocian zone (in the words of Dawkins (1916:112)), and to Axó, Central Cappadocian zone, to a lesser extent, nouns are assigned a specific grammatical gender value on the basis of the categorical semantic distinction of animacy. Human nouns seem to bear a masculine or feminine grammatical gender value, while

non human nouns become neuter, which marks the lack of gender. See examples under (14):

(14)	a. t∫obanus.MASC 'shepherd'	(Delmesó, North Cappadocian)	< çoban.Ø 'shepherd'
	b. patiʃahos.MAS0 'king'	C (Delmesó, North Cappadocian)	< padišah.Ø 'king'
	c. herifos. MASC 'man'	(Axó, Central Cappadocian)	< herif.Ø 'man'
	d. γərəxos.MASC 'scorpion'	(Axós, Central Cappadocian)	< kuyruk.Ø 'scorpion'
	e. balduza.FEM 'sister-in-law'	(Axós, Central Cappadocian)	< baldız.Ø 'sister-in-law'
	f. tʃiftʃis.MASC 'farmer'	(Malakopí, North Cappadocian)	< çiftçi.Ø 'farmer'
	g. ast∫is.MASC 'cook'	(Malakopí, North Cappadocian)	< aşçı.Ø 'cook'
	h. γ/goltʃ/dʒis.MA 'guard'	SC (Axós Central Cappadocian)	< kolcu.Ø 'guard'

As shown in (14) above, human male loanwords ending in a consonant in Northern Cappadocian are assigned a masculine grammatical gender value and are mainly integrated into the *-os* subgroup of nouns (examples 14a-d), while human loanwords ending in a vowel or loan agentive nouns in *-cI*¹¹, mainly into the *-is* (very few ending in *-a*(*s*) e.g. *arkadaf* 'friend' into the *-as* subgroup) subgroup of nouns (examples 14f-h).

On the contrary, non animate nouns are integrated into the originally neuter subgroup, the one in -i, whether consonant final -which constitutes the vast majority of Turkish loan elements from Turkish- or vowel final and attach to the originally neuter generalized -ja -ju markers (which is usually called in the literature 'agglutinative inflection' cf. Dawkins 1916; Janse 2004, forthcoming etc.). It should be noticed that in this case neuter subgrouping

¹¹ In Turkish, the suffix -cI is subject to vowel harmony whereby the final vowel can equally appear as 1/ul/, u/u/ or $\ddot{u}/y/$ as well, depending on the preceding vowel.

marks the characteristic [-human or -animate] and more generally the lack of gender. E.g.:

(15)	tfadir(i).NEU (Delmesó, North Cappadocian)	< çadır.Ø
	'tent'	'tent'
	diken(i).NEU (Delmesó, North Cappadocian)	< diken.Ø
	'thorn'	'thorn'
	varmax(i).NEU (Delmesó, North Cappadocian)	< parmak.Ø
	'finger'	'finger'
	γazan(i).NEU (Axós, Central Cappadocian)	< kazan.Ø
	'copper'	'copper'
	irmax(i).NEU (Axós, Central Cappadocian)	< irmak.Ø
	'river'	'river'
	yara.NEU (Axós, Central Cappadocian)	< yara.Ø
	'wound'	'wound'

Crucially, the addition of this innovative categorical distinction in the Cappadocian system – which is absent from Turkish – is not uniform in all Cappadocian subvarieties. Signs of de-systematization appear already in the Central Cappadocian zone. In Axó, human loanword elements marked as masculine on the basis of their animacy (see the examples c., d., e. under 14) co-occur with loanwords, which, although bearing the same semantic characteristic, are marked as neuter. E.g.:

(16)	arkada J.NEU (Axós, Central Cappadocian)	<	arkadaş
	'friend'		'friend'
	musafir.NEU (Axós, Central Cappadocian)	<	misafir
	'guest'		'friend'
	bektſis.NEU (Axós, Central Cappadocian)	<	bekçi
	'field guard'		'field guard'
	misedzis.NEU (Axós, Central Cappadocian)	<	meşeçi
	'lumberjack'		'lumberjack'

This instability of grammatical gender assignment in loanwords can be seen as a transitory stage (cf. Poplack & Sankoff (1984: 124)) paving the way towards the re-structuring of the specific category as exemplified in South Cappadocian.

In the more 'corrupted' in the words of Dawkins's (1916: 112) South Cappadocian zone, this distinction appears to become completely extinct and all nouns, both loan and native elements, either plus or minus human are formally neuter, marking the lack of gender, establishing thus a totally genderless system. E.g.:

(17)	tſoban.NEU (Ulağáç, Fertek, South Cappadocian)	< çoban.Ø
	'shepherd'	'shepherd'
	padisax.NEU (Ulağáç, South Cappadocian)	< padišah.Ø
	'king'	'king'
	baldəza.NEU (Ulağáç, South Cappadocian)	< baldız.Ø
	'sister-in-law'	'sister-in-law'
	bizelik(i).NEU (Ulağáç, South Cappadocian)	< bilezik.Ø
	'bracelet'	'bracelet'

What can be seen in Cappadocian is that a totally new categorical distinction emerges, that of animacy as an inflectional classifier, a distinction that is totally absent both from Greek and Turkish. Assuming thus, that intradialectal variation mirrors the gradualness of linguistic change, the addition of this extra category of animacy, present in North and -to some extent- in Central Cappadocian but extinct in the South Cappadocian zone could best, in our view, be accounted for as a temporary resolution, a repair strategy, one of the greater or lesser re-arrangements in the structure of the system in order to pave the way to its reshaping according to the new dynamics and tendencies, due to the prevailing -but not exclusive- influence of the dominant Turkish language; namely towards acquiring a totally genderless status. In this vein, all loanwords are accommodated as neuters - which marks the lack of gender - and are inflected via the attachment to the generalized - originally most productive neuter – inflectional -ja - ju markers, the so called 'agglutinative inflection'. These markers as already shown in Karatsareas (2011), Melissaropoulou (2014) form part of the one and only inflectional paradigm that tended to generalize and substitute the several original subgroups of nouns¹².

4. Discussion

¹² Table 1	The uniform	inflectional	paradigm	of Cappadocian
raule r.	The unitorni	minectional	paradigin	or Cappadooran

	Singular	Plural	Singular	atropos	Plural	atropos
			'man'		'man'	
Nom	Ø	-ja	atropos		atropoz-ja	
Gen	-ju	-(ja)ju	atropoz-ju		atropoz-(ja)-ju
Acc	Ø	-ja	atropos		atropoz-ja	

(adapted from Melissaropoulou 2014: 327)

Although our data involve two totally divergent case studies, important generalizations focusing both on commonalities and particularities can arise.

Emphasizing commonalities, in both cases what seems to play a very important role in grammatical gender assignment as part of the morphological adaptation process is the semantic feature of animacy. Either in contact between gendered languages or between gendered vs. genderless systems, the most compelling mechanism at work, the one that could be argued to have a universal basis is the correspondence with natural gender, offering further support to the claim that gender has a semantic basis/core (cf. Aksenov 1984: 17-18). In this sense, one of the most important functions of gender seems to be the grammatical encoding of sex and animacy as a means of nominal classification.

Apart from the notion of animacy, the other important facet of gender is the formal one, i.e. as an inflectional classifier in the organization of nominal classification types. As illustrated by the data on Grico, apart from the compelling mechanism of animacy, the other important parameter governing grammatical gender assignment in loanwords is the formal correspondences. This factor seems to be activated – mainly but not exclusively – when structural compatibility among the systems in contact is involved. Both Italian and Grico are gender-inflection class systems, notwithstanding that the grammatical gender values are not identical in both of them. What seems to play a crucial role is that the gender-inflection class classification is present in both systems. In the case of loanword integration into Grico, the notion of gender is strongly related to the notion of inflection class since assignment of gender entails membership in a specific inflection class.

On the contrary, in Cappadocian where contact between an originally genderinflection class system and a genderless agglutinative one non displaying inflectional classes is at play, the morphological facet is not realized, only the semantic one, based on animacy which serves -at least at a particular stage- as a classificator of loanwords into the different inflection classes and takes over the formal function as well. In this case, animacy takes over the classificatory function of integrating humans to the inflection classes that originally contained human nouns, i.e. where marked as masculine or feminine.

Crucially, the progressive loss of the different grammatical gender values and the temporary resolution strategy of the animacy based classification seem to go hand in hand with the progressive loss of the different inflection classes. The direction towards the establishment of a completely genderless system coincides with the direction towards the establishment of a single and uniform inflection class for nouns, remarkably the one coinciding with the most productive neuter inflection class. Admittedly, there is no general consensus in the relevant literature on the sources of these innovations, i.e. the loss of grammatical gender distinctions and of the different inflection classes with the development of 'agglutinative' inflection in Cappadocian. Previous research has overwhelmingly accounted for them as instances of contact-induced change, (see, among others, Thomason & Kaufman (1988: 215-222); Johanson (2002:104), Winford (2005: 402-409), (2010: 181)) resulting from the influence of Turkish. Karatsareas (2011: 8-9), on the other hand, treated them in strictly language internal terms, i.e. as internal developments dating back to a linguistic precursor of the Modern Asia Minor Greek dialects.

Loss of grammatical gender and of the different inflection classes were accounted for by Melissaropoulou (2014) as contact-induced simplification phenomena (cf. Nichols 1992; Trudgill 2009, 2011) that were adjusted to the system main intra-linguistic characteristics and tendencies aiming to balance out the system. In this spirit, the addition of the extra category of animacy is seen as a temporary repair complexification strategy paving the way towards the simplification of inflectional organization under the influence of Turkish.

Whatever the primary or the secondary cause of change, it seems that the loss of the one category – gender – in Cappadocian entails the loss of the other one as well since the basic function - contribution in the organization of grammar i.e. classification of nouns need not be served anymore, paving the way towards grammar simplification. It is true that complex morphology is not a sufficient condition for the realization of grammatical gender, since there are languages with complex agglutinating morphology and no grammatical gender. Crucially, things seem to go the other way around offering further support to the claim that the distinction of different inflectional classes entails the realization of different grammatical gender values, while genderless languages generally tend to have no distinction of (macro)classes (see Dressler & Thorton 1996: 26), leading to a simpler morphology. Further support to this claim is offered by data on Slavonic languages, Germanic languages and many of the German dialects, Bantu languages or English (cf. Dressler et al. 1996; Corbett 1991; Hickey 1999).

In the case of Grico, on the other hand, the strong correlation between gender and inflection class is strongly corroborated in its positive aspect, since it was shown that in loanword integration a specific form (phonological shape) entails assignment of a specific grammatical gender value and membership in a specific inflection class. Crucially, in loanword integration the mismatches between gender assignment and inflection class membership are minimal, establishing a one to one correspondence between a specific gender value and a specific inflection class. Thus, our findings show that in Grico loanwords gender has priority over inflection class, i.e. follows the 'normal' direction in terms of Aronoff (1994: 74), corroborating the claim that inflection class membership depends on extra-morphological factors such as gender and phonology (cf. Wurzel 1984; Aronoff 1994). Furthermore, our data seem to verify only partially the established claims in the literature that the source language as well may employ special strategies such as the preference for a default gender (see Kilarski 2003; Stolz 2009) or for a special gender-noun class. Our data show that in cases of structural compatibility among the involved systems, i.e. the case of Grico, the formal correspondences take priority over a default gender. In the case of Cappadocian, on the other hand, the situation seems to be more complicated in the sense that Cappadocian adopts the neuter, marking the lack of gender, under the influence of the dominant genderless Turkish language, indicating thus a kind of preference for a default gender value even in its negative realization. On the other hand, the emergence of a unique inflection class, known as the 'agglutinative inflection', is viewed as well as a consequence of the loss of grammatical gender under the Turkish influence. However, the preference for the prevalence of this specific neuter class over the other available ones appeals again to reasons of formal correspondences between the original members of this class and the vast majority of Turkish loanwords (since, after final unstressed -i deletion both native and loan words end in a consonant).

Lastly, based on our data we cannot postulate that there are different or additional mechanisms which apply to the assignment of borrowings and not to that of native words. All operative mechanisms (animacy, formal correspondences, analogy) can apply equally efficiently in both native and loanword elements both in cases of structural compatibility and incompatibility among the systems involved (cf. Christophidou 2003; Ralli 2005). In other words, recipient systems seem to allude to their available mechanisms, and try to treat and incorporate loanword elements with the same means as native words. These findings offer further support to the status of gender and inflection class as integral parts of the organization of grammar and not just as the burden of diachrony or as what language evolution has not make disappear yet.

References

- Aikhenvald, A. Y. 2000. Classifiers: A typology of noun categorization devices. Oxford: Oxford University Press.
- Aksenov, A. T. 1984. "K Probleme èkstralingvističeskoj motivacii grammatičeskoj kategorii roda". *Voprosy jazykoznanija* 1, 14-25.
- Anastasiadi-Symeonidi, A. 1994. Νεολογικός δανεισμός της Νεοελληνικής [Neological borrowing in Modern Greek]. Thessaloniki.

- Aronoff, M. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Cambridge. MA/London: MIT Press.
- Bichakjian, B.H. 1999. "Language evolution and the complexity Criterion". *Psycologuy*, 10 (033).
- Braunmüller, K. 2000. "Gender in North Germanic: A diasystematic and functional approach". *Unterbeck & Rissanen*, 25-53.
- Caratzas, S. C. 1958. L'origine des dialectes néogrecs de l'Italie méridionale. Paris: Les Belles Lettres
- Carstairs, A. 1987. Allomorphy in inflection. London: Croom Helm.
- Carstairs-McCarthy, A. 1994. "Inflection Classes, Gender, and the Principle of Contrast". *Language* 70, 737–787.
- Christophidou, A. 2003. "Γένος και Κλίση στην Ελληνική (Μια Φυσική Προσέγγιση)" [Gender and Inflection in Greek (A Natural Approach)] in A. Anastasiadis-Symeonidis, A. Ralli & D. Cheila Markopoulou (eds.), *To Γένος* [Gender]. Athens: Patakis, 100-131.
- Corbett, G. 1991. Gender. Cambridge: Cambridge University Press.
- Corbett, G. 2005. "The number of genders" in M. Haspelmath, M. S. Dryer,D. Gil & B. Comrie (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 126-129.
- Corbett, G. 2007. "Gender and Noun Classes" in T. Shopen (ed.), Language Typology and Syntactic Description: III: Grammatical categories and the lexicon. Cambridge: Cambridge University Press, 241-279.
- Corbett, G. 2008. "Canonical Inflectional Classes" in F. Montermini, G. Boyé & J. Tseng (eds.), Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux. Somerville, MA, USA: Cascadilla Proceedings Project, 1-11.
- Corbett, G. & N. Fraser. 2000. "Default genders" in B. Unterbeck, M. Rissanen, T. Nevalainen & M. Saari (eds.), *Gender in Grammar and*

Cognition (Trends in Linguistics: Studies and Monographs 124). Berlin: Mouton de Gruyter, 55-97.

- Dahl. Ö. 2000a. "Animacy and the notion of semantic gender" in B. Unterbeck, M. Rissanen, T. Nevalainen and M. Saari (eds.), *Gender in grammar and cognition*, *I: approaches to gender*. Berlin: Mouton de Gruyter, 99–115.
- Dahl. Ö. 2000b. "Elementary gender distinctions" in B. Unterbeck, M. Rissanen, T. Nevalainen and M. Saari (eds.), Gender in grammar and cognition, II: manifestations of gender. Berlin: Mouton de Gruyter, 577–593.
- Dawkins, R. M. 1916. Modern Greek in Asia Minor: a study of the dialects of Sílli, Cappadocia and Phárasa with grammar, texts, translations and glossary. Cambridge: Cambridge University Press.
- Dressler, W. 1987. "Word formation (WF) as part of natural morphology" *in*W. Dressler, W. Mayerthaler, O. Panagl, W. Wurzel, (eds.), *Leitmotifsin Natural Morphology*. Amsterdam: Benjamins, 99-126.
- Dressler, W. U. & Thornton. A. M. 1996. "Italian nominal inflection". *Wiener Linguistische Gazette* 57-59: 1-26.
- Dressler, R., Drazyk, D., Dziubalska K. & Jagla. E. 1996. "On the earliest stages of acquisition of Polish declension". Wiener Linguistische Gazette 53-54: 1-21.
- Filieri, V. G. 2001. Ivò milò to griko. Βασική μέθοδος των Ελληνικών του Salento σε σύγκριση με τα Νέα Ελληνικά [I speak Grico. A basic method on Greek of Salento in comparison with Standard Modern Greek]. University of Ioannina, Center of Greek Language and Culture. INTERREG II, Greece - Italy - Action 5.4.
- Fanciullo, F. 2001. "On the origins of Modern Greek in Southern Italy" in
 M. Janse, B. Joseph & A. Ralli (eds.), Proceedings of the 2nd

International Conference on Modern Greek Dialects and Linguistic Theory. Patras-Unversity of Patras, 67-77.

- Haugen, E. 1950. The analysis of linguistic borrowing. *Language* 26, 210-331.
- Haugen, E. 1953. The Norwegian Language in America: A Study in Bilingual Behavior. Vol. 1: The bilingual community; Vol II. The American dialects of Norwegian. Bloomington: Indiana University Press.
- Hickey, R. 1999. "On the phonology of gender in Modern German" in Rissanen, M. and B. Unterbeck (eds.), *Gender in Grammar and Cognition.* Berlin: Mouton de Gruyter, 621-663.
- Ibrahim, M. H. 1973. *Grammatical gender: Its origin and development*. The Hague: Mouton.
- Janse, M. 2004. "Animacy, definiteness and case in Cappadocian and other Asia Minor Greek dialects". *Journal of Greek Linguistics* 5, 3-26.
- Janse, M. 2009. "Greek-Turkish language contact in Asia Minor". *Études Helléniques/Hellenic Studies* 17, 1: 37-54.
- Janse, M. forthcoming. "Cappadocian". In Tzitzilis (ed.), Νεοελληνικές Διάλεκτοι [Modern Greek Dialects]. Thessaloniki: INS (Manolis Triantafyllides Foundation).
- Johanson, L. 1992. *Strukturelle Faktoren in türkischen Sprachkontakten* [Structural factors in Turkish language contacts] Stuttgart: Franz Steiner.
- Johanson, L. 2002. "Contact-induced linguistic change in a code-copying framework" in M. C. Jones and E. Esch (eds.), Language Change: The Interplay of Internal, External and Extra-linguistic Factors. (Contributions to the Sociology of Language, 86). Berlin: Mouton de Gruyter, 285–313.

- Karanastasis, A. 1984. Ιστορικόν Λεζικόν των ελληνικών ιδιωμάτων της Κάτω Ιταλίας [Historical Lexicon of the Greek dialects of Southern Italy]. Vol 1. Athens: Academy of Athens.
- Karanastasis, A. 1997. Γραμματική των Ελληνικών Ιδιωμάτων της Κάτω Ιταλίας [Grammar of the Greek dialects of Southern Italy]. Athens: Academy of Athens.
- Katsoyannou, M. 1996. "Ελληνικά της Κάτω Ιταλίας: μορφολογία των ονομάτων και εξέλιξη του κλιτικού συστήματος" [Greek in Southern Italy: morphology of nouns and evolution of the nominal system]. *Studies in Greek Linguistics 17*, 328-341. Thessaloniki: Afoi Kyriakidoi.
- Katsoyannou, M. 1999. "Ελληνικά της Κάτω Ιταλίας: η κοινωνιογλωσσολογική άποψη" [Greek in Southern Italy: the sociolinguistic perspective] in A. Moser (eds.), Greek Linguistics 97' Proceedings of the 3rd International Conference on Greek Language [Ελληνική Γλωσσολογία '97: Πρακτικά του Γ' Διεθνούς Συνεδρίου για την ελληνική γλώσσα]. Athens: Greek Letters, 605-613.
- Karatsareas, P. 2011. A study of Cappadocian Greek Nominal morphology from a diachronic and dialectological perspective. Ph.D. dissertation. University of Cambridge.
- Kilarski, M. 2003. "Gender assignment in Danish, Swedish and Norwegian: a comparison of the status of assignment criteria". *Nordlyd* 31.2, 261-274.
- Maiden, M. & Parry, M. (eds.). 1997. The dialects of Italy. London: Routledge.
- Mavrochalyvidis, G. & Kesisoglou. I. I. 1960. Τὸ γλωσσικὸ ἰδίωμα τῆς Ἀζοῦ [The dialect of Axos]. Athens: Institut Français d'Athènes.

- Mavrochalyvidis, G. 1990. Η Αξό της Καππαδοκίας [Cappadocian Axo] Vol.
 2. Athens: Orestis, Prodromos & Anthoula G. Mavrochalyvidi.
- Manolessou, I. 2005. "The Greek dialects of Southern Italy: an overview". *Cambridge Papers in Modern Greek* 13, 103-125.
- McWhorter, J. H. 2001. "The world's simplest grammars are creole grammars". *Linguistic Typology*, *5*, 125–166.
- Melissaropoulou, D. 2014. "Reorganization of grammar in the light of the language contact factor: a case study on Grico and Cappadocian" in M. Janse, B. Joseph, A. Ralli & M. Bağrıaçık (eds.), Proceedings of the 5th International Conference of Modern Greek Dialects and Linguistic Theory, Ghent 20-22 September 2012, 311-334. Available on line at

http://lmgd.philology.upatras.gr/en/research/downloads/MGDLT5_proc eedings.pdf

- Minas, K. 1994. Η γλώσσα των δημοσιευμένων μεσαιωνικών ελληνικών εγγράφων της Κάτω Ιταλίας και της Σικελίας [The language of published medieval manuscripts of South Italy and Sicily]. Athens: Academy of Athens.
- Minas, K. 2004. Μελέτες Νεοελληνικής Διαλεκτολογίας [Studies in Modern Greek Dialectology]. Athens: Typothito.
- Newton, B. 1963. "The grammatical integration of Italian and Turkish substantives into Modern Greek". *Word* 19, 20-30.
- Nichols, J. 1992. *Linguistic diversity in space and time*. Chicago: Chicago University Press.
- Poplack, S. & Sankoff. D. 1984. Borrowing: The synchrony of integration. Linguistics 22, 99-135.
- Poplack, S., Pousada, A. & Sankoff. D. 1982. "Competing influences on gender assignment: variable process, stable outcome". *Lingua* 57, 1-28.

- Profili, O. 1985. "La romanisation d' un grec parler de l' Italie du Sud par les parlers romans environants". Actes du XVIIe Congrés International de Linguistique et de Philologie romans, Aix-en-Provence 29 août – 3 septembre 1983. Aix-en-Provence: Université de Provence, 129-139.
- Profili, O. 1999α. "Η Ελληνική στη Νότια Ιταλία" [Greek in Southern Italy] in A.-Ph. Christidis (ed.), Dialectal enclaves of the Greek Language [Διαλεκτικοί θύλακοι της ελληνικής γλώσσας]. Athens: YPEPTH– Centre for Greek Language, 31-37.
- Profili, O. 1999β. "Η αναζωογόνηση της grico στην Grecia Salentina" [Revitalization of Grico in 'Grecia Salentina'] in A.-Ph. Christidis (ed.), Dialectal enclaves of the Greek Language [Διαλεκτικοί θύλακοι της ελληνικής γλώσσας]. Athens: YPEPTH– Centre for Greek Language, 47-54.
- Ralli, A. 2000. "A feature-based analysis of Greek nominal inflection". Γλωσσολογία/Glossologia 11-12, 201-227.
- Ralli, A. 2002. "The role of morphology in gender determination: evidence from Modern Greek". *Linguistics* 40, 3: 519-551.
- Ralli, A. 2003. "Ο καθορισμός του γραμματικού γένους στα ουσιαστικά της νέας ελληνικής" [Grammatical gender determination in Modern Greek nouns] *in* Anna Anastassiadis-Symeonidis, Angela Ralli and Despina Chila-Markopoulou (eds.), *Το Γένος* [Gender]. Athens: Patakis, 57-99.
- Ralli, A. 2006. "On the Role of Allomorphy in Inflectional Morphology: Evidence from Dialectal Variation" in G. Sica (ed.), Open Problems in Linguistics and Lexicography. Milano: Polimetrica, 123-151.
- Rolhfs, G. 1977. *Grammatica Storica dei Dialetti Italogreci*. Munchen: C.H. Beck'sche Verlagsbuchhandlung.

- Sankoff. G. 2001. "Linguistic Outcomes of Language Contact" in P. Trudgill, J. Chambers & N. Schilling-Estes (eds.), *Handbook of Sociolinguistics*. Oxford: Basil Blackwell, 638-668.
- Sasse, H. J. 1992. "Language decay and contact-induced change: similarities and differences" in M. Brenzinger (ed.), Language death. Berlin/New York: Mouton de Gruyter, 59-80.
- Spencer, A. 1999. Gender as an inflectional category. *Essex Research Reports in Linguistics* 25, 35-72. Essex: University of Essex.
- Stomeo, P. 1996. *Racconti greci inediti di Sternatia*. Lecce: Edizioni "La nuova Ellade".
- Stolz, C. 2009. "Loan word gender in Maltese, with a special focus on Gender Copy" in B. Comrie, R. Fabri, E. Hume, M. Mifsud, T. Stolz and M. Vanhove (eds.), *Introducing Maltese linguistics*. Proceedings of the first international conference on Maltese linguistics. Amsterdam/Philadelphia: John Benjamins, 321-353.
- Thomason, S.G., & Kaufman. T. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Tommasi, S. 1996. Katalisti o Kosmo. Ghetonia: Kalimera.
- Thornton, A. 2001. "Some Reflections on Gender and Inflectional Class Assignment in Italian" in C. Schaner-Wolles, J. Rennison, and F. Neubarth (eds.), *Naturally! Linguistic Studies* in Honour of Wolfgang Ulrich Dressler Presented on the Occasion of his 60th Birthday. Torino: Rosenberg & Sellier, 479-487.
- Trudgill, P. 1999. "Language contact and the function of linguistic gender". *Poznań Studies in Contemporary Linguistics* 35:133-152.
- Trudgill, P. 2009. "Sociolinguistic typology and complexification" in S. Geoffrey, D. Gil & P. Trudgill (eds.), Language Complexity as an Evolving Variable. Oxford: Oxford University, 98-109.

- Trudgill, P. 2011. Sociolinguistic Typology: Social Determinants of Linguistic Complexity (Oxford Linguistics). Oxford: Oxford University Press.
- van Coetsem, F. 1988. Loan phonology and the two transfer types in language contact. Dordrecht: Foris.
- Winford, D. 2005. "Contact-induced changes: classification and processes". *Diachronica* 22, 373-427.
- Winford, D. 2010. "Contact and borrowing" in Raymond Hickey (ed.), The Handbook of Language Contact. Malden, MA/Oxford: Wiley-Blackwell, 170-187.
- Wurzel, W. 1984. *Flexionsmorphologie und Naturlichkeit*. Berlin: Akademie-Verlang.

FROM LATIN TO OLD SPANISH: ON THE POLYSEMY OF DENOMINAL PARASYNTHETIC VERBS PREFIXED WITH A-¹

Isabel Pujol Payet Universitat de Girona, Spain

Abstract

The main aim of this paper is to examine the word formation of denominal parasynthetic verbs with prefix *a*- in Old Spanish. The analysis relies on the lexical semantics viewpoint of the Generative Lexicon. I argue that the polysemy in denominal parasynthetic verbs can essentially be attributed to the semantic features of the nominal stem. Regarding the data under study, the paper focuses on the verbs contained in Nebrija's *Vocabulario* (1495). This information is compared with the one provided by the Spanish textual corpora CORDE, CE and CDH.

1 Introduction

Since the seventies several authors have discussed the extraordinary polysemy of Spanish denominal parasynthetic verbs (*cf.* Reinheimer-Rîpeanu 1974, Rainer 1993, Serrano Dolader 1995, Rifón 1997, and Lavale 2013, among others).² The last grammar edited by the Spanish Royal Academy still states: "Es compleja y múltiple la aportación semántica del sustantivo a la interpretación de estos verbos derivados", vid. RAE (2009: 607). The aim of this paper is to provide a historical account of how these verbs were generated in Late Latin and how they became the pattern followed by further Spanish neologisms. In view of this, the paper mainly focuses on the verbs contained in Nebrija's *Vocabulario* (1495), the first bilingual dictionary of

¹ Research for this paper has been supported by the *Ministerio de Ciencia e Innovacion* through projects FFI2011-29440-CO3-02 (Periferias y cambio lingüístico: descripción, teoría y aplicaciones) and FFI2011-24183 (Portal de léxico hispánico: documentación y morfología derivativa); and by the *Departament de Filologia i Comunicació* of the *Universitat de Girona* through a grant for research (Ajut a la recerca 2013).

I am grateful to Dra. Montserrat Batllori for her help and comments. All errors are my own reponsability.

 $^{^2}$ For an excellent description of the current state of the issue of the classification of denominal verbs, vid. Lavale (2013: 307-344).

Spanish, with special attention to those verbs attested in the 13^{th} century and in the 15^{th} c.

I am going to argue that the diversity of meanings in denominal parasynthetic verbs can be attributed essentially to the semantic features of the noun. This noun, like any noun, has a minimum meaning that legitimates its possible combinations with other lexical units, as an independent word, and it accounts for the derived verb meaning as well.

In order to explain the genesis of polysemy, I will consider the proposal concerning the Qualia Structure of nominals presented in Pustejovsky's (1995) *Generative Lexicon Theory.*³ The author puts forward the idea that new would-be meanings are already conveyed as a possibility in the Lexicon definition of the word, cf. de Miguel (2009: 341).

In this paper I will not deal with the incidence of prefix a- in these verbs, because I believe that this question should be studied throught a comparison with other verbs formed with different prefixes.⁴

2 Lexical information within the Generative Lexicon

It is not my intention to submit a detailed description of the Generative Lexicon framework in this section (for this topic vid. Pustejovsky 1995, De Miguel 2009, and Batiukova 2009, among others). However, I would like to highlight some key aspects of this proposal which will be taken into account in my analysis. According to Pustejovsky (1995), semantic information concerning a lexical item can be defined at four levels of representation:

³ In Pustejovsky's words "What qualia structure tells us about a concept is the set of semantic constraints by which we understand a word when embedded within the language.", cf. Pustejovsky (1995: 86).

⁴ Vid. Schroten (1997) for the contrast between Spanish prefixed verbs with *a*- and with *en*-; vid. Acedo (2006) for Catalan change-of-state verbs with prefix *a*- and *en*- in contrast with prefixed verbs with *es*-; vid. Iacobini (2010) for Late Latin denominal and deadjectival verbs with preverbs *ad*-, *de*-, *dis*-, *ex*- and *in*-.

(1) Levels of representation (Pustejovsky 1995: 61)

- ARGUMENT STRUCTURE (AS): Specification of number and type of logical arguments, and how they are realized syntactically.
- EVENT STRUCTURE (ES): Definition of the event type of a lexical item and a phrase. Sorts include STATE, PROCESS, and TRANSITION, and events may have subeventual structure.
- QUALIA STRUCTURE (QS): Modes of explanation, composed of FORMAL, CONSTITUTIVE, TELIC and AGENTIVE roles.
- LEXICAL INHERITANCE STRUCTURE (LIS): identification of how a lexical structure is related to other structures in the type lattice, and its contribution to the global organization of a lexicon.

QUALIA STRUCTURE (QS) specifies essential aspects of a word's meaning in four roles: CONSTITUTIVE, FORMAL, TELIC and AGENTIVE, vid. (2). This information is highly relevant for the morphological analysis of words, since it allows us to explain different types of polysemy, as well as the possible senses of a derivative word, vid. Schroten (1997), Batiukova (2008), Salazar (2011), Adelstein (2012), and Berri & Adelstein (2012).

(2) QUALIA STRUCTURE = [CONSTITUTIVE = what x is made of FORMAL = what x is TELIC = function of x AGENTIVE = how x came into being]

3 From verbal stems to nominal stems in Latin: an example of reanalysis

In Classical Latin, prefixed verbs with *ad*- usually display a verbal stem (*admovēre* < *movēre*, *adoptare* < *optare*, *adluere* < *luere*). Acedo (2013) states that preverbation was responsible for the production of a large number of verbs in both Archaic and Classical Latin (200 BC - end of 300 AD). With a verbal stem, the preverb may operate either as a main predicate —*curro* 'to run' vs. *accurro* 'to run to a place'; *caedo* 'to fall' vs. *accido* 'to start to fall, drop'—, or as a modifier of the main predicate (*amo* 'to love' vs. *adamo* 'to love intensely'). In the first case, the verbal stem denotes the Manner in which the event occurred, vid. Acedo & Mateu (2009).⁵

⁵ Acedo (2013) and Acedo & Mateu (2009) are based on Talmy's 1985 typology in motion events. According to Talmy (1985), a motion event consists of several elements: Motion, Figure, Ground, Path, Manner and Cause.

The Oxford Latin Dictionary (OLD) offers only 3 examples of verbs that derive from a nominal stem: accuso < [ad- + cavsa + -o], adaero [ad- + aera + -o] and admoenio [ad- + moenia + -o]). In his study based on the Thesaurus linguae latinae corpus, Iacobini (2010) considers the following examples from Classical Latin denominal verbs: accommodo 'to fit, apply, adapt' < commodum 'benefit, profit, interest'; accumulo 'to heap up, pile up, accumulate' < cumulus 'a heap, pile, mound'; adglutino ~ agglutino 'to glue, paste on' < gluten 'glue'; aggrego 'to join together' < grex, gregis 'a herd of domestic animals'; and adumbro 'to shade' < umbra 'shadow'. Iacobini (2010) also points out the following Classical Latin verbs derived from adjectives: acclaro 'to reveal, make manifest' < clarus; adnubilo 'to become cloudy' < nubilus 'covered with cloud, cloudy'; adsevero ~ assevero 'to assert emphatically, affirm' < severus 'severe, serious'; and appropinquo 'to come near' < propinguus 'near'.

Nevertheless, as Latin evolved, new verbs with *ad*- were formed from a nominal stem (*accorporo < corpus*; *adhospito < hospes*) or an adjectival stem (*adlasso < lassus* 'tired', *allevio < levis* or *levius*). There are fewer examples of verbs with adjectival stems, vid. Iacobini 2010.

Iacobini (2010) analyses the evolution of nominal and adjectival verbs with preverbs *ad-*, *de-*, *dis-*, *ex-* and *in-* from Classical Latin to Late Latin. In his study, the author asserts that in Classical Latin there is a corresponding non-prefixed verb for 45,6 % of these prefixed verbs (for example, *animo* and *exanimo*, *curuo* and *incuruo*), while there are no corresponding non-prefixed verbs for the remaining 54,4% (for example, *decortico* but not **cortico*, *accommodo* but not **commodo*). In Late Latin, the number of pairs of prefixed verb forms and non-prefixed verb forms decreases: only 23% of verbs show this relationship while the remaining 77% do not.⁶ As stated by this author the existence of nominal and adjectival prefixed verbs without a corresponding to Iacobini, prefixed verbs which have a non-prefixed partner usually have the same meaning as the non-prefixed verb (*accumulo* vs. *curulo*, *incurvo* vs. *curuo*, *intitulo* vs. *titulo*). Hence, nominal and adjectival verbs.

(3) "Un [...] appauvrissement sémantique [du préverbe] a déterminé la synonymie effective entre verbes préfixés et verbes non-préfixés et il a permis la réinterprétation de verbes préfixés comme s'ils étaient directement formés par des noms ou des adjectifs.", Iacobini (2010).

⁶ One of the reviewers suggests that we should consider that this is common with oriented verbs such as *dismember* where the possible intermediate verb has no pragmatic use.

In my opinion, the change from verbal stems to nominal/adjectival stems can also be understood as an example of formal reanalysis in which the listeners relate prefixed verbs to nominal or adjectival stems and not to verbal stems. This is possible when there is a close formal relationship between a verb and a noun or adjective, as in (4) and (5). Once reanalysis has occurred, the prefix *ad*- can combine with nouns or adjectives to create new verbs independently of the existence of a previous verb.⁷

- (4) a. $adaquo \leftrightarrow aqua, -ae (aqvor)$ b. adiugo \leftrightarrow iugum, -i (ivgo) c. administro ~ amministro \leftrightarrow minister, -tri (ministro) d. admurmuro \leftrightarrow murmur, -ris (mvrmvro) e. adnumero ~ annumer \leftrightarrow numerus, i (nvmero) f. adnuntio \leftrightarrow nuntium, -ii (nvntio) g. adoro \leftrightarrow os, oris (oro) h. aduelo \leftrightarrow velum, -i (velo) i. aduerbero \leftrightarrow uerber, -ris (verbero) j. aduigilo \leftrightarrow vigil, -ilis (vigilo) k. $aduoco \leftrightarrow vox$, -cis (voco) 1. adumbro \leftrightarrow umbra, -ae (vmbro) m. affiguro \leftrightarrow figura, -ae (figvro) n. agglomero ~ adglomero \leftrightarrow glomus, -eris (glomero) o. agglutino ~ adglutino \leftrightarrow gluten, -inis y glutinum, -i (glvtino) (5) a. acclaro 'to reveal, make manifest' \leftrightarrow clarus (claro) b. *adamplio* 'to enlarge' \leftrightarrow *amplius* (*amplio*) c. *adapto* 'to adapt or modify' \leftrightarrow *aptus* (*apto*) d. addenso 'to thicken' \leftrightarrow densus (denso) e. admaturo 'to hasten' \leftrightarrow maturus (matvro)
 - f. aggrauo ~ adgrauo 'to weigh down' \leftrightarrow gravis (gravo)

The situation described above demonstrates how nouns with very different semantic features can be perceived as stems of prefixed verbs (people:

⁷ Acedo (2006), in line with Talmy (1985) and (2000), regards the reanalysis of prefixed verbs from Latin to Romance as follows: "The satellite-framed > verb-framed typological change plays a special role in the diachronic interpretation of the difference between Latin and Romance prefixed verbs. In fact, this change has been suggested to have favoured the reanalysis of typical satellite-framed predicates with conflation of a manner Co-event as new change-of-state predicates of the Romance type which conflate the Ground component into the verb."
minister, *vigil*; parts of the body: *os*; objects or instruments: *velum*, *glomus*, *iugum*, *uerber*; materials or substances: *gluten*, *aqua*; sounds: *murmur*, *vox*; quantity: *numerus*). This flexibility gives great freedom when it comes to the formation of new denominal parasynthetic verbs.

Iacobini (2010) provides us with the following examples from Late Latin which exemplify this diversity of meaning (people: hospes 'guest' > adhospito 'to entertain as guest'; parts of the body: geniculum 'knee' > aggeniculor 'to bow the knee before, to kneel before', pectus 'breast, chest' > appectoro 'to press to the breast'; objects or instruments: aes 'coin' > adaero 'to value, to put price', glomus 'a ball-shaped mass' > agglomero 'to mass together, join forces (with others)', 'to pile up in masses'; quantity: nihil 'nothing' > adnihilo 'to bring to nothing, to annihilate', pretium 'a price, money value, value in exchange'> appretio 'to value or estimate at a price, to appraise, rate', 'to purchase'). Other examples from Iacobini (2010) denote 'part of something': decima > addecimo 'to pay / to receive the tithe', titulus > adtitulo 'to name, entittle'. As the preverbation of the verbal stem (vid. Acedo 2013 and Acedo & Mateu 2009), the nominal stem of most of these verbs still denotes the Manner in which the event occurs: adhospito 'to entertain as a guest'; appectoro 'to press to the breast'; adaero 'to estimate by money'; agglomero 'to wind on as a ball'; addecimo 'to take by the tenth part'. In the case of deadjectival verbs, the stem denotes a state.

4 Verbal patterns: the semantic types of nominal stems

In many recent morphological studies of derived verbs in Spanish it is assumed that the semantic type to which the stem of the derived form belongs determines the latter's syntactic and semantic characteristics to a considerable degree.⁸ Most parasynthetic verbs prefixed with *a*- denote a change in quality or state or a change of place (cf. Malkiel 1941 for examples in Latin and Old Spanish and Gràcia *et al.* 2000 for Modern Spanish, among others).

It should also be borne in mind that verbs like those that appear in (4), which could be interpreted from a formal viewpoint as denominal, serve as models for neological denominal verbs. This is shown by new formations in Late

⁸ Martín García (2007: 280) states "[...] las propiedades sintácticas y semánticas del sustantivo que interviene en la formación de verbos en *–ear* determinan la estructura argumental y aspectual de los verbos derivados, así como su significado." Batiukova (2008), in her study of verbs ending with *–izar*, emphasises the direct relationship between the semantic information specified in the QS of the nouns and adjectives acting as the stem for these verbs and the meaning of the derived verb: "[...] el tipo semántico de la base predetermina la lectura [del verbo derivado] en gran medida."

Latin (cf. the examples provided by Iacobini 2010) and the earliest examples documented in Spanish.

In this section I am going to deal with the semantic types of nouns forming the stem of the first parasynthetic verbs in Spanish (nouns denoting instruments, properties, objects and places) with a view to describing their polysemy.

4.1 Nouns that denote an instrument

Parasynthetic verbs whose stem is an instrumental noun⁹ (such as *acuchillar*) express an event which denotes a change in quality or state in their internal argument (IA). In Classical Latin verbs such as *advelo* 'to use a veil to cover or hide something' are already documented. This verb is formed from the verb *velo* (with the same meaning) derived in turn from the noun *velum*, *-i* 'veil, covering'. As a result, in the series *advelo* – *velo* – *velum*, the prefixed verb *advelo* is closely related to the noun *velum*, both in form and meaning. Other verbs of the same type as *advelo* are *adiugo* 'to use a yoke to join', 'to join as a yoke does', *adverbero* 'to use a whip to flog', etc., which contain nouns designating instruments. Following this model, Spanish has, since its origins, created denominal verbs like those in (6), based on instrumental nouns:

a. *amolar* 'to use a grindstone to sharpen a knife or a cutting object' < muela 'grindstone'

b. *arrendar* 'to use reins to tie an animal' < *rienda* 'rein'

c. *acuchillar* 'to use a knife to stab or to kill somebody' < *cuchillo* 'knife'

d. *ahorcar* 'to use the gallows to hang somebody' < *horca* 'gallows'

e. *aserrar* 'to use a saw to cut something' < *sierra* 'saw'

f. *atrancar* 'to use a bar to shut the door' < *tranca* 'bar'

The model was active throughout the Middle Ages, as can be seen in the following verbs created in the fifteenth century:

⁹ For a definition of the concept of instrument, vid. Bolaños (2011).

a. *abotonar* 'to use a button (or buttons) to do up a garment' < *botón* 'button'
b. *abrochar* 'to use a clasp, fastener to do up a garment' < *broche* 'clasp, fastener'
c. *acepillar* 'to use a brush or plane to smooth wood or metals' < *cepillo* 'brush or plane'
d. *atenazar* 'to use tongs to torture someone plucking out pieces of flesh' < *tenazas* 'tongs'
e. *atraillar* 'to use a rope to tie up dogs' < *trailla* 'rope to tie up dogs in hunting'

According to the Generative Lexicon theory, the objects and instruments nouns form part of the unified or functional types. As such, their Telic role (the purpose or function of the object) is specified. The information related to a noun like *cuchillo* is thus specified as 'an instrument that can be used to cut', vid. Pustejovsky (1995: 10) and Batiukova (2009: 238-9), among others.

(8)	cuchillo 'knife'	
	Argument Structure =	[ARG1 = x: tool/instrument]
	-	D-ARG 2 = y: object]
	Qualia Structure =	[FORMAL = x]
		TELIC = cut (e, x, y)]

In contrast with this information, we can see that in verb formation the value of stem nouns can undergo changes in some cases. The purpose of the *cuchillo* in *acuchillar* is not to cut but 'to wound or kill'. Similarly, while *muela* means a 'stone disc used for grinding', its use in *amolar* refers to a 'stone disc used to sharpen cutting implements'. Although tenazas refers to a metal tool with two arms used to seize, pull or cut something, the use of tenazas in atenazar means 'an instrument used to torture people'. This fact is related to restrictions on the choice of verb with respect to the internal argument (IA). We thus find that the verb *acuchillar* selects an IA responding to category 'living being' (one stabs living beings but cuts material objects). This leads to a series of changes in the semantic properties of the stem noun *cuchillo*, vid. (9): the knife becomes a 'weapon that is used to wound or kill living beings'. These more specific values will be the ones projected in the semantic properties of the derived verb.¹⁰

¹⁰ In some cases the change affects only the Telic role of the verb. We find this in *arrendar*. While *rienda* refers to a 'strap used to control horses' the sense of *rienda* in *arrendar* is a 'strap *to tie up* horses'.

(9)	cuchillo 'knife' > acuch Argument Structure =	illar 'to knife, stab' [ARG1 = x: weapon D-ARG2 = y: living creature
	Qualia Structure =	[FORMAL = x TELIC = hurt, death (e, x, y)]

As shown in (10), in verbalization the stem noun becomes a hypernym. We thus see that one can stab (*acuchillar*) with a knife (*cuchillo*) but one could also do it with a sword; we can fasten (*abrochar*) a garment with a brooch (*broche*) but also with buttons, buckles, etc.

(10) a. "este rey ... acuchilla con espada estrecha ..." [CDH: 1442, Juan de Mena, *Homero romanzado*]
'this king... stabs (someone) with a narrow sword'

b. "... una capa de escarlata que con *hevillas* y *ojales* de oro se *abrochava*." [CDH: 1482-1492, Garci Rodríguez de Moltalvo, *Amadís de Gaula*]

'... a scarlet cape done up with gold buckles and buttonholes'.

4.1.1 Nouns that are interpreted as instruments

In verbalization, base nouns which do not themselves denote instruments can be used if they did, because of the fact that their function or purpose (*i.e.*, their Telic role) is prominent enough to be transferred to the verb. This is the case of nouns designating substances or parts of the body.¹¹

In connection with the former, we find that in Latin a verb like *adaquo* already has various meanings, cf. Valbuena s.v.: in Suetonius it means 'to water cattle'; in Caesar 'to obtain supplies of water'; another meaning is 'to wet, soak, moisten, spray'. Note that the three meanings of the derived verb are related to a specific aim or purpose for which water is used, with the information contained in the Telic role. Following the model of *adaquo* < *aqua*, Spanish forms *amelezinar* 'to give medication in order to cure' < *melezina*.

¹¹ Schlesinger (1995) proposes a classification of instruments in ten types, those indicating CAUSE being a particularly important group: tools (*e.g.* Jack cut the cake with a *knife*), means of transport (*e.g.* She came by *plane*), body parts (*e.g.* He peeled the apple with his left *hand*), abstract instruments (*e.g.* You ought to persuade him with nice *words*) and material (*e.g.* We washed the dishes with *soap*), vid. Bolaños (2011: 28).

Regarding nouns that refer to a part of the body, following the model of *aggeniculor* 'to use one's knees to kneel', other verbs are created, including *abraçar* 'to use one's arms to greet, show affection, etc.' < *brazo* 'arm', *acodar* 'to lean on one's elbows' < *codo* 'elbow', *apearse* 'to use one's foot to alight from a horse' < *pie* 'foot', *arrodillar* 'to use one's knees to kneel' < *rodilla* 'knee'.

4.2 Nouns that denote a specific property

Nouns can be classified into different semantic types. Bosque (1989: 36) states that nouns may define "objetos físicos, como *casa*, pero también procesos como *envejecimiento*; estados como *inocencia*; o acciones como *destrucción*." In the case of denominal verbs, several authors have suggested that the nominal stem may refer to certain properties inherent to an object or entity (and not to the object or entity itself). From this point of view, the nominal stem may be perceived as being closer to an adjective, vid. Serrano Dolader (1995: 117-118), Rifón (1997: 126), Gràcia *et al.* (2000), and Lavale (2011). According to Acedo (2006: 51) and Lavale (2011: 120), the nominal stems of denominal verbs must be understood as states.¹²

In Latin very few denominal verbs with prefix *ad*- in which the nominal stem refers to a quality have been attested. In the derived verb this property can affect either the Internal Argument (e.g. *admoenio*) or External Argument (e.g. *administro*, *advigilo*), vid. Lüdtke (2011: 190-199).

4.2.1 Qualities that are transferred to the Internal Argument of the verb

The example of *admoenio*, *-ire* 'to surround with walls, to fence in, to besiege (a city)' < *moenia* 'walls' shows a transitive verb of action in which a property of the stem noun (the wall that surrounds a place) is transferred to the Internal Argument (IA) of the verb (the city). Thus, the verb denotes a change of state. The walls are seen as walls surrounding a place. Therefore, in the verbalization, nominal stems project semantic information on a prototypical property of the noun to an Internal Argument of the verb. As a result, the city may have been perceived as a walled city (or even, metaphorically, a besieged city).

¹² From a syntactic perspective of the morphology, Acedo (2006: 67) questions the concepts of both morphological derivation and any inheritance of the Argument Structure: "En el cas dels verbs denominals en concret, no existeix, al nostre parer, una relació de derivació lèxica entre *ploma* i *esplomar*, per exemple, sinó que ambdós mots són el resultat de combinar la mateixa arrel, "plom, amb configuracions diferents de morfemes a la sintaxi."

Following the *admoenio* pattern, the verb *affiliare* 'to take someone as a son, to adopt' < filius, *-i* 'son' (vid. Niermeyer) is attested in the 10th century; in Spanish the first attestation of this verb is found in Aragonese documents in the 11th century (DCECH). Old Spanish creates other verbs that share this pattern as the examples below illustrate:

(11) afilar 'to sharpen, to make sharp o sharper'

First attestation: 13th C. (DCECH) Example: *afilar los cuchillos* 'to sharpen the knives' Derived from *filo* 'sharp edge of a cutting instrument' Prototypical property of *filo*: sharp **filo**

QUALIA STRUCTURE = [CONSTITUTIVE = x FORMAL = sharp]

The prototypical property of *filo* is transferred to the IA of the verb *afilar* (cutting instrument with *sharp* edge). The QUALIA STRUCTURE of the IA encodes this semantic information in the FORMAL role).

(12) *apolillar* 'to be eaten by moths (clothing, wool, etc.)'¹³

First attestation: 15th C. (DCECH)

Example: *apolillar(se) una ropa, lana* Derived from *polilla* 'moth (any of various moths of the family Tineidae, whose larvae feed on wool, hair, fur, and feathers)'

Prototypical property of the *polilla*: to feed where it nests. **polilla**

QUALIA STRUCTURE = [FORMAL = x]

TELIC = to feed on (e, x, y)

The prototypical property of the moth is transferred to the IA of the verb *apolillar* (clothing which is *moth-eaten*). The Qualia Structure of the IA encodes this semantic information in the FORMAL role).

(13) *abrasar* 'to burn', 'to be destroyed by fire'

First attestation: 15th C. (DCECH)
Example: *el incendio abrasó la cabaña* 'the fire burned the hut'
Derived from *brasa* 'embers; a glowing or smouldering piece of coal or wood, as in a dying fire'
Prototypical property of *brasa(s)*: the quality of hot **brasa**

¹³ In Nebrijas's *Vocabulario*, the Latin translation of *apolillar* is "tinea pertundo. is." 'the moth, to make holes in', translation which highlights the typical activity of the moth.

QUALIA STRUCTURE = [CONSTITUTIVE = wood, coal FORMAL = hot, incandescent TELIC = burn (e, x, y) AGENTIVE = be on fire (e, x)]

The properties of embers are transferred to the IA of the verb *abrasar* (something burned, reduced to embers). The Qualia Structure of the IA encodes this semantic information in the FORMAL role.

(14) aislar 'to isolate, to place apart, to cause to be alone'¹⁴

First attestation: 15th C.

Example: *la crecida del río aisló los pueblecitos de la capital* 'the swelling river isolated villages from the capital'

Derived from *isla* 'island, a mass of land that is surrounded by water'

Prototypical property of *isla*: isolation; the fact that an island is surrounded by water on all sides motivates a metaphorical interpretation in wihich the island is perceived as an isolated land.

isla

QUALIA STRUCTURE = [CONSTITUTIVE = mass of land FORMAL = surrounded by water]

The prototypical property of the island is transferred to the IA of the verb *aislar* (something isolated). The Qualia Structure of the IA encodes this semantic information in the FORMAL role.

Thus we can see that properties transferred to the IA of the verb may refer to different aspects of the semantic information of the nominal stem (qualities themselves, but also states caused by a typical function).

Among denominal verbs in which the stem noun transfers a property to the IA, we can distinguish, from the Early Spanish, a sub-group of verbs which denote a change in a person's state of mind. The presence of such verbs was already conspicuous in the thirteenth century (*atormentar*, *airar*, *abiltar*, *avergonzar*, *agradar*), while new forms were generated in later centuries, such as *apasionar*, in the fifteenth century, vid. (15). The nouns on which these verbs are based designate emotional states (torment, anger, abjection, shame, pleasure, passion) and these properties are transferred to the IA of the verb.

¹⁴ In Nebrija's dictionaries the Latin equivalent of *aislar* is *intercludo*, *-ere* (< *inter* + *claudo*) 'close or block the path or way in', cf. Valbuena s.v.

(15) a. "Non ha cosa que *abilte* a onbre como çaherimiento y rretraerle muerte que fizo." [*CORDE*: c 1250, *Libro de los buenos proverbios que dijeron los filósofos y sabios antiguos*]

'The thing that most *degrades* a man is to say to him or to do something with which he feels humiliated and to say to him that he killed someone'

b. "d'esta batalla que avemos arrancado; / al rey Alfonso, que me á *airado*, / quiérol' enbiar en don treinta cavallos." [*CORDE*: c 1140, *Poema de Mio Cid*]

'From this battle that we have won, I would like to send to King Alfonso, who broke off his friendly relationship with me, thirty horses as a gift'

c. "... vna vez enel año cortaua los cabellos que le cresçian mucho & le *agradauan* [...]." [*CORDE*: 1293, *Castigos*. BNM ms. 6559] 'once a year, he used to cut his hair: it grew a lot and he *liked* it'

d. "E bien que el abad non fuese agrauiado por otra enfermedad, pero la gota, que *atormentaua* sus manos e pies, non çesaua de lo fatigar." [*CORDE*: c 1255, *Crónica de Sahagún*]

'And although the abbot was not affected by any other illness, the gout that *tormented* his hands and his feet gave him no rest'

e. "[...] el padre desta moça [...] a ella con su fijo *auergonço*" [*CORDE*: a 1452, Alfonso Gómez de Zamora, *Morales de Ovidio*. BNM ms. 10144]

'the father of this young woman embarrased her with her son'

Although these verbs have existed in Spanish since its beginnings, I would like to point out that I see no exact parallel in Latin, where changes of state, both physical and mental, are usually expressed by inchoative verbs in *-escere: erubesco* 'to be ashamed' < ex + rubesco; insolesco 'to take pride' < insolens + -sco; ditesco 'to become rich' < dis, -itis + -esco; flaccesco 'to become thin' < flacceo + -sco; pinguesco 'to grow fat' < pinguis, -e + -esco; senesco, insenesco and consenesco 'to grow old'; albesco 'to become white' <math>< albeo + -sco. According to Malkiel (1941: 432) between the fourth and fifth centuries verbs in *-escere* developed factitive-causative values, meaning that from this time onward they could express two types of process: inchoative, in which case the subject acquired a property, or causative, in which case the subject brought about the acquisition or modification of a property in the object.

Emotional verbs such as *airar* coexist in the language with other parasynthetic or deadjectival verbs which present a formal structure of the type [*en*- + adjective + -*ecer*] and with which they share the same semantic characteristics: *embravecer*,¹⁵ *envilecer*,¹⁶ *enloquecer*, *entristecer*.

(16) a. "El bravo *enbravece* a los hombres contra si." [CORDE: c 1430, *Floresta de philósophos*]

' The angry man infuriates people against him'

b. "[...] el Señor al omne recto crio etc., e el pecado lo *envilesçe*." [CDH: c 1422-1433, Mose Arragel de Guadalfajara, *Traducción y glosas de la Biblia de Alba*]

'God created the upright human being, and sin degrades him'

c. "[...] la fornicaçion *enloqueçe* al sabidor, la enbriagueza pone en captiuo los sentidos del onbre [...]". [CDH: *Traducción del Soberano bien de San Isidoro*]

'fornication *drives* the wise man *mad*, drunkenness captivates a man's senses'

The [*en*- + adjective + -*ecer*] pattern in verbs shows great vitality in Old Spanish —vid. Batllori & Pujol (2012), and Sánchez González de Herrero (1992: 1316-17)— although according to Malkiel (1941) the most productive deadjectival pattern at this stage and the one that allows the formation of most neologisms is [*a*- + adjective + -*ar*], as against [*en*- + adjective + -*ecer*], [*a*- + adjective + -*ecer*] (*aflaquecer*, *ablandecer*)¹⁷ and [*en*- + adjective + -*ar*] (*engrosar*, *ensordar*, *encortar*).¹⁸

4.2.2 Properties that are transferred to the External Argument of the verb In the stem of the Classical Latin verbs administro, -are and advigilo, -are we find an agent noun (*minister*, -tri 'person who serves, servant' and vigil, -ilis 'person who watches or guards', respectively). The prototypical information in these cases indicates the way in which these individuals act. When the

¹⁵ Note the synonymy with *airar*, mentioned above.

¹⁶ Note the synonymy with *abiltar*, mentioned above.

¹⁷ In her study of parasynthetic verbs in Spanish medical text of the fourteenth and fifteenth centuries, Sánchez González de Herrero (1992: 1317) points out that a large proportion of verbs prefixed in *en*- have a less used variant prefixed in *a*-: this is the case of *enmollecer / amollecer*, *enflaquecer / aflaquecer*, *ennegrecer / anegrecer*, *emblandecer / ablandecer*, *enclarecer / aclarecer* or *endormecer / adormecer*.

¹⁸ Vid. Batllori (2012, in press) for a diachronic analysis of verbs prefixed in *a*- and ending with *-ecer*.

prefixed verbs are formed, this information is transferred to the external argument (EA). In this way the subject of the verb *administro* 'assist, help', 'take care of', 'oversee' acquires the property of the servant and the subject of *advigilo* 'watch', the property of the watchman.

Following the model of these verbs *amaestrar* is documented in Old Spanish.¹⁹ In examples like this there is no change of state in the IA of the verb, as the property denoted by the stem is not projected onto it (a person does not become a master as a result of being taught), vid. Batiukova (2008).²⁰

(17) amaestrar ' to train, to teach'

Examples: *amaestrar un muchacho desde que era un niño* 'teach a boy since he was a child'; *el hombre amaestra la bestia* 'the man trains the beast'

Derived from *maestro* 'teacher, person whose occupation is teaching others'

Prototypical property of maestro: teach

maestro

QUALIA STRUCTURE = [FORMAL = x: human being TELIC = teach (e, x, y)]

The prototypical property of teacher (to teach) is transferred to the EA of the verb *amaestrar* 'to act as a teacher'. The Qualia Structure of the EA encodes this information.²¹

4.3 Nouns that denote objects

In contrast with the discussion in the previous section, the nominal stem in certain parasynthetic verbs such as *anidar* 'to nest, make one's nest' may denote an object (and not certain properties of the object itself). In these structures the parasynthetic verb expresses the genesis of the object. This information is encoded in the Qualia Structure of the nominal stem and, precisely, in the AGENTIVE role, vid. (18).²² The verbs following this pattern

¹⁹ CE gives an example of *amaestrar* from the thirteenth century in Berceo. However, according to the DCECH (s.v. *maestro*) the variant recorded in Berceo does not have the prefix: *maestrar*. CORDE documents exemples of *amaestrar* as early as the fourteenth century.

²⁰ Vid. Batiukova (2008) for an explanation of similative verbs in -izar, such as *tiranizar*.

²¹ Batiukova (2008) states that the way people act (eg. to act as a teacher) is also associated with verbal event type "de la misma forma que la manera de moverse está relacionada con el desplazamiento en verbos como *gatear*, *volar*, *trepar*, etc.".

 $^{^{22}}$ Vid. Batiukova (2008) for some verbs ending in *-izar* that are intrepreted as performative verbs.

are generally intransitive verbs or behave like intransitive ones, vid. (19). Observe that Nebrija defines these verbs, in his *Vocabulario*, with *hacer* 'to do', vid. (20).

(18)	nido 'nest'	
	QUALIA STRUCTURE =	[FORMAL = x: object
		AGENTIVE = generate]

(19) a. "et vosotros todos non çesedes de *aventar* con vuestras alas et de soplar el fuego" [CDH: 1251, *Calila e Dimna*], *aventar* 'hacer viento' < *viento* 'and don't you all stop beating your wings and blowing upon the fire', *aventar* 'to make, produce a draught of air' < *viento* 'wind, air'

b."las aves empollan, los ganados *ahijan*" [CDH: 1513, Gabriel Alonso de Herrera, *Obra de agricultura*], *ahijar* 'crear hijos' < *hijo* 'birds lay eggs, cows have calves', *ahijar* 'to give birth to children' < *hijo* 'child'

(20) a. "Ahoiar hazer hoio. [lat.] scrobem fodio" *'Ahoiar* to make a hole'

> b. "Ahumar hazer humo. [lat.] fumigo. as" *Ahumar* to give out smoke'

c. "Anidar hazer nido. [lat.] nidifico. as. nidiculor. aris." 'Anidar to make one's nest'

d. "Arraigar hazer raizes. [lat.] radico. as" 'Arraigar to take roots'

e. "Aventar hazer viento. [lat.] uentilo.as" *'Aventar* to make, produce a draught of air'

4.4 Nouns that denote a place

The intransitive verb of motion ARRIPARE < AD + RIPA 'riverbank', 'coast', 'shore' is attested in Late Latin. This verb will bring about Old Spanish *arribar*, which preserves its etymological meaning of 'to reach the shore'

throughout the Middle Ages. In this case, the noun *riba* indicates the goal of the movement.²³

(21) "(...) en las barcas son metidos, / van buscar a Valencia, a mio Cid don Rodrigo; / arribado an las naves, fuera eran exidos." [CDH: c 1140, *Poema de Mio Cid*]
'they have got on the boats, / they are going to Valencia to look for mio Cid don Rodrigo; / the ships *have reached the shore*, they have got out of them'

However, the verb *arribar* may also combine with a locative prepositional phrase, as the following examples show. In (22a) the prepositional phrase conveys once again the locative value of the nominal stem. Nevertheless in (22b) the nominal stem has no locative meaning, because it is expressed in the prepositional phrase. So, it could be argued that the nominal stem of the verb *arribar* lexicalises the Manner of the motion event, the Manner in which the goal is reached.

(22) a. "a las aguas de Duero ellos arribados son" [CDH: c 1140, Poema de Mio Cid]
 'they have reached the Duero shore'

b. "[...] el barón [...] *arribó* a la corte del rei don Fernando" [CDH: c 1236, Gonzalo de Berceo, *Vida de Santo Domingo de Silos*] 'baron *came* to King Fernando's court'

The verb *asomar* < *somo* 'the highest point, summit' < lat. SUMMUS 'highest' exemplifies parasynthetic verbs with a locative stem that were created in Old Spanish, vid. DCECH s.v. *somo*. The historical dictionary (DHist.) defines the first meaning of this verb as 'to reach the top or to climb' and provides the following example that illustrates the figurative meaning of this verb:

(23) "El ome cobdicioso que non sabe guardar / ciégalo la cobdicia, faze lo asomar, / faze lo de la cima, caher en mal lugar" [*Dhist: Libro de Alexandre*].
'The greedy man who cannot wait / greed blinds him, greed puts him

at the top, / greed topples him forward from the top into the wrong place'

²³ According to Talmy (1985), the locative stem of this verb lexicalises the Ground of motion, vid. Acedo & Mateu (2009: 489).

However I don't find examples of this meaning in CORDE where the semantic values given by the DCECH are commonly attested: 'to appear at the top of a place', 'to appear at a great distance' and 'to start showing something'. As in the previous example of *arribar*, the nominal stem of this verb doesn't include the locative meaning, which is expressed in the prepositional phrase as well. So, it could be argued that the nominal stem of the verb *asomar* lexicalises the Manner of the motion event 'as if he were on the top, that is, far away'.

(24) "Quand assomó Achiles a unos campos planos, / conoçiéronlo luego en los gestos loçanos" [CDH: c 1240-1250, Libro de Alexandre]
'When Achiles appeared in the flat fields, they knew him immediately by his good-looking appearence'

The examples in (25) show two meanings of the verb *aventar* (< *viento* 'wind'): on the one hand, 'to take something into the air', specially 'to winnow, to separate grain from chaff by means of a wind or current of air'; and, on the other, 'to shoot into the air, to throw', vid. DCECH s.v. *viento*. In both cases the noun *viento* has a locative meaning. The same is true in the example of (26) that means 'to leave, expose to the wind'.

(25) a.Trillava don Agosto las miesses por las eras, aventava las parvas"
 [CDH: 1240-1250, Libro de Alexandre]
 'Mr. Agosto threshed the grain throughout the threshing floor, he winnowed the grain'

b. "Aventó un venablo que le avié fincado" [CDH: 1240-1250, Libro de Alexandre] 'he threw a dart [...]'

(26) a. "Aventar el pan al viento. uentilo.as." [Nebrija, Vocabulario] 'to leave the bread to the air, to air the bread'

> b. "Euentilo. as. por *aventar* el pan." [Nebrija, *Lexicon*] 'to leave the bread to the air, to air the bread'

It is worth noticing the polysemy of the verb *aventar* in its intransitive use with performative meaning, as shown in (19a) 'to make, produce a draught of air', that contrasts clearly with the transitive examples with locative meaning in (25) and (26).

5 Conclusion

As shown in section 4, Old Spanish denominal parasynthetic verbs with prefix *a*- provide evidence in favour of showing that the semantic type of the noun determines the meaning of the derived verb to a considerable degree. As a result, among change of state verbs, nominal stems that denote an instrument generate instrument verbs (*cuchillo* > *acuchillar*), and nominal stem that express a particular property generate causative-resultative verbs (*filo* > *afilar*, *polilla* > *apolillar*, *brasa* > *abrasar*, *isla* > *aislar*, *tormento* > *atormentar*). In regard with change of location verbs, the nominal stem displays a locative value —*viento* > *aventar* in (25) and (26)—. Hence, it can be stated that the event of these verbs is a Transition, vid. (1).

Among the verbs that do not convey change of state, there are verbs like *amaestrar* (< *maestro*), the nominal stem of which refers to an individual who has specific properties (who acts in a particular way). Hence, in these cases, the derived verb has a similative meaning 'to act as a x'. In addition, other verbs like *anidar* (< *nido*), with a nominal stem that refers to an object generated in the process of verbalization, bear a performative reading 'to produce x'. Thus, it is clear that the event of these verbs is a Process, vid. (1).

Consequently, it could be argued that, depending on the different semantic types of nominal stem, denominal parasynthetic verbs may lexicalise either some aspects related to the Manner of the event, or some information associated with the final result of the event. Instrument and similative verbs are examples of the first case while causative-resultative and performative verbs are examples of the second one. Therefore, the property denoted by the stem in causative-resultative verbs expresses the final state of the change, and the nominal stem of the performative verbs refers to the final object produced in the process of verbalization.

In the Generative Lexicon framework, polysemy of denominal parasynthetic verbs is explained by the generative mechanism called *selective binding*. According to it, the meaning of these verbs depends on the selection of the Qualia information by the IA/EA of the derived verb. Ultimately this information depends on the semantic type of the nominal stem, vid. (27). So, causative-resultative verbs (e.g. *afilar*, *apolillar*, *atormentar*) saturate the Formal role of their IA; locative verbs —like *aventar* in (25) and (26)—saturate the Constitutive role of their IA; similative verbs (*amaestrar*), the Formal role of their EA; and performative verbs (*anidar*), the Agentive role of the theme incorporated in the verb.²⁴ Instrument verbs behave differently,

 $^{^{24}}$ Vid. Batiukova (2008) for a synchronic analysis of the verbs ending with *-izar* within the Generative Lexicon framework.

though, because their nominal stem refers to a default argument in the Argument Structure of the verb.

Meaning of the derived verb	Exemple	Nominal Stem	Qualia
Causative- resultative	<i>afilar</i> los cuchillos `to sharpen the knives'	<i>filo</i> 'sharp edge'	The Formal role of the IA is saturated with semantic information from the nominal stem
Locative	<i>aventar</i> las parvas 'to winnow the grain'	viento 'wind'	The Constitutive role of the IA is saturated with semantic information from the nominal stem
Similative	<i>amaestrar</i> a alguien 'to train, to teach someone'	<i>maestro</i> 'teacher'	The Formal role of the EA is saturated with semantic information from the nominal stem
Performative	<i>anidar</i> 'to nest, make one's nest'	<i>nido</i> 'nest'	The Agentive role of the theme is incorporated in the derived verb
Instrumental	<i>acuchillar</i> a alguien 'to knife, stab someone'	<i>cuchillo</i> 'knife'	In this case cuchillo is not a property of an argument, but the syntactic argument itself

(27) Meaning of the derived verbs

Finally, in this paper I have provided evidence in favour of the proposal that the arguments with which the verb combines in the sentence may influence possible interpretations, vid. *ahijar* in section 4.2.1. in contrast with (19b) and *aventar* in (25) and (26) in constrast with (19a).

6 Corpus

abezar, abiltar, abollar, abollonar, abotonar, abraçar, acompañar, abrasar and abrasarse, abrochar, acepillar, acodar, acostarse,²⁵ acuchillar, adeudar, afilar, agradar, ahijar, ahocinarse, ahoiar, ahorcar, ahumar, airarse, aislar, alastrar, alastrarse, alindar, aliñar, alumbrar, amaestrar, amañar, amassar, amelezinar, amenazar, amenguar, amolar, amontarse, amontonar, anidar, aojar, apassionar and apassionarse, apearse, apolillar, apremiar, apresurarse, arrastrar, arrebatar, arrendar, arribar, arrimar and arrimarse, arrodillar, arropar, asserrar, asestar, asomar, asombrar, atenazar, atollar, atormentar, atraillar, atrancar, aventar.

References

Acedo Matellán V. 2006a. "Una aproximació sintàctica als verbs prefixats en català". *Estudios Catalanes* 4: 41-78.

- 2006b. "Prefixes in Latin and Romance and the satellite-/verb-framed distinction" in Actes del VII Congrés de Lingüística General, CD-ROM. Barcelona. Universitat de Barcelona.
- 2013. "Preverbs llatins: aspectes morfosintàctics i semàntics". Talk on October 25, 2013, Universitat Autònoma de Barcelona.
- Acedo Matellán V., J. Mateu 2009. "L'expressió dels esdeveniments de canvi: del llatí al català" in J. Rafel (ed), *Diachronic Linguistics*. Girona. Documenta Universitaria, 473-496.
- Adelstein A. 2012. "Delimitación espacial y formación de nombres relacionales" in E. Bernal, C. Sinner, M. Emsel (eds), *Tiempo y espacio* en la formación de palabras del español. München. Peniope.
- Batiukova O. 2008. "Morfología: del léxico a la sintaxis oracional" in Actas del VIII Congreso de Lingüística General, CD-ROM. Madrid. Universidad Autónoma de Madrid.

 $^{^{25}}$ Vid. Fernández Jaén (2007) for a study on the semantic evolution of the verb acostarse.

- 2009. "Aplicaciones lexicográficas de la teoría del Lexicón Generativo" in
 E. de Miguel *et al.* (eds), *Fronteras de un diccionario. Las palabras en movimiento*. San Millán de la Cogolla. Cilengua, 233-270.
- Batllori Dillet M. (2012, in press): "Evolución e historia de los verbos con prefijo a- y sufijo –esçer", in Actas del IX CIHLE (Cádiz, 2012). Cádiz. Universidad de Cádiz.
- Batllori Dillet M., I. Pujol Payet 2012. "El prefijo a- en la formación de derivados verbales", in Actas del VIII CIHLE (Santiago de Compostela, 2009). Madrid. Arco Libros, 1595-1607.
- Beltrán J.A. 1999. *Introducción a la morfología latina*. Zaragoza. Universidad de Zaragoza.
- Berri M., A. Adelstein 2012. "Polisemia regular en nombres con sentidos locativos" in E. Bernal, C. Sinner, M. Emsel (eds), *Tiempo y espacio en la formación de palabras del español*. München. Peniope.
- Bolaños Navalón J. 2011. "Instrumental verb formation: A conceptual approach" in J.L. Cifuentes Honrubia, S. Rodríguez Rosique (eds), Spanish Word Formation and Lexical Creation. Amsterdam – The Netherlands. John Benjamins.
- Bosque I. 1989. Las Categorías Gramaticales. Madrid. Editorial Síntesis.
- Gràcia Ll. et al. 2000. Configuración morfológica y estructura argumental: léxico y diccionario. Resultados del proyecto de investigación DGICYT PB93-0546-C04. Gipuzkoa. Servicio Editorial de la Universidad del País Vasco.
- Fernández Jaén J. 2007. "Prototypes, meanings and Motion: Cognitive Evolution of Spanish Acostarse". International Journal of English Studies 7 (1): 1-15.
- Iacobini C. 2010. "Les verbes parasynthétiques: de l'expression de l'espace à l'expression de l'action". *De lingua Latina* 3.
- Lavale Ortiz R.M. 2013. *Verbos denominales causativos en español actual.* PhD Dissertation. Universidad de Alicante, Alicante.

- Lüdkte J. 2011. La formación de palabras en las lenguas románicas: su semántica en diacronía y sincronía. México. El Colegio de México.
- Malkiel Y. 1941. "Atristar-entristecer: adjectival verbs in Spanish, Portuguese and Catalan". Studies in Philology 38: 429-461.
- Martín García J. 2007. "Verbos denominales en *-ear*: caracterización léxicosintáctica". *Revista Española de Lingüística (RSEL)* 37: 279-310.
- Miguel E. de 2009. "La Teoría del Lexicón Generativo" *in* E. de Miguel (ed), *Panorama de la lexicología*. Barcelona. Ariel, 337-368.
- Pustejovsky J. 1995. The Generative Lexicon. Cambridge. The MIT Press.
- Rainer F. 1993. Spanische Wortbildungslehre. Tübingen. Niermeyer.
- Rifón Sánchez A. 1997. Pautas semánticas para la formación de verbos en español mediante sufijación. Santiago de Compostela. Universidad de Santiago de Compostela.
- Real Academia Española 2009. "La derivación verbal. La parasíntesis" in Nueva gramática de la lengua española. Madrid. Espasa-Calpe, vol. I : 577-626.
- Reinheimer-Ripeanu S. 1974. *Les dérivés parasynthétiques dans les langues romanes*. The Hague-Paris. Mouton.
- Salazar Burgos H.R. 2011. Descripción y representación de los adjetivos deverbales de participio en el discurso especializado. PhD Dissertation. Universidad Pompeu Fabra, Barcelona.
- Sánchez González de Herrero N. 1992. "Derivados verbales contenidos en textos médicos medievales castellanos" *in Actas del II CIHLE*. Madrid. Pabellón de España, vol. I: 1315-1321.
- Schlesinger, I.M. 1995. Cognitive space and linguistic case. Semantic and syntactic categories in English. Cambridge. Cambridge University Press.
- Schroten J. 1997. "On Denominal Parasynthetic Verbs in Spanish" *in* J. A. Coerts and H. de Hoop (eds), *Linguistics in the Netherlands*: 195-206.
- Serrano Dolader D. 1995. Las formaciones parasintéticas en español. Madrid. Arco/Libros.

- Talmy L. 1985. "Lexicalization patterns: Semantic Structure in Lexical Forms", in T. Shopen (ed), Language typology and Syntactic Description III: Grammatical Categories and the Lexicon. Cambridge. Cambridge University Press, 36-149.
- 2000. Toward a cognitive semantics. Cambridge, Mass. MIT Press.

Sources

- *CDH* = Real Academia Española: *Corpus del Nuevo diccionario histórico del español*, in http://www.rae.es
- CE = Davies M. Corpus del español, in < http://www.corpusdelespanol.org>
- *CORDE* = Real Academia Española: *Corpus diacrónico del español*, in http://www.rae.es
- DCECH = Corominas J., J.A. Pascual 1980-1991. Diccionario Crítico Etimológico Castellano e Hispánico. Madrid: Gredos.
- Dhist. = Real Academia Española 1933-1936. Diccionario histórico de la lengua española, in http://www.rae.es
- Nebrija A. de 1495. *Dictionarium hispanum latinum* o *Vocabulario españollatino*. Madrid. RAE [facsimile of first edition, 1989].
- Niermeyer J.F. 1976. *Mediae Latinitatis Lexicon Minus*. *Abbreviationes et Index Fontium*. Leiden New York Köln. Brill.
- *OLD* = Glare P. G. W. (ed.) 1982. *Oxford Latin Dictionary*. Oxford. Clarendon Press.
- Valbuena M. de 1793/1860. *Diccionario universal latino-español*. París. Librería de Rosa y Bouret [1846, sixth edition].

ACCOUNTING FOR AFFIX POLYSEMY WITH SEMANTIC MAPS – A DIACHRONIC STUDY OF -AGE SUFFIXATION IN ENGLISH

Marion Schulte Bielefeld University

Abstract

This paper introduces a new way of accounting for the semantics of derivational affixes, namely a significantly adapted semantic map approach. The semantic map method has several advantages, such as great flexibility and openness, but it has until now mainly been used for cross-linguistic investigations of inflectional categories (*e.g.* Haspelmath 2003). The adapted method presented in this paper is optimised to account for the semantics of derivational affixes in a single language. The substantial adaptations are described in detail and compared to some previous semantic map approaches. The adapted method is then used to account for the semantics of the English derivational suffix *-age*. This investigation is diachronic and compares the semantics of *-age* in Middle English to *-age* in Present Day English.

1 Introduction

Affix semantics is, of course, not a new topic. It has been approached from many different theoretical directions and addressed by numerous scholars (*e.g.* Haspelmath 2003, Lehrer 2003, Lieber 2004, Plag 1999). In general, many discussions of affix semantics are predominantly concerned with inflection, and derivation has traditionally been less at the centre of attention. Lieber even claims that "[i]t seems safe to say that the most neglected area of morphological theory in the last three decades has been derivational semantics" (2012: 2108). Some recent work in this area (*e.g.* Dalton-Puffer 1996, Lieber 2004, Uth 2011) has started to focus more on the semantics of derivation, however.

The present study is also concerned with the semantics of derivational affixation: It analyses the derivatives of the English suffix *-age* that are attested in the *Oxford English Dictionary* (OED). This investigation aims to find out how the different senses of a polysemous morphological category, exemplified by *-age* derivatives in English, are structured. The first question to be asked is whether there is evidence for a single polysemous category

with multiple related senses rather than a number of homonymous suffixes that have the same form but unrelated senses. Another question is how the different senses are organised: Do they, for example, cluster around a single core sense, or do they display a different structure?

A method that has been used before to account for polyfunctional morphological categories is the semantic map approach. It has the crucial advantages of being open and flexible, and not making *a priori* assumptions with regard to the relations between different senses. As it has until now mainly been used in cross-linguistic studies of inflectional categories it needs to be adapted significantly to be used to describe a derivational suffix in a single language. This paper describes the adaptations of the semantic map method used in a diachronic study of an English derivational suffix, and shows that this adapted version can help to gain valuable insights into the semantics of derivational affixes in general.

In section 2, previous ways of creating semantic maps are introduced, followed by a description of the new, adapted semantic map method. Section 3 describes the data source for the diachronic investigation of *-age* derivatives in English, and the methods employed in the analysis. The results of this study are discussed in section 4. The semantic maps created for the different time periods under investigation are also included there. Finally, section 5 contains a summary and conclusion.

2 Semantic Maps

The semantic map approach is a relatively recent method of cross-linguistic comparison, and there is no unified procedure in producing semantic maps. A good introduction is provided in Haspelmath (2003). He defines a semantic map as "a geometrical representation of functions in "conceptual/semantic space" that are linked by connecting lines and thus constitute a network" (Haspelmath 2003: 213).

Haspelmath himself creates a number of maps for different inflectional categories in various languages. He arranges the functions that a single grammatical morpheme expresses in such a way that they form a continuous area on a larger semantic map. This larger semantic map is aimed to be universal, so it should ideally incorporate all the readings of corresponding grammatical categories across languages. These functions are then joined by connecting lines, which represent the degree of similarity between them. Such lines are added to a map if two functions are expressed by the same grammatical category in a single language. The basis for these maps is cross-linguistic data: Haspelmath claims that "it is generally sufficient to look at a dozen genealogically diverse languages to arrive at a stable map" (2003: 217).

Maps generated in this way have a number of advantages for cross-linguistic comparison. An important point is that they make no *a priori* assumptions, but are very much data-driven. For example, if a new reading is encountered in the data, it can be easily added to the map as a new function. Maps or parts of maps can also be proven wrong if they don't account for the data. Also, there are no conceptual limitations on the number or type of functions and connecting lines. And, crucially for the purposes of this investigation, semantic maps do not make a priori judgements on the structure of a morphological category. It is often claimed that word meanings cluster around a single sense (e.g. Tyler & Evans 2001 for spatial prepositions), and this assumption is also sometimes held by researchers working on grammatical morphemes. Lieber (2004), for example, admits that affixes are polysemous, but claims that they have a single core sense. Additional readings are then regarded as sense extensions of this core sense, and the core sense is the only one described by her metalanguage. It is, however, unclear how such a core sense should be established. Despite Tyler & Evans's (2001) efforts to develop criteria for finding a core sense of prepositions, it is not clear how one would find a core sense of bound morphemes. Even if there were an accepted procedure for this, it is doubtful whether every single morphological category would be structured in such a way, and to assume such a structure from the outset is a serious and unnecessary limitation. Semantic maps do not make such assumptions, but they can account for a prototypical structure of senses, or functions, if this emerges from the data. If a category exhibits a different structure though, this can also be represented on a semantic map. So no matter how a morphological category is structured, a semantic map can account for that structure.

Another major advantage for the purposes of this study is the ability of semantic maps to represent semantic change. If semantic change occurs, new functions can be added to the map or functions that have become extinct can be removed from it. The ways in which a morphological category has changed are then immediately apparent when two semantic maps are compared.

But semantic maps that are created in the way described above have at least one major disadvantage: they do not incorporate any information on the frequency of the different functions. In traditional implicational maps, a function is added if it is attested, so this is a simple yes/no question. It doesn't matter whether that function is extremely common or very rare, every function is represented in the same way. Also, the difference between unattested readings and rare, but attested ones is drastic, although this distinction may just be due to a small sample size (*cf.* Cysouw 2007: 232).

Cysouw (2007) improves the traditional semantic map design by incorporating frequency information. He mentions several different

possibilities for doing so, *e.g.* drawing the connecting lines with varying thickness depending on the frequency of co-occurrence of two functions. But the method he favours is multidimensional scaling. Here, similar functions, or analytical primitives as Cysouw calls them instead, are positioned close to each other on the map. This has the additional advantage of being able to quantify the similarity between two readings, something that is not done on traditional maps. Semantic maps generated with this technique (*e.g.* Wälchli & Cysouw 2012) incorporate frequency information, but are also different from implicational semantic maps in other respects. For example, they don't contain connecting lines between functions.

Both implicational semantic maps and those created by multidimensional scaling have been developed for cross-linguistic comparison. Questions that such analyses address are, for example, which functions, or analytical primitives, are so similar that they are often expressed by the same form, or which functions are not usually expressed by the same form. The questions that drive the present investigation are quite different, as this study is concerned with the structure of a single morphological category in one individual language. Because the aims of this study and, as a result, the data taken into account, are quite different from large scale cross-linguistic comparisons, the methods employed by Haspelmath (2003), Cysouw (2007), and Wälchli & Cysouw (2012) cannot be used here. In order to account for the semantics of a single suffix in one language, the semantic map approach had to be adapted significantly.

In the adapted semantic map approach, the main elements of traditional semantic maps, semantic functions and connecting lines, are retained. Instead of using the term *functions*, I will use *readings*, however. These readings are mnemonic labels for the senses expressed by derivatives. For example, a frequently encountered reading in *-age* derivatives is ACTION, as in *spillage* 'the action of spilling' or *marriage* 'the action of getting married'. These words refer to various kinds of actions, which are all subsumed under this label. As the creation of semantic maps is a bottom-up method, these readings are not specified beforehand. If a derivative cannot be classified into one of the already existing groups, a new label is created. The labels thus established in this investigation are ACTION, CONDITION, POSITION, GENERAL ABSTRACT, AMOUNT, CHARGE, RIGHT, TENURE, COLLECTIVE, PERSON, LOCATION, and OBJECT.

The second element of a semantic map, the connecting lines, represents the overlaps of polysemous derivatives. Most formations have more than one reading, so they are members of more than one reading group. A connecting line between each of the readings expressed by a derivative is drawn in these cases.

The representation of both readings and connecting lines is affected by how often they occur in the data. The boxes surrounding individual readings are scaled to size. If a reading occurs only rarely, it is small, if it occurs frequently, it is larger. Similarly, the connecting lines vary in thickness depending on their frequency: if an overlap is only attested once or twice, the line is thinner than if an overlap occurs more often. The details of this scaling process are described in section 3.3.

The semantic maps that are created in this way retain many of the key advantages of traditional semantic maps, most importantly the ability to account for language change, the flexibility to incorporate or remove individual readings, and the lack of *a priori* assumptions regarding the structure of a morphological category. But the adapted model also improves the traditional method by incorporating information on the frequency of occurrence of individual readings. This provides further clues on the structure of the category in question.

3 Data and Method

The data for this study come from the *Oxford English Dictionary online* (OED), an extremely large historical dictionary. The OED is a suitable source for morphosemantic studies not just because of its size – it also contains valuable meta-information on its entries. One of these is the date of first attestation of each entry, which enables the researcher to search for neologisms from a certain period. This feature is exploited in the present study. Neologisms from Middle English (ME) with first attestation dates from 1100-1499 can thus be compared to neologisms from Present Day English (PDE) with attestation dates after 1900.

Only transparent derivatives of the suffix *-age* are considered for this study. As the search mechanism of the OED can only search for strings of letters and not for affixes, the results had to be cleaned up. The search for word-final < age > in nominal headwords yields many items that are not derivatives of the suffix *-age*, for example *age* or *cage*. Opaque derivatives of *-age* like *language* and *message* are also excluded from the study. These words are *-age* derivatives in French and were borrowed as such into English, but their bases are not attested outside of these formations in the receiving language. Such words cannot provide information on the semantics of the suffix, as the suffix cannot be easily separated from the derivative. In order to be kept in the result file, derivatives have to contain a base that is either independently attested in the OED at around the same time, or the base has to be present in other derivatives with comparable semantics. An example for a transparent derivative is *teacherage* 'a house or lodgings provided for a teacher. But

formations like *vintage* 'the produce or yield of the vine', for example, are also kept, because other derivatives like *vinter* 'a vintner' or *vintry* 'a place where wine is sold or stored' are attested as well. As the *vint*- element in these formations has a comparable meaning to that in *vintage*, it is likely that *vintage* is perceived as semantically transparent.

3.1 Classification of readings

All of the transparent -age derivatives then have one or more readings assigned to them, depending on the semantic paraphrases provided in the OED. These paraphrases contain formulations that can be used to classify the derivatives into groups. The expressions 'action of doing' or 'process of doing', for example, are used frequently in the OED, and derivatives described by these are classified as having an ACTION reading in this study. Similarly, the phrases 'state of being', or 'condition of being' indicate a CONDITION reading, and collectives are usually described by the expressions 'collectivity of or 'collectively'. In most cases, the classification on the basis of these paraphrases is thus quite straightforward. Some descriptions require a certain amount of interpretation, however, because they do not provide any obvious indicators like the ones outlined above. A good example for this is victorage 'victory'. This paraphrase is only minimal and doesn't use any of the formulaic expressions that are found in other entries. In such cases the quotations provided in the OED for each entry and sense, and the paraphrases for the words that are mentioned, for example *victory*, are considered as well. In spite of this, most words can be assigned to reading groups without problems.

3.2. Core sense

The term *core sense* is often used without specifying exactly what is meant by it. Tyler & Evans (2001) have lamented the lack of reliable criteria in establishing different senses and the relations between these with regard to prepositions. For spatial prepositions, they offer a set of criteria to do just that, but unfortunately these cannot merely be replicated and reused for the different senses of derivational affixes.

In this study, a core sense is assumed to be a productive sense, which means that this sense is expressed by neologisms. An obvious indicator for the productivity of a reading is its type frequency in neologisms – if a reading is common in newly coined derivatives, it can be assumed to be productive. Another factor is the transparency, or semantic uniformity of the derivatives in a reading group, because transparency is a prerequisite for productivity. If most of the derivatives with a certain reading are similar regarding form and/or semantics, this reading is transparent and thus potentially productive. Good examples are the derivatives with an AMOUNT reading in PDE. The

analysis in 4.2 shows that this group of derivatives is both structurally and semantically unified. Most bases refer to units of measurement, *e.g.* in *headage, minutage*, or *gallonage*, and the derivatives all refer to an amount of the unit that is denoted by the base. The derivatives with a GENERAL ABSTRACT reading, however, are much more diverse, which makes this reading far less transparent than AMOUNT. And due to this lack of transparency, GENERAL ABSTRACT cannot be considered productive, in spite of its high type frequency (see the analyses in section 4 for a more detailed discussion).

Indicators for the transparency of a reading are the amount of hybrid formations with a Germanic base in ME, as the suffix itself is borrowed from French, and hybrid formations prove that a word formation pattern is used independently of mere borrowing. Other factors are the amount of monosemous items in a reading group and the regularity of additional senses, as both of these contribute to the systematic nature of a reading.

Although different methods for calculating the productivity of affixes have been proposed in the literature (*e.g.* Baayen & Renouf 1996), none of these can be used in the context of the present study. Such measurements rely on corpora and the information that can be extracted from them, for example the number of hapax legomena. The token frequencies of different types, which are crucial for most measurements of productivity, cannot be obtained from dictionary data, however. Dictionary-based studies thus have to rely on other methods to determine the productivity of word formation processes.

3.3 Creation of semantic maps

The readings established in the way outlined above provide the functions used in the semantic map. Depending on the number of times each of them occurs in the data, their size on the map varies. In order to be able to directly compare different maps with each other the size is determined by the relative number of the readings, *i.e.* the percentage of the number of types with one reading out of all types. The reading AMOUNT, for example, contains 26% of all types in PDE, but only 5% of all types in ME. Comparing the absolute numbers, 20 in PDE and 7 in ME, would be problematic, because the overall number of neologisms in ME is much higher than that in PDE.

The size of the readings on the map is staggered in steps of 5%:

0,1% – 4.9%: 1cm x 0.5cm 5% – 9.9 %: 2cm x 1cm 10% – 14.9%: 3cm x 1,5cm 15% – 19.9%: 4cm x 2cm etc.

A connecting line between two readings is drawn if a single derivative

expresses both of these readings. The lines vary in thickness according to the absolute number of times this overlap occurs:

1-2 occurrences: dashed line

- 3-5 occurrences: continuous line, thickness 0.0
- 6-10 occurrences: continuous line, thickness 0.1
- 11-15 occurrences: continuous line, thickness 0.2
- 16 20 occurrences: continuous line, thickness 0.3

4 Results and Discussion

4.1 -age derivatives in Middle English

Altogether 142 formations can be analysed as transparent *-age* derivatives coined in Middle English. These derivatives are classified into 12 different readings. The distribution of these readings according to type frequency and examples for each reading are shown below in table 1. The total number of types is bigger than 142, because many derivatives are polysemous and are put into more than one reading group.

reading	no. of types	examples
ACTION	59	lighterage, mockage, rivage
CHARGE	48	hidage, murage, stallage
LOCATION	24	baronage, cottage, reclusage
GEN. ABSTRACT	20	advantage, alliage, lovage
COLLECTIVE	19	cordage, cousinage, lastage
CONDITION	19	bondage, cousinage, marriage
OBJECT	14	altarage, fardellage, murage
RIGHT	10	coinage, passage, pickage
AMOUNT	7	portage, superplusage, usage
TENURE	5	bondage, thanage, villeinage
POSITION	5	bondage, parentage, thanage
PERSON	4	hostage, marriage, personage

Table 1: Distribution of readings according to type frequency (ME)

It is immediately evident from this table that there are significant differences in the type frequency of the readings encountered in ME neologisms. The two most frequent readings are ACTION, which accounts for more than a third of all types, and CHARGE, which is expressed by a third of all types. The next most common interpretation is LOCATION, but its type frequency is much lower than that of CHARGE and ACTION. From here, the type frequency decreases steadily to the rare reading PERSON, which contains only four types. Such a distribution could, of course, be illustrated by other methods than semantic maps -a bar chart, such as that in figure 1, also shows these differences in frequency. But the semantic map in figure 2 has the added benefit of also depicting the overlaps of polysemous derivatives, which exposes the structure of a morphological category in a completely new way. The differences in frequency would, however, be lost on a traditional semantic map. The adapted method is able to illustrate them, and is thus a much richer representation of the data.



Figure 1: Distribution of readings according to type frequency (ME)

The connecting lines on the semantic map also show considerable variation regarding frequency. Given the high number of types with an ACTION or CHARGE reading it is perhaps not surprising that the thickest lines are connected to those two interpretations. But RIGHT, which is one of the less common readings, is connected to both of these readings by very thick lines. This shows that a reading doesn't have to be extremely frequent to exhibit very regular semantic overlaps with other groups.

Apart from the varying thickness of the lines it is obvious that there are a large number of these that connect virtually all readings with one another. No reading is isolated, which means that all contain polysemous derivatives. Middle English *-age* neologisms thus clearly form a highly interconnected set, which suggests a polysemous category instead of multiple homonymous affixes.



Figure 2: Semantic map of ME neologisms

This answers the initial research question posed in section 1, but it does not provide evidence for the finer details of the structure of this morphological category. To find out whether there are a single or maybe multiple core senses of *-age*, and how the other senses are related to that/these core senses, each of the readings has to be analysed separately. As the space for such a detailed analysis of each reading is limited in an article such as this one, only the most striking examples are discussed in the following section.

Three different readings, ACTION, CHARGE, and COLLECTIVE, are analysed in an exemplary fashion. The first two have a high type frequency, and they are thus potentially productive readings. COLLECTIVE has a lower type frequency, but it is often claimed to be a core interpretation of *-age (e.g.* Lieber 2004).

The reading ACTION is the most frequent reading in ME neologisms, and it is also connected to nearly all other readings by polysemous derivatives, as can be seen on the map in figure 2. This clearly makes ACTION a candidate for a core reading of this morphological category. Many of the derivatives in this group are coined in early ME, i.e. before 1350. Examples are pilgrimage 'the action or practice of making [...] a journey to a sacred place' (c1275) and towage 'the action or process of towing or being towed' (a1327). This is below the share of early ME coinages among all -age neologisms, so that this reading cannot be considered particularly productive in the earliest period of the suffix's adoption into English. The number of neologisms with this reading rises significantly towards the late ME period, however. A good indicator for the transparency of this reading is the high amount of hybrid formations in this group. Nearly a quarter of the ACTION derivatives contain bases that are definitely Germanic, which equals almost half of all hybrids in the data. This pattern was thus probably perceived as a transparent means of word formation by the 15th century at the latest. The share of monosemous formations also points into this direction. Almost half of all items with an ACTION reading are monosemous, and most of the others show highly regular overlaps especially with CHARGE. This makes the derivatives highly predictable and increases their semantic transparency. Another interesting feature concerns the nature of the bases in this group. Many, slightly less than a quarter, are deverbal, as might be expected for action nouns. About the same number are clearly denominal derivatives though, e.g. putage 'harlotry, prostitution' from pute n. 'a prostitute', or brokerage 'the action [...] of a broker'. Most of the remaining items can be interpreted as either deverbal or denominal. The frequency and regularity of denominal formations with an ACTION reading clearly show that -age suffixation does not merely transpose verbs into nouns.

These facts suggest that ACTION is a highly transparent reading by late ME.

Taken together with the high number of types with this interpretation, this reading is clearly productive and can thus be considered a core sense of *-age*. The next most frequent reading, CHARGE, gives rise to 48 derivatives. This reading is expressed by some of the earliest transparent *-age* derivatives in English. Also, more than 20% of the CHARGE formations are hybrids, which is higher than the average and comparable to the share of hybrids in the ACTION group. CHARGE contains mostly monosemous items. The derivatives with additional readings have highly regular polysemies and mostly refer to either an ACTION or a RIGHT as well as a CHARGE. Structurally, the derivatives in this group are most likely based on nouns, *e.g. barbicanage* 'tribute paid for the construction and maintenance of barbicans' from *barbican* n. 'an outer fortification or defence to a city or castle', or *pontage* 'a toll for the use of a bridge' from *pont* n. 'a bridge'.

The derivatives with this reading are very early and are clearly semantically transparent formations. As they are also quite frequent, this reading can be assumed to be productive and is thus a core reading of *-age* in ME.

A reading with medium frequency is COLLECTIVE. Most of the 19 derivatives with this reading are first attested in late ME. The derivatives in this group are overwhelmingly denominal, consider for example *baggage* or *peerage* 'the body of peers; peers as a class'. The base nouns usually denote countable objects or persons, and the derivatives can generally be interpreted as denoting the collectivity of objects or persons denoted by the base, so their interpretation is highly systematic. However, only one hybrid formation, *lastage* 'the ballast of a ship', can be found in this group, and only a quarter of derivatives are monosemous.

Because the interpretation of the COLLECTIVE derivatives is so systematic and their structural make-up is so regular, they cannot be considered semantically opaque. The type frequency of this reading is quite low, however, and there are hardly any hybrid formations attested in the data, so this reading cannot be considered productive in ME.

An analogous analysis of all other readings found in the data reveals that ACTION and CHARGE are the only clearly productive readings in ME. These two can be considered core senses of *-age* in that period. It is interesting that the two core readings are linked closely by a high number of polysemous derivatives. Such formations often express an action and the charge that has to be paid for that action, *e.g. cranage* 'the use of a crane to hoist goods' (ACTION) and 'dues paid for the use of a crane' (CHARGE). ACTION is probably the more generic reading of the two, and the derivatives with CHARGE readings often implicitly refer to an action even when they are monosemous, *e.g. pickage* 'a fee paid for breaking the ground and setting up a booth, stall, tent [...]'. CHARGE can therefore be seen as a sense extension of

ACTION. Some other readings, like COLLECTIVE or LOCATION, are transparent, but due to their relatively small type frequency they are not as productive as ACTION and CHARGE. A number of the low-frequency readings like PERSON cannot be considered transparent, and they are therefore certainly not productive. The many connecting lines linked to ACTION also show that this reading in particular lies at the heart of the morphological category.

4.2 -age derivatives in Present Day English

Since 1900, 78 transparent *-age* derivatives are recorded in the OED. They have similar readings as the ME neologisms, but not all of the readings attested in ME are present in PDE, as table 2 shows.

reading	no. of types	examples
ACTION	40	creepage, flamage, mud pilotage
AMOUNT	20	headage, minutage, pointage
GEN. ABSTRACT	20	air mileage, frottage, voidage
COLLECTIVE	11	screenage, signage, twiggage
OBJECT	11	pre-package, spillage, stillage
LOCATION	6	coverage, parachutage, teacherage
CONDITION	3	plaçage, problemage, victimage
CHARGE	1	warehouseage
RIGHT	0	
TENURE	0	
POSITION	0	
PERSON	0	

Cable 2: Distribution	of readings	according to t	ype frequency	(PDE)

The most frequent reading in PDE neologisms is ACTION – more than half of all types express this reading. The next most frequent readings are AMOUNT and GENERAL ABSTRACT, but these already have a far lower type frequency than ACTION. Two of the readings found in ME neologisms, CONDITION, and CHARGE are extremely rare in PDE coinages, and four ME readings are not attested in PDE derivatives at all. CHARGE, which is a productive reading in ME with a high type frequency only gives rise to a single new formation in PDE. There has clearly been significant change in this word formation pattern since ME.

The semantic map in figure 3 below illustrates these differences in type frequency. The reading ACTION clearly dominates the map, and most of the other recorded readings, apart from AMOUNT and GENERAL ABSTRACT, are quite small. This is a difference from the ME map, where readings of all sizes, *i.e.* frequencies, are attested.

Another obvious difference between ME and PDE is the number of connecting lines. The PDE map contains much fewer of these, which is, of course, partly due to the overall smaller number of readings found in that period. But it is also the case that each of the readings found in PDE has much fewer connections to other readings than before. Most readings show two or three connecting lines, and two are completely isolated. Only ACTION and GENERAL ABSTRACT are connected to more than three other readings. Most of the connecting lines are also continuous, representing three or more derivatives with a particular polysemy. In ME, most lines are dashed, which indicates that only one or two derivatives show that semantic overlap. Given the differences in the number of neologisms in these two periods – there are twice as many coinages in ME – this is even more significant. Although the overall number of types is smaller in PDE, the polysemies shown by individual derivatives are more regular than in ME.

All of this leads to a much simpler map and a clearer picture of the polysemies shown by *-age* neologisms in PDE. The dominance of ACTION in terms of type frequency and the smaller number, but higher regularity of polysemous derivatives suggests that *-age* suffixation is more concentrated on a smaller number of readings and polysemies in PDE than it is in ME. A semantic analysis of the readings ACTION and AMOUNT in the remainder of this section will throw more light on these changes. These two are the most frequent readings in PDE.

It was already stated above that ACTION clearly dominates this morphological category in terms of type frequency. This reading was already found to be a core sense in ME, and, given its prevalence, it is likely that it still has that status in PDE. As this period is much shorter than that of ME, a consideration of the attestation dates of PDE neologisms is not particularly interesting. Also, hybrid formations are not analysed in this part, as the etymology of base words is likely unknown to speakers, and this feature does not contribute information on the transparency of the formations.

Almost 30% of the ACTION derivatives contain exclusively nominal bases, and only 16% are clearly deverbal. Most of the remaining derivatives are either denominal or deverbal, as the bases are attested in both forms. The share of denominal formations is thus very high, even higher than the share of denominal formations with the same interpretation in ME. Nearly all of the base nouns denote a physical object or person, and the resulting derivative refers to an action connected with that object or person. A good example for this is *victimage* 'the practice of seeking out a victim'. These derivatives are highly transparent, as the ACTION reading of the derivative is clearly due to the addition of the suffix.



Figure 3: Semantic map of PDE neologisms

Almost a third of the ACTION derivatives contain exclusively nominal bases, and only 13% are clearly deverbal. Most of the remaining derivatives are either denominal or deverbal, as the bases are attested in both forms. The share of denominal formations is thus very high, even higher than the share of denominal formations with the same interpretation in ME. Nearly all of the base nouns denote a physical object or person, and the resulting derivative refers to an action connected with that object or person. A good example for this is victimage 'the practice of seeking out a victim'. These derivatives are highly transparent and systematic. More than half of the derivatives are monosemous, which also contributes to the transparency of this pattern. ACTION is clearly a transparent reading, and because it is also extremely frequent, it should be considered a productive reading as well. It also gives rise to different sense extensions, *e.g.* to OBJECT, as is indicated by the high number of derivatives with both readings shown in figure 3, and is generally closely connected to a number of different readings through polysemous derivatives. ACTION lies at the heart of this morphological category and is a core sense of -age in PDE.

AMOUNT is a reading that is not particularly frequent in ME, but shows the second highest number of neologisms in PDE. 60% of the AMOUNT formations are monosemous, and most of the remaining words show additional GENERAL ABSTRACT readings. The monosemous derivatives usually denote a number or amount of measurement units, which are denoted by the base. Good examples are *minutage* 'the amount of time for which a commercial television company is permitted to broadcast advertisements' and metreage 'the aggregate amount of metres'. The polysemous derivatives have the same AMOUNT interpretation, but they also denote an abstract concept that is the result of the amount denoted. Air mileage, for example, refers to an 'extent or distance in air miles' (AMOUNT), but also to the 'rate or efficiency of travel through the air' (GENERAL ABSTRACT). The majority of the derivatives in this group, namely 75%, are denominal formations, and the bases usually denote measurement units like gallon, watt, or metre. The remaining quarter of formations may also be denominal, and then largely conform to the semantics of the clearly denominal formations, but other parts-of-speech are also possible in these cases.

The derivatives in this group are very regular, both in their semantics and in their morphological structure, and can be considered highly transparent. This reading is also frequently found in neologisms, so it can be seen as a productive reading, and thus a core sense of *-age* in PDE.

ACTION and AMOUNT are the only two core senses of *-age* in PDE. Most of the other readings are either not frequent enough to be considered productive, or they are not transparent. The two core senses exhibit significant

differences in their morphological structure, so that the two patterns don't share a large number of polysemous derivatives. Although both readings have a large share of nominal bases, the semantics of these are quite different. The AMOUNT derivatives are based on nouns that refer to units of measurements, while ACTION formations are mostly coined on nouns denoting persons or objects. The structural differences make these two patterns quite distinct and increase each reading's transparency. This also makes it possible to predict the semantics of new *-age* derivatives. If these coinages are deverbal, or if they are denominal and the base denotes a person or an object, it is most likely that they denote an ACTION. If they are denominal and the base is likely to be counted, they probably refer to an AMOUNT of the entities denoted by the base.

5 Conclusion

This analysis has shown that adapted semantic maps can be used successfully to account for the different senses and the relations between the senses of a highly polyfunctional derivational affix. It is a data-driven and flexible method, which can illustrate the structure of a morphological category.

Due to the data considered in this particular case, namely neologisms from different periods of English, productivity judgements could be made. It turned out that *-age* does not have a single, but two core senses. In ME, the core senses are ACTION and CHARGE, and in PDE they are ACTION and AMOUNT. The change of core senses, as well as the developments of the remaining senses show that there has been significant semantic change in this word formation pattern. This change can be illustrated in a straightforward manner by the semantic maps. As the maps are enriched with frequency information, they can also help to find these core senses. The second element of the semantic maps, the connecting lines, provides further information on the relations between readings.

This method can also be used to compare the semantics of different derivational affixes in a single language, or across different languages. Such a comparison could expose regularities regarding semantic change or common sense extensions. It would, for example, be interesting to see whether other affixes with similar core senses have the same sense extensions and thus similar polysemies.
References

- Baayen, R. H., A. Renouf. 1996. "Chronicling the Times: Productive Lexical Innovations in an English Newspaper". *Language* 72: 69-96.
- Cysouw M. 2007. "Building Semantic Maps: The Case of Person Marking" in M. Miestamo, B. Wälchli (eds), New Challenges in Typology: Broadening the Horizons and Redefining the Foundations. Berlin: de Gruyter, 225-247.
- Dalton-Puffer C. 1996. The French Influence on Middle English Morphology. A Corpus-Based Study of Derivation. Berlin: Mouton de Gruyter.
- Haspelmath M. 2003. "The Geometry of Grammatical Meaning: Semantic Maps and Cross-Lingusitic Comparison" *in* M. Tomasello (ed.), *The New Psychology of Language*. Vol 2. Mahwah, NJ: Erlbaum, 211-242.
- Lehrer A. 2003. "Polysemy in Derivational Affixes" in B. Nerlich et al (eds). Polysemy. Flexible Patterns of Meaning in Mind and Language. Berlin: Mouton the Gruyter, 217-232.
- Lieber R. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge UP.
- 2012. "Semantics of derivational morphology" in C. Maienborn, K. von Heusinger, P. Portner (eds), *Semantics* (HSK 33.3). Berlin: de Gruyter, 2098-2119.
- Oxford English Dictionary 2014. Retrieved from http://www.oed.com
- Plag I. 1999. *Morphological Productivity. Structural Constraints in English Derivation*. Berlin: Mouton de Gruyter.
- Tyler A., V. Evans 2001. "Reconsidering Prepositional Polysemy Networks: The Case of over". *Language* 77: 724-765.
- Uth M. 2011. Französische Ereignisnominalisierungen. Abstrakte Bedeutung und regelhafte Wortbildung. Berlin: de Gruyter.

Wälchli B., M. Cysouw 2012. "Lexical Typology Through Similarity Semantics: Toward a Semantic Map of Motion Verbs". *Linguistics* 50: 671-710.

FROM EMERGENT AVAILABILITY TO FULL PROFITABILITY? THE DIACHRONIC DEVELOPMENT OF THE ITALIAN SUFFIX -ZIONE FROM THE 16TH TO THE 19TH CENTURY

Pavel Štichauer Charles University in Prague

Abstract

This paper¹ describes the quantitative development of the Italian suffix *-zione* within the time span that goes from the 16^{th} to the 19^{th} century. The diachronic period covered is characterised by the first codification of the *italiano letterario* and the moment in which it becomes the national language of the newly unified Italy. The article is based on the corpus data drawn from four corpora covering the traditionally defined periods. It uses a quantitative approach to morphological productivity to show, on the four different-sized corpora, the main quantitative characteristics. The paper establishes that the productivity of the suffix *-zione* is, across the four centuries, rather constant. On the basis of some complementary lexicographic evidence, it is shown that the situation might be due to the internal structure of the *-zione* formations, being mostly direct Latin borrowings, where the major verbal inputs, *i.e.* verbs in *-izzare* and *-ificare*, can be traced back to as late as the 19^{th} and 20^{th} centuries.

1 Introduction: the premises and the aim of the study

This article is a direct continuation of a previous paper (*cf*. Štichauer 2009) where the productivity of three Italian suffixes, *-zione*, *-mento* and *-gione* was investigated – dichronically, from the 13^{th} to the 16^{th} century – within the approach to morphological productivity based on Baayen's work. The present article takes up only one of these suffixes, namely *-zione*, and attempts to show its subsequent evolution from the 16^{th} to the 19^{th} century using the same methodological tools introduced in the previous work.

¹ The study is part of a larger project *Word-formation in Italian from the 16th to the* 20^{th} century financed by GAČR (Czech Science Foundation), n. P406/12/0450.

1.1 Availability/profitability

It is widely assumed that Corbin's (1987: 177) distinction between *disponibilité* and *rentabilité* (availability/profitability) has settled the basis for a distinction between a qualitative and a quantitative approach to morphological productivity. One of the possible quantitative approaches there is undoubtedly Baayen's (1992; 2001) probabilistic conception which tries to capture the productivity of a word-formation process in terms of the ratio between the number of *hapax legomena* and the overall number of tokens.

1.2 Diachronic approach

The aim of this paper is, precisely, to adopt such a probabilistic approach, using the tools specifically designed for this purpose (to be introduced below), and to apply it *diachronically*. In fact, the paper attempts to investigate the productivity of the suffix *-zione* across four centuries, from the 16th to the 19th century, and it tries to *compare* the results for each century separately in order to see whether there is or not a significant change diacronically.

1.3 The Italian suffix -zione

Even though the diachronic development of *-zione* is interesting for various reasons, it is problematic in several aspects. First, the *-zione* formations are, in most cases, directly connected to their Latin counterparts, thus being considered rather as typical loanwords (*cf.* Thornton 1990: 200: "*La stragrande maggioranze dei derivati in -zione* (...) consiste di voci entrate in *italiano come prestiti colti dal latino.*"). Second, the morphosemantic transparency is highly variable ranging from opaque formations, as an outcome of a massive borrowing process, tend to have a constant structure from the 15th century (*cf.* Štichauer 2009) until as late as the 19th century, as witnessed by the lexicographic evidence (see section 6). In view of this, the aim of the paper is thus straightforward: to offer a piece of corpus-based evidence of the evolution of *-zione* formations which would be complementary to the well-known facts from the literature.

2 Corpus / subcorpus selection

In order to carry out the intended research, the collection of Italian literary texts LIZ 4.0 was used along with the further subdivison into four subcorpora corresponding to the centuries in question. The time span covered is characterised, as already mentioned, by the first codification of the *italiano letterario* at the beginning of the 16^{th} century, and by a first (and gradual)

establishment of a *national language* at the end of the 19th century; it is thus supposed to represent a sort of homogeneous diachronic phenomenon.

However, not all the texts present in LIZ 4.0 can be assumed to be representative of *italiano letterario* as there are various dialectal works. Therefore, we proceeded to a further selection eliminating texts such as the following: Ruzante (16th century, Paduan dialect), some of Goldoni's works (18th century, some plays written in the Venetian dialect), Carlo Porta's poetical works (19th century, Milanese dialect) and Giuseppe Gioacchino Belli's *Sonetti* written in the 19th century Roman dialect.

The overall quantitave structure of the resulting four subcorpora is summarised in Table 1.

Period/subcorpus	Number of tokens	Number of texts
16 th century	10 459 937	221
17 th century	3 028 291	49
18 th century	4 600 141	237
19 th century	10 285 433	217

 Table 1: The parameters of the subcorpora

3 The data elaboration and the quantitative overview

As far as the technical details are concerned, the starting point was a *frequency list* of all types with the suffix *-zione* for each period/subcorpus separately. The traditional *TYPE/TOKEN* distinction is assumed, but the definition and identity of certain TYPES is particularly tricky due to some lemmatisation problems; the procedure was the following:

1) A simple sequence-based corpus search (all words ending in -zion/-sion/-tion etc.).

2) The *manual* lemmatisation of orthographic and phonomorphological variants (*e.g.*, *imaginazione / immaginazione / immaginazion* etc.).

3) The elimination of formations considered not to be part of the WFR in question; following Gaeta & Ricca 2002; 2003; 2006 five major elimination criteria have been adopted:

a) strong opacity (*e.g.*, *frazione* – not directly connected to *frangere*)

b) baseless formations (*e.g.*, *stazione*)

c) nominal bases (those having a collective meaning, *e.g. fogliazione*)

d) derivational inner cycles (e.g., *indeterminazione* – the basis for the prefix *in*- is *determinazione*, which is in turn derived from

determinare by means of *-zione*, *cf*. Gaeta & Ricca 2006: 79-83, Štichauer 2009: 60-61)

e) strict borrowings (*e.g.*, *interiezione*) – this is a controversial class since virtually all *-zione* formations can be said to be borrowings, but the group is here defined as containing only those words which lack completely any verbal base; in this sense this class could also be subsumed under b) *baseless formations*.²

The overall number of types (V), tokens (N) and *hapax legomena* (V_1) obtained after this post-processing is given in Table 2.

Period/subcorpus	V	V ₁	Ν
16 th century	722	194	32031
17 th century	490	96	11127
18 th century	550	94	14334
19 th century	883	130	53089

 Table 2: The number of types (V), hapaxes (V1) and tokens (N) in the four subcorpora

4 How to compare the four different-sized corpora

It is clear that the empirical data are not directly comparable both *qualitatively* and *quantitatively*. First, the subcorpora could be considered to be representative since all major texts of the *italiano letterario* are well present. However, the corpora are not of the same size, their dimension ranging from 3 to 10 milions of tokens; besides, they are also not *balanced* (the amount of texts transmitted and their typology is highly restricted due to extralinguistic reasons; *cf., e.g.*, Claridge 2008).

To obviate these problems, a principled solution has been adopted consisting in the creation of *coherent and comparable* theoretical models on the basis of *LNRE models of word frequency distributions* (*cf.* Baayen 2001, 2008). One of these models, the *Zipf-Mandelbrot* (ZM) model, implemented in the *ZipfR* package (*cf.* Baroni & Evert 2006), has been adopted to produce *expected values of V*, V_1 at a unified number of N tokens, thus making the data across the four subcorpora at least theoretically comparable.

² The difference is clearly not sufficiently justified. I maintain the the principal difference lies in the diachronic relation between *stare* – *stazione*, in which the verb is derivationally related to the noun even if the regular semantic relation is completely lost. In the case of *interiezione*, the verbal relation holds only for the starting point which is Latin *intericio*, *interiectus*, *interiectio*. In any case, the borrowing class could obviously be a special (and not numerous) case of the *baseless formations*.

5 Assessment of the productivity on the basis of expected values

The ZM model has been used throughout to produce expected values at the unified value of N = 50000, which is the empirical value of the 19th century (and so the values of the other subcorpora are *extrapolated to the highest reliable N*).

Even though this is only a *statistical model* (whose reliability can of course be measured, but the precise results are, *faute de place*, left aside), the outcomes are worth being considered. They are summarised in Table 3, in which E indicates *expected values* of the number of types (V) and *hapaxes* (V₁).

Period/subcorpus	Ν	E(V)	E (V ₁)	$P(V_1/N)$
<i>-zione</i> 16 th century	50000	813	216	0.0043
<i>-zione</i> 17 th century	50000	657	117	0.0023
<i>-zione</i> 18 th century	50000	738	165	0.0033
-zione 19 th century	50000	873	155	0.0031

Table 3: Expected values of V, V₁ at N = 50000 on the basis of the ZM model

6 Interpretation: the structure of the real data

The estimated figures show that in the time span in question there is *no significant* increase of productivity. The qualitative and quantitative aspects of the empirical data lead to several important observations:

a) Even though there is vocabulary growth (the number of types is higher in the 19^{th} century than in the 16^{th} century in two corpora of the same size), the contribution of *hapax legomena* seems to be marginal.

b) The *-zione* formations tend to have the same token frequency with the exception of the 19th century (the relative token frequencies for the four subcorpora are 3,06 ‰, 3,67 ‰, 3,11 ‰ and 5,16 ‰, respectively) where the token frequency is increasing – a sign of a future evolution (*cf.* Gaeta & Ricca 2006 who report that *-zione is* by far the most frequent suffix in present-day Italian).

c) As is well known from the literature (Gaeta & Ricca 2006; Thornton 1990; 1991), among the new formations ending in *-zione* there are, in the present-day Italian, mainly the *-izzazione* nouns. This is confirmed also by the lexicographic evidence, as summarised in Table 4 below. It is interesting to note that the increasing productivity of *-zione* is, in this sense, driven by one particular subclass of verbal bases, those in *-izzare*, the productivity of which is thus a *condition sine qua non*...

Dictionary	Verbs in -	Nouns in -
	izzare	izzazione
Vocabolario degli Accademici	55	0
della Crusca (1612)		
Il Tommaseo (1857-1879)	288	35
Dizionario Italiano Sabatini-Coletti	636	522
(2004)		

 Table 4: Lexicographic data for the verbs in -izzare and nouns in

 izzazione

The corpus data confirm this picture. The 16^{th} century subcorpus contains just one *-izzazione* formation (*particolarizzazione*), while in the 19^{th} century corpus 24 nouns can be found (*e.g.*, *centralizzazione*, *democratizzazione*, *generalizzazione*, *settentrionalizzazione*, etc.), not all of them being among the *hapaxes* (but certainly fall within low-frequency items).

The structure of the *hapax legomena* is also variable across the four centuries. There are, of course, new *loanwords* from Latin, but also genuine Italian neologisms which, in both cases, compete with their *-mento* counterparts (*e.g., investigazione - investigamento*, 17th century; *migliorazione - miglioramento*, 18th century; *collegazione - collegamento*, 19th century).

The structure of the *most frequent types* tends to be constant over the four periods under investigation. It is undoubtedly in line with what can be found in the literature (*cf.* above). The diversities between the centuries can be ascribed to the presence of specific texts.

The most frequent types represent, in virtually all cases, formations which are still part of the present-day Italian lexicon (and some of them are, of course, part of a larger *international Latinate basis of the lexicon*).

7 Conclusions

The corpus data mostly confirm the well-known lexicographically established situation about the quantitative and qualitative nature of *-zione* formations (*cf.* Thornton 1990; 1991). A detailed scrutiny of the *hapax legomena* is to be carried out, especially where a given *hapax* can safely be said to be a genuine Italian formation, not a direct Latin borrowing, and especially where the formation in question is no longer in use but has been superseded by a rival formation (*e.g. abituazione* which is a frequent formation of the 18th century, but totally absent in the following century; *perdonazione - perdono* etc.).

This detailed scrutiny is left for future elaboration (*cf.* Štichauer, *in preparation*).

References

- Baayen H. R. 1992. "Quantitative aspects of morphological productivity" *in*G. Booij, J. van Marle (eds). *Yearbook of Morphology 1991*. Dordrecht: Kluwer, 109-149.
- 2001. Word frequency distributions. Dordrecht: Kluwer.
- 2008. Analyzing Linguistic Data. A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press.
- Baroni M., S. Evert 2006. "The *zipfR* package for lexical statistics: A tutorial introduction" (http://zipfr.r-forge.r-project.org/).
- Claridge C. 2008. "Historical Corpora" in A. Lüdeling, M. Kytö (eds), Corpus Linguistics. An International Handbook, Berlin: Mouton de Gruyter, vol. 1, article 14, 242-259.
- Corbin D. 1987. *Morphologie dérivationnelle et structuration du lexique*, 2 voll. Tübingen: Niemeyer (reprint 1991: Villeneuve d'Ascq: Presses Universitaires du Septentrion, coll. Sens et structures).
- Gaeta L., D. Ricca 2002. "Corpora testuali e produttività morfologica: i nomi d'azione in due annate della *Stampa*" in R. Bauer, H. Goebl (eds), *Parallela IX. Testo variazione informatica. Text Variation Informatik.* Wilhelmsfeld: Gottfried Egert Verlag, 223-249.
- 2003. "Frequency and productivity in Italian derivation: A comparison between corpus-based and lexicographical data." *Italian Journal of Linguistics/Rivista di Linguistica* 15, 1: 63-98.
- 2006. Productivity in Italian word formation: A variable-corpus approach. Linguistics 44, 1: 57-89.

- LIZ 4.0. 2001. Letteratura italiana Zanichelli. CD-ROM dei testi della letteratura italiana, a cura di Pasquale Stoppelli ed Eugenio Picchi. Bologna: Zanichelli.
- Štichauer P. 2009. "Morphological productivity in diachrony: the case of the deverbal nouns in *-mento*, *-zione* and *-gione* in Old Italian from the 13th to the 16th century". *in* F. Montermini, G. Boyé, J. Tseng (eds), *Selected Proceedings of the 6th Décembrettes*. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2241, 138-147.
- Štichauer P. in preparation. La formazione delle parole in diacronia. Studi di morfologia derivativa dell'italiano tra il Cinquecento e l'Ottocento. Prague: Charles University Press.
- Thornton A. M. 1990. "Sui deverbali italiani in -mento e -zione (I)". *Archivio glottologico italiano*, LXXV/II: 169-207.
- 1991. "Sui deverbali italiani in -mento e -zione (II)". Archivio glottologico italiano, LXXVI/I: 79-102.

MULTIPLE DERIVATION IN FRENCH DENOMINAL ADJECTIVES

Jana Strnadová

Paris Diderot University & LLF, Charles University Prague

Abstract

This paper provides an overview of the morphological and phonological properties of nouns from which French adjectives are derived. Two questions arise with respect to the data. First, is the relation between an adjective and its base independent of previous morphological processes? Second, how is the derivation history reflected in subsequent derivation? Three situations with complex suffixed base nouns—deadjectival, deverbal and denominal nouns—are investigated in detail. We show that the attractiveness of some suffixal combinations together with the semantic type of the base play a central role in multiple derivation. Therefore, we argue that the derivation history impacts on the possibility of adjectival derivation insofar as it constrains the semantic and morphophonological properties of base nouns.

1 French Denominal Adjectives

In recent years, French adjectives derived from nouns have been the focus of much research. The word-formation patterns involving the suffixes *-esque* (Plénat, 1997), *-ien* (Lignon, 2000), *-ier* (Roché, 2004) or *-eux* (Fradin, 2007) have been studied from morphophonological and semantic points of view and denominal adjectives have often been discussed in connection with the notion of relational adjective (Mélis-Puchulu, 1991, Roché, 2006, Fradin, 2008). However, some suffixes have not been investigated in sufficient detail (*-aire, -al, -el, -ique*) and an overall quantitative and qualitative analysis based on extensive data is missing.

In this paper, we examine the types of nouns from which French adjectives are derived, with special focus on multiple derivation. This could help explain why not all types of nouns are easily adjectivized in French, contrary to what happens for example in Slavic languages. Example (1) illustrates the case of deverbal suffixed and converted nouns from which no adjective has been derived. This raises the question of the impact of the derivation history on subsequent derivation and the possibility of deriving adjectives from complex base nouns. In contexts such as those in (2) and (3), French uses prepositional phrases where English tends to use NN compounds and Czech has denominal adjectives which precede a head noun.

- a. DÉCOLLER 'take off' → DÉCOLLAGE 'take off' → ?
 b. TRANSPORTER 'to transport' → TRANSPORT 'transport' → ?
- (2) *piste de décollage* (fr) *runway* (en) *vzletová dráha* (cz)
- (3) avion de transport (fr) transport plane (en) dopravní letadlo (cz)

After a brief presentation of the data set in section 2, we examine some properties of nouns from which French adjectives are derived in section 3, providing an overview of their morphological and phonological properties. We then proceed to three case studies in section 4, with base nouns derived from a verb, an adjective or a noun. In section 5, we summarize and discuss our conclusions about how the derivation history is reflected in subsequent derivation and whether the relation between an adjective and its base is independent of previous morphological processes.

2 Data set

This work is based on a list of 4,422 lemmas which have been obtained from the combination of two sources. First we used the lexical resource DenALex (Strnadová & Sagot, 2011), which contains regular derivational relations between French nouns and derived adjectives. Candidate derivatives were first constructed from existing French nouns by using derivational rules for each of the suffixations mentioned in Table 1¹. These rules deal with suffix variations (*-el/-iel/-uel*) and they were completed by morpho-graphemic rules based on addition, deletion or change of characters and dealing with base variations (*-eur* \rightarrow *-or*), so that candidates such as those in (4) could be constructed. These candidates were then searched automatically in the lexica Morphalou (Romary *et al.*, 2004) and Lefff (Sagot, 2010) and in three corpora : a newspaper corpus (l'Est républicain), French Wikipedia and GoogleNGrams. They were then manually validated.

(4) a. FONCTEUR 'functor' \rightarrow FONCTORIEL 'functorial'

¹Other suffixes can also be used but either they impose specific constraints on the semantic type of the base (*-ais, -ain, -ois* occur with toponymic bases: ISLANDE 'Iceland' \rightarrow ISLANDAIS 'Icelandic', AFRIQUE 'Africa' \rightarrow AFRICAIN 'African', SUÈDE 'Sweden' \rightarrow SUÉDOIS 'Swedish') or they occur only with few lexemes (*-ard*: MONTAGNE 'mountain' \rightarrow MONTAGNARD 'from the mountain').

Suffix	Base Noun	Derived Adjective
-aire	CONSUL 'consul'	CONSULAIRE 'consular'
-al	PARENT 'parent'	PARENTAL 'parental'
$-el^2$	CULTURE 'culture'	CULTUREL 'cultural'
-esque	CARNAVAL 'carnival'	CARNAVALESQUE 'of carnival'
-eux	ANGINE 'angina'	ANGINEUX 'anginal'
-ien	MICROBE 'microb'	MICROBIEN 'microbial'
-ier	CÔTE 'coast'	CÔTIER 'coastal'
-ique	MÉTHODE 'method'	MÉTHODIQUE 'methodical'
- <i>u</i>	FEUILLE 'leaf'	FEUILLU 'leafy'

b. RÉSIDENCE 'residence' \rightarrow RÉSIDENTIEL 'residential'

Table 1: Examples of Adjectival Suffixation	ı in	French
---	------	--------

We complemented this data set with adjectives from Lexique3 (New, 2006), a database of French words containing frequency information. The result is that we have both well established (from Morphalou, Lefff and Lexique3) and rare or recently coined (from corpora and the web) lexemes. The distribution of our data according to the sources is presented in Table 2.

Source	Adjectives
DenALex - lexica	2,176
DenALex - corpus, web	1,643
Lexique3	603
Total	4,422

 Table 2: Data Sources

This study considers only free forms as bases and thus leaves aside those adjectives from Lexique3 which exhibit idiosyncratic allomorphy of the base or

²The suffixes *-el* and *-al* are sometimes considered allomorphs of one suffix, as they both come from Latin *-alis* (*-al* was borrowed from Latin whereas *-el* was inherited). They also present the same allomorphs in further derivation (NATIONAL 'national' \rightarrow NATIONALITÉ 'nationality', MENSUEL 'monthly' \rightarrow MENSUALITÉ 'monthly rate'). We present them separately because in synchrony, they do not have a complementary distribution (they both attach to nouns ending with *-eur* for example : ÉQUATEUR 'equator' \rightarrow ÉQUATORIAL 'equatorial', FONCTEUR 'functor' \rightarrow FONCTORIEL 'functorial') and there are doublets with contrasting meanings, such as STRUCTUREL and STRUCTURAL 'structural'.

suppletion, such as the examples in $(5)^3$. We also limited our attention to suffixation, even though denominal adjectives can also be created through noun \rightarrow adjective conversion (6a) or by prefixation (6b). Finally, we exclude adjectives derived from proper nouns.

- (5) a. MOINE 'monk' \sim MONACAL 'monastic'
 - b. OISEAU 'bird' ~ AVIAIRE 'avian'
- (6) a. PIÉTON 'pedestrian' in rue piétonne 'pedestrian street'
 - b. ANTIRIDES 'anti-wrinkle' in crème antirides 'anti-wrinkle cream'

3 Properties of Base Nouns

3.1 Morphological Complexity of Base Nouns

Since no adjective is derived from deverbal nouns like DÉCOLLAGE 'take off' or TRANSPORT 'arrival', morphological complexity could be a possible constraint on input for French adjectives. Therefore, we established the morphological type of the nouns which serve as inputs for adjectival derivation, combining the use of *Dérif* (Namer, 2009) and manual validation.

The results we obtained are presented in Table 3. As the Ratio column shows, 54% of adjectives from our data are derived from simple nouns, 31% from compounds⁴ and only 13% from suffixed nouns, for which three subcategories are distinguished: denominal, deadjectival and deverbal nouns.⁵

Naturally, without knowing the distribution of the types in the lexicon in general, one cannot evaluate these results. Thus, Figure 1 presents the following comparison: the light grey columns represent the types of nouns occurring as bases of denominal adjectives, the dark grey columns correspond to the distribution of morphological types of nouns in the lexicon in general. These counts are based on 4 samples of 500 French nouns randomly chosen from Lexique3 (New, 2006).

³The \sim sign stands for a relation between two lexemes which is not derivational but rather lexical. Derivational relations are represented by the \rightarrow sign.

⁴It is worth noting that no native VN or NN compound is present in our corpus as a base, whereas 20% of all adjectives are derived from the so-called neoclassical compounds (Namer, 2007), the analyzability of which is questionable.

⁵For 5% of the corpus, the type cannot be determined due to the uncertainty of the orientation of conversion (Tribout, 2010): in the pair TRIOMPHE 'triumph' \rightarrow TRIOMPHAL 'triumphant', the noun can be analyzed either as a simplex or as a deverbal noun converted from TRIOMPHER 'to triumph'. Another problem is that such a typology implies a unitary base approach which is not always appropriate (Bochner, 1993, Strnadová, to appear).

MORPH. TYPE	RATIO	BASE NOUN	DERIVED ADJECTIVE
Simple N	54%	DENT 'tooth'	DENTAL 'dental'
Compound N			
Neo-Classical	31%	LYMPHOCYTE	LYMPHOCYTAIRE
		'lymphocyte'	'lymphotic'
Native	0%	LAVE-VAISSELLE	
		'dishwasher'	
Profived N	10/	ANTICYCLONE	ANTICYCLONAL
riciixeu îv	1 70	'anticyclone'	'anticyclonic'
Suffixed			
Denominal N	4%	FIBRILLE 'fibril'	FIBRILLEUX 'fibrilous'
Deadjectival N	J 2%	ÉGALITÉ 'equality'	ÉGALITAIRE 'egalitarian'
Deverbal N	7%	ORNEMENT 'ornamen	t'ORNEMENTAL 'ornamental'
Other	1%	OVNI 'ufo'	OVNIESQUE 'ufonian'

 Table 3: Distribution of Morphological Types of Base Nouns



Figure 1: Distribution of morphological types in the full nominal lexicon and in the set of bases of denominal adjectives

Two striking tendencies can be observed. On the one hand, simple nouns and neoclassical compounds are overrepresented as bases of French derived adjectives. On the other hand, suffixed nouns, and in particular deverbal nouns, are strongly underrepresented.

For that reason, we propose to explore in section 4 the cases with derivation history $X \rightarrow Nsfx \rightarrow A$. The case where X is an adjective is unusual because it entails a derivational cycle, which is not normally expected $(A \rightarrow N \rightarrow A)$. Where X is a verb, the observation in need of an explanation is the unexpectedly low prevalence of deverbal nouns (7%) as bases of adjectives given the proportion of deverbal nouns in the overall nominal lexicon (17%). In both cases, there is a category change (A \rightarrow N, V \rightarrow N), which may have an impact on subsequent derivation. Finally, the situation where X is a noun is interesting inasmuch as it involves two lexemes of the same category (N \rightarrow N) which will have repercussions on the polysemy of the adjective.

3.2 Length of Base Nouns

The length of the phonological form of a base is a possible factor hampering subsequent derivation. As Plénat & Roché (2003) observe, in French, the length of the optimal base for suffixation by *-esque* or *-issime*, has two syllables. Suffixed nouns are in general longer than simple nouns and this could explain why suffixed nouns are less frequent as bases of denominal adjectives.

If we examine the number of syllables of base nouns in French denominal adjectives, as presented in Table 4, we observe that most of the adjectives are derived from disyllabic (29%) or trisyllabic (29.5%) nouns. 4 syllable nouns are also well represented (20.5%), but 70% of them are bases for adjectives derived with *-ique*. Almost all (90%) adjectives derived from even longer bases are suffixed in *-ique*. This fact correlates with the high number of neoclassical compounds from which *-ique* adjectives are derived.

Suffix	1syll	2syll	3syll	4syll	5yll	6syll	7syll	8syll	Total
-aire	52	141	175	86	21	3			478
-al	62	170	152	43	8				435
-el	14	91	169	93	13	3			383
-esque	17	42	50	6					115
-eux	173	241	94	18					526
-ien	13	54	66	22	6	2			163
-ier	63	134	31	5					233
-ique	97	361	566	632	282	91	15	7	2,051
- <i>u</i>	32	5	1						38
Total	523	1,239	1,304	905	330	99	15	7	4,422
					-				

Table 4: Distribution of denominal suffixes by length of the base noun

Once more, we should compare this data with the number of syllables of nouns in general. Figure 2 illustrates this comparison: the light grey line represents base nouns of denominal adjectives, the dark grey line corresponds to the 30,435 nouns of Lexique3.

Surprisingly, the comparison does not show any important difference in the preferences of the adjectival suffixation. The proportion of monosyllabic nouns



Figure 2: Distribution of nouns by length in syllables, in the full nominal lexicon and in the set of bases of denominal adjectives

is the same in the full nominal lexicon and in the sublexicon of bases. Likewise, a higher number of syllables does not seem to preclude further derivation, in particular, in the case of neoclassical compounds, which constitute most of the cases with more than 5 syllables.

3.3 Morphological Niches

It has been observed in previous studies that some affixes favor the attachment of other affixes (Williams, 1981, Lindsay & Aronoff, 2013). This phenomenon is known as potentiation and it seems to have an important role in the derivation of adjectives from nouns in French.

Table 5 shows the frequencies of the most important affix combinations. The columns represent the final segment of the nouns. The rows stand for adjectival suffixes. The grey cells highlight several niches in the derivation of French denominal adjectives. The table should be read as follows: nouns ending with -(i)te' potentiate the attachment of *-aire* and the resulting derivatives end with -(i)taire.

The following patterns of potentiation can be inferred from the table: $-ie \rightarrow -ique$, $-iste \rightarrow -istique$, $-ion \rightarrow -ionnel$, $-ion \rightarrow -ionnaire$, $-ment \rightarrow -mentaire$, $-ment \rightarrow -mentai$, $-Ance^{6} \rightarrow -Antiel$, $-eur \rightarrow -orial$, $-ite \rightarrow -itaire$, illustrated by the examples in (7).

⁶We borrow the representation of *-Ance* with the capital *-A* to Dal & Namer (2010). It stands for the two orthographs *-ence* and *-ance* which are pronounced in the same way $/\tilde{as}/.$

	-ie	-eur	-iste	-(i)té	-Ance	-ion	-ment	-ule	-ite
-aire	8	0	0	64	1	65	25	64	9
-al	15	37	0	3	3	10	16	0	2
-el	8	7	0	0	54	224	5	0	1
-esque	0	0	0	0	0	0	0	0	0
-eux	3	2	1	1	1	6	0	0	3
-ien	17	0	0	1	0	0	0	0	0
-ier	0	0	0	0	3	0	1	0	0
-ique	1039	0	99	2	1	7	0	0	38
- <i>u</i>	0	0	0	0	0	0	0	1	0

Table 5: Morphological Niches of French Adjectives

- (7) a. ANALOGIE 'analogy' \rightarrow ANALOGIQUE 'analogical'
 - b. PIANISTE 'pianist' \rightarrow PIANISTIQUE 'pianistic'
 - c. OPTION 'option' \rightarrow OPTIONNEL 'optional'
 - d. ÉVOLUTION 'evolution' \rightarrow ÉVOLUTIONNAIRE 'evolutionary'
 - e. RÈGLEMENT 'regulation' \rightarrow RÉGLEMENTAIRE 'regulatory'
 - f. GOUVERNEMENT 'gouvernment' \rightarrow GOUVERNEMENTAL 'gouvernmental'
 - g. RÉSIDENCE 'residence' \rightarrow RÉSIDENTIEL 'residential'
 - h. DICTATEUR 'dictator' \rightarrow DICTATORIAL 'dictatorial'
 - i. PUBLICITÉ 'advertisement' \rightarrow PUBLICITAIRE 'advertising'

All these examples show that the phonological make-up of a noun is a partial predictor of the suffix used to derive an adjective from that noun. However, the remaining question is whether the attachment of a suffix is ruled only by the phonological shape or whether it also depends on some morphological or lexical information. If we look at two nouns finishing with *-ion*, like those in (8), they do not take the same adjectival suffix. The attachment of *-el* to *-ion* applies to nouns which have the deverbal suffix *-ion* (8a). Otherwise, another suffix is preferred such as *-ique* in (8b). This permits us to claim that these niches are of a morphological rather than phonological nature.

(8) a. CITER 'cite' → CITATION 'citation' → CITATIONNEL 'citation'
b. ANION 'anion' → ANIONIQUE 'anionic'

The last two columns in Table 5 represent final sequences which are suffixoids rather than true suffixes but which nevertheless function as attractors: $-ule \rightarrow$ *-ulaire*, *-ite* \rightarrow *-itique*, exemplified in (9).

(9) a. CELLULE 'cell' \rightarrow CELLULAIRE 'cellular'

b. SATELLITE 'satellite' \rightarrow SATELLITAIRE 'satellite, A'

We can now turn back to morphological types of base nouns in denominal adjectives introduced in Table 3 and look at the proportion of each type for each suffix, presented in Table 6. Remarkable combinations are highlighted in grey. The suffixes *-esque*, *-eux*, *-ier* and *-u* attach mostly to simple nouns, which corresponds to the fact that for these suffixations, we haven't identified any morphological niches. The suffix *-ique* appears with simple nouns and especially with neoclassical compounds, in 58%, which represents *-ie* \rightarrow *-ique* pattern. The 5% of denominal suffixed base nouns for *-ique* adjectives all instantiate the *-iste* \rightarrow *-istique* pattern. Concerning adjectives suffixed with *-aire*, 10% of them have a deverbal base noun and 10% have a deadjectival base noun, representing the niches *-ion* \rightarrow *-ionnaire* and *-ment* \rightarrow *-mentaire* for deverbal nouns and *-ité* \rightarrow *-itaire* for deadjectival nouns. Finally, the suffixe *-el* attaches to a deverbal noun in 52% of the cases. This high proportion is due to the *-ion* \rightarrow *-ionnel* pattern.

Morph.Type %	-aire	-al	-el	-esque	-eux	-ien	-ier	-ique	- <i>u</i>
Simple N	67	72	36	84	89	65	83	35	89
Compound N									
Neo-classical	8	14	2	5	2	31	1	58	0
Native	-	0	0	1	0	0	0	0	0
Prefixed N	2	2	3	1	1	0	0	1	3
Suffixed N									
Denominal N	4	3	1	6	4	2	8	5	5
Deadjectival N	10	1	5	0	2	0	0	1	0
Deverbal N	10	7	52	0	2	0	6	0	3
Other	0	1	0	3	0	2	1	0	0

Table 6: Distribution of morphological types across suffixes

We conclude that the relation between an adjective and its base noun is not formally independent of previous morphological processes and the derivation history is reflected in the coining of new derived lexemes.

4 Multiple Derivation

In this section, we examine what happens from a semantic point view when the base of a denominal adjective is a suffixed noun. Three situations appear and they are illustrated in (10) in the form of derivational chains.

- (10) a. $A \rightarrow N \rightarrow A$: suffixed deadjectival noun base
 - b. $V \rightarrow N \rightarrow A$: suffixed deverbal noun base
 - c. $N \rightarrow N \rightarrow A$: suffixed denominal noun base

We argue that in contemporary French denominal adjectives are semantically built on their base nouns according to their semantic type. For (10a) and (10b), the semantics of the adjective is built on that of the noun, which is immediate base, but not directly on the semantics of the ultimate base, be it verbal or adjectival. On the other hand, for (10c), adjectives construct their semantics on that of one or the other noun in their derivation history. Thus, previous morphological processes are reflected not only in the internal structure of the noun but also in its semantic type. This fact has consequences for subsequent derivation.

4.1 Suffixed Deadjectival Nouns

For deadjectival nouns, two patterns are noteworthy: $-it\dot{e} \rightarrow -itaire$ (11) and $-Ance \rightarrow -Antiel$ (12).

We can evaluate the potentiation effect of these patterns by comparing two proportions. Let us for concreteness focus first on the case of *-itaire*. First we divide the number of derived adjectives suffixed with *-itaire* by the number of nouns finishing in *-ité*. This proportion estimates the likelihood for a noun in *-ité* to have a corresponding derived adjective in *-aire*. Second we divide the number of all adjectives suffixed with *-aire* by the number of all nouns. This second proportion estimates the likelihood for any noun to have a corresponding derived adjective in *-aire*. If the two proportions are roughly equal, there is no potentiation effect. A first proportion that is notably higher than the second one indicates potentiation: there are more adjectives in *-itaire* than one would expect. A first proportion that is notably lower than the second one would correspond to the opposite situation of *preclusion* of a suffix. We used Lexique3 as the source of the data.

We present the results of this evaluation in (13) for *-itaire* and in (14) for *-Antiel* which confirm the potentiation of *-ité* \rightarrow *-itaire* and *-Ance* \rightarrow *-Antiel* patterns.

- (11) $\acute{E}GAL \rightarrow \acute{E}GALIT\acute{E} \rightarrow \acute{E}GALITAIRE$ 'equal' 'equality' 'egalitarian'
- (12) CONFIDENT \rightarrow CONFIDENCE \rightarrow CONFIDENTIEL 'confident' 'confidence' 'confidential'
- (13) $\frac{\text{Aitaire}}{N\text{ite}} = 0,112 > \frac{\text{Aaire}}{N} = 0,0201$
- (14) $\frac{Aantiel}{Nance} = 0,1703 > \frac{Ael}{N} = 0,0139$

Nouns with the suffix *-Ance* are specific in the sense that they have either an adjectival (12) or a verbal base (15). As a result, it is often impossible to decide what the base is, when the morphological family contains both the verb and the corresponding adjective (Dal & Namer, 2010), as exemplified in (16).

- (15) $PRÉFÉRER \rightarrow PRÉFÉRENCE \rightarrow PRÉFÉRENTIEL$ 'prefer' 'preference' 'preferential'
- (16) ABONDER / ABONDANT \rightarrow ABONDANCE \rightarrow ABONDANTIEL 'abound' / 'abundant' 'abundance' 'abundance_{AZR}'

Therefore, we focus on the first pattern: $-it\acute{e} \rightarrow -itaire$. Only recent formations exhibit this pattern: all *-itaire* lexemes date from the 19th or 20th century, whereas many other *-aire* adjectives are inherited from Latin.

It may seem unusual to derive an adjective from a noun which had been derived from an adjective, as this amounts to creating a kind of cycle $A \rightarrow N \rightarrow A$. For example, Roché (2009a) talks about 'paradoxical formations' when an adjective is built on a noun denoting a property, which is the case of most deadjectival nouns. According to him, it is only possible either if the first adjective in the chain has lost its original meaning (17a) or has never existed in French (17b). This semantic paradox could explain why other types of property nouns do not appear as bases of French adjectives. For example, there is no adjective derived from nouns in *-eur* as exemplified in (18)⁷.

- (17) a. VAIN 'vain' → VANITÉ 'vanity' → VANITEUX 'conceited'
 b. [lat. *piger* >] PARESSE 'laziness' → PARESSEUX 'lazy'
- (18) BLANC \rightarrow BLANCHEUR \rightarrow *BLANCHEUREUX, *BLANCHORIAL 'white' 'whiteness'

However, out of 64 *-ité* nouns with a corresponding *-itaire* adjective in our data set, 44 have a transparent adjectival base, as illustrated in (19)-(29).

We argue that this cyclic derivation is not a problem because *-itaire* adjectives built on *-ité* have the semantics of relational adjectives. Thus, the meaning of the derived *-itaire* adjective is distinct from the meaning of the first adjective in the chain. These adjectives are used as modifiers rather than in predicative position and their meaning is often restricted to a special domain or to a particular discourse situation.

⁷The reason cannot be phonological here as there are adjectives finishing with *-eureux*, such as MALHEUREUX, PEUREUX, CHALEUREUX, VAPOREUX.

The adjective in (20) does not relate to the basic reading of the base noun 'the property of being equal', but rather to the reading 'the doctrine of the equality of mankind and the desirability of political, social, and economic equality'. The same observations can be made for the less well established adjective MODERNI-TAIRE in (21), which means 'relative to the doctrine of modernity', not 'relative to the property of being modern'.

- (20) a. $\acute{E}GAL \rightarrow \acute{E}GALIT\acute{E} \rightarrow \acute{E}GALITAIRE$ 'equal' 'equality' 'egalitarian'
 - b. le mariage égalitaire 'egalitarian marriage'
- (21) a. MODERNE \rightarrow MODERNITÉ \rightarrow MODERNITAIRE 'modern' 'modernity'
 - b. une vision modernitaire 'vision of modernity'
 - c. Pour l'heure, la pression modernitaire conduit la direction du PCF, qui valorise désormais les terrains sociétaux, à abandonner notamment trois concepts...
 'For the moment, the pressure of modernity leads the direction of PCF, which promotes from now societal fields, to abandon 3 concepts...' http://www.decitre.fr/livres/bastille-republique-nation -9782841861279.html

These *-itaire* adjectives have a regular semantic relationship $N \rightarrow A$ 'relative to N', where the N is viewed as a topic for debate in society in a given time, which applies also to the examples (22) and (23) which are commonly found in news reports.

- (22) a. IDENTIQUE \rightarrow IDENTITÉ \rightarrow IDENTITAIRE 'identical' 'identity'
 - b. une crise identitaire 'identity crisis'
- (23) a. AUSTÈRE \rightarrow AUSTÉRITÉ \rightarrow AUSTÉRITAIRE 'austere' 'austerity'
 - b. politique austéritaire 'austerity policy'
 - c. Europe austéritaire 'austerity Europe'

Example (24) illustrates another type of *-itaire* adjectives. Most *-ité* nouns denote the property corresponding to the adjectival base such as the example in (25). However, as noted by Koehl (2009) and by Roché (2009b), some *-ité* nouns can express proportional relation which is semantically related to a noun and which can be measured, as shown in (26).

(24) a. NAISSANCE ~ NATAL \rightarrow NATALITÉ \rightarrow NATALITAIRE 'birth' 'native' 'birth rate'

- b. la propagande natalitaire 'birth rate propaganda'
- (25) a. MORTEL \rightarrow MORTALITÉ 'mortal' 'property of being mortal'
 - b. la mortalité de l'âme 'the mortality of the soul'
- (26) a. MORT \rightarrow MORTEL \rightarrow MORTALITÉ 'death' 'mortal' 'death rate'
 - b. *la mortalité infantile* 'infant mortality'

We observe that *-itaire* adjectives can be derived from the proportional relation interpretation: in (27a), DENSITÉ is a property noun without any denominal adjective built on it; in (27b) the adjective DENSITAIRE relates to the relationship noun which expresses a measurable proportional relation, as illustrated in (28). In this example, derivational possibilities reflect different semantic interpretations of the base noun.

- (27) a. DENSE 'dense' \rightarrow DENSITÉ 'property of being dense'
 - b. DENSE 'dense' \rightarrow DENSITÉ 'quantity of people or things in a given place' \rightarrow DENSITAIRE
- (28) Eu égard à la détérioration des conditions écologiques et édaphiques aucune **évolution densitaire** des populations de criquet ne pourrait se faire sentir.

'Despite the deterioration of ecological and edaphic conditions, no density evolution of grasshopper population can appear.'

www.clcpro-empres.org/Library/Bulletins_pays/.../2009_FEVRIER.doc

It just goes to show an emergence of a new regularity with *-itaire* adjectives, belonging to administrative, cultural or sociopolitical language, mentioned already by Tournier (2001).

Finally, as is often the case for denominal adjectives, *-itaire* adjectives can be specialized for a particular domain of knowledge or activity, as exemplified in (29) where the adjective derived from the noun RÉCIPROCITÉ is specific to the domain of fishing.

- (29) a. RÉCIPROQUE \rightarrow RÉCIPROCITÉ \rightarrow RÉCIPROCITAIRE 'reciprocal' 'reciprocity'
 - b. Les accords réciprocitaires : Le souhait d'une réciprocité nationale est unanimement exprimé par l'ensemble des pêcheurs.
 'Reciprocal Agreements: the wish of national reciprocity is expressed unanimously by the whole of the fishermen.' http://www.federationpeche.fr/

4.2 Suffixed Deverbal Nouns

The most frequent patterns for subsequent adjectival derivation from deverbal nouns $(V \rightarrow N \rightarrow A)$ are *-ion* \rightarrow *-ionnel* (30) and *-ion* \rightarrow *-ionnaire* (31). Some other patterns, such as *-ure* \rightarrow *-ural* (32) or *-ment* \rightarrow *-mental* (33), occur with fewer examples.

- (30) ORGANISATION 'organization' \rightarrow ORGANISATIONNEL 'organizational'
- (31) RÉACTION 'reaction' \rightarrow RÉACTIONNAIRE 'reactionary'
- (32) PROCÉDURE 'procedure' \rightarrow PROCÉDURAL 'procedural'
- (33) ORNEMENT 'ornament' \rightarrow ORNEMENTAL 'ornamental'

Here we examine the first pattern $-ion \rightarrow -ionnel$, which has the highest number of newly coined lexemes. Only 38% of our *-ionnel* adjectives are listed in the *TLFi*. Among these, some have been borrowed from Latin or, arguably, English. However, a majority are regularly derived from French nouns ending with *-ion*. The estimation of the potentiation effect (see section 4.1) is given in (34). We can see that the likelihood that an adjective in *-ionnel* is derived from a noun in *-ion* is unexpectedly high.

(34)
$$\frac{Aionnel}{Nion} = 0,1254 > \frac{Ael}{N} = 0,0143$$

Considering now the remaining 62% of adjectives ending with *-ionnel* not found in the *Tlfi*, they mostly have a regular relational meaning, as shown by the examples in (35).

- (35) a. *bilan motivationnel* 'motivational balance'
 - b. comportement organisationnel 'organizational behavior'
 - c. système prostitutionnel 'prostitution system'

We observe that the adjective is not semantically linked to the eventive meaning of the verb. Rather, it has a regular meaning 'relative to N'. This is exemplified in (36), where the noun ASSOCIATION is clearly felt as analogous to the non deverbal nouns PROFESSION 'profession', SYNDICAT 'union' and CLUB 'club'.

- (36) a. ASSOCIER \rightarrow ASSOCIATION \rightarrow ASSOCIATIONNEL 'associate' 'association' 'associational'
 - b. Du point de vue professionnel, syndical, associationnel, clubiste, etc., le membre participant ne peut enfreindre les règlement du groupe sans compromettre sa liberté d'en faire partie.
 'From the professional, union, associational, club point of view, the participant cannot break the rules of the group without compromising their freedom to be a member of it.'

 $\label{eq:logspot} $$ tr/2008_05_25 $$ exclusion-clouseau.blogspot.fr/2008_05_25 $$ exclusion-clouseau.blogsp$

These adjectives are often restricted to a specialized field or discipline, as in example (37) where the adjective is coined with a specific use in mathematics.

 (37) a. VARIER → VARIATION → VARIATIONNEL 'vary' 'variation' 'variational'
 b. calcul variationnel 'variational calculus'

Two remarks can be made about the examples above and the system organization in French. On the one hand, we observe that *-ionnel* adjectives are in competition with prepositional phrases (10,500 hits for (35c) versus 33,300 for (38a), data picked up on Google.fr on November 1st 2013). On the other hand, some of the adjectives have a corresponding adjective ending with another suffix, especially *-if* or *-oire*, such as (38b). These adjectives can derive their meaning to that of a verb or a noun ending with *-ion*, and are then less transparent than regular denominal adjectives.

- (38) a. système de prostitution 'prostitution system'
 - b. affichage associatif 'association display'

The data we deal with shows that speakers can coin new adjectives from deverbal nouns following the frequent patterns. However, some other strategy can develop in the system. Roché (2009a) handles some gaps with a 'principle of economy': no adjective is derived from the noun in (39a) insofar as the system uses a lexeme which already exists derived from another related lexeme, (39b), instead of coining a new lexeme⁸. If this principle were strictly followed, it would rule out examples such as (36).

- (39) a. ÉLECTION 'election' \rightarrow °ÉLECTIONNEL, °ÉLECTIONNIQUE, °ÉLECTIONNAIRE
 - b. ÉLECTEUR 'voter' \rightarrow ÉLECTORAL 'electoral'

Still, some types of deverbal nouns never occur as bases for adjectives, such as nouns suffixed with *-aison* and *-age* or nouns converted from verbs as we have seen in (1). In these cases, using a prepositional phrase seems to be the only possibility.

⁸The only other adjective in the morphological family is ÉLECTIF 'elective' which was borrowed from Latin and is motivated also with respect to the verb ÉLIRE 'to elect'. The meaning of this adjective is not simply 'relative to an election', but rather 'who elects' or 'based on an election'.

4.3 Suffixed Denominal Nouns

In this section, we argue that adjectives derived from suffixed denominal nouns can derive their semantics not only from that of their immediate base noun, but also from the semantics of the first noun in the derivational chain $N \rightarrow N \rightarrow A$. This situation is illustrated in (40). Here we reach the limits of the linear representation of derivation which does not link directly SÉNAT and SÉNATEUR.

(40)	SÉNAT	\rightarrow	SÉNATEUR	\rightarrow	SÉNATORIAL
	'senate'		'a member of		'relative to a senator,
			a senate'		or a senate'

The *-eur* \rightarrow *-orial* pattern is restricted to *-eur* nouns which denote human agents. This goes together with the fact that there are no adjectives in *-orial* derived from property nouns in *-eur* which are derived from adjectives, as it was shown in (18). It confirms also our hypothesis about the morphological nature of the niches.

Another pattern is represented by *-iste* \rightarrow *-istique*. Roché (2009b) analyzed *-istique* adjectives as derived directly from a noun by adding *-istique*. This analysis cannot account for the fact that *-istique* adjectives can also construct their semantics on that of an *-iste* noun and therefore they can have two (41) or even more meanings (42).

UITARISTE \rightarrow guitarist'	GUITARISTIQUE 'relative to guitar,	to guitarist'
URNALISTE \rightarrow 'who writes newspaper'	JOURNALISME \rightarrow 'the domain of newspaper'	JOURNALISTIQUE 'relative to newspapers, journalists, iournalism'
	UITARISTE \rightarrow guitarist' URNALISTE \rightarrow 'who writes newspaper'	UITARISTE \rightarrow GUITARISTIQUE guitarist' 'relative to guitar, URNALISTE \rightarrow JOURNALISME \rightarrow 'who writes 'the domain of newspaper' newspaper'

For the examples in (40), (41) and (42), a binary relationship is not sufficient to deal with the facts, as argued for other cases by Bochner (1993) who introduced cumulative-pattern formations or by Namer (2013) who proposed ternary relations for similar data in French.

Example (43) illustrates a cumulative pattern à la Bochner. This pattern should be interpreted as stating the existence of mutual implicative relations between the three items. Pattern (43) allows not only for $X \sim Xiste$ and $Xiste \sim Xistique$ relation, but it allows also for $X \sim Xistique$ relation. Thus it makes it possible to have an adjective in *-istique* without having any corresponding intermediate noun in *-iste* like in (44).

Because it amounts to dropping the presupposition of a unique base for a derived lexeme, a cumulative pattern allows one to account for ternary morphological relations without stipulating cross-lexemic borrowings of stems (Namer, 2013, Roché, 2011). This representation is based on surface forms and does not need to stipulate a borrowed theme /senator/ from SÉNATEUR to deal with the relation between SÉNAT and SÉNATORIAL.

$$(43) \left\{ X, Xiste, Xistique \right\}$$

(44) football \rightarrow °footballiste \rightarrow footballistique 'football' 'relative to football'

Semantically, there is no change in denotation type (all the nouns denote individuals), hence there is no problem for the interpretation of the adjective. What seems to be important is the semantic relation between the nouns. For (45), it is a profession and an institution, for (46), it is a profession and a domain of activity.

- (45) a. élection sénatoriale 'election to the Senate' b. siège sénatorial 'senator's seat'
- (46) a. carrière journalistique 'journalistic career'
 - b. activité journalistique 'journalistic activity'
 - c. industrie journalistique 'newspaper industry'

This multiple relationship is crucial to the understanding of the competition of denominal adjectives with syntactic means of modification in terms of prepositional phrases: the adjective SÉNATORIAL is in competition with both SÉNATEUR and SÉNAT.

Example (47) illustrates a slightly different situation: the lexemes are not in a derivational chain, but they are still part of the same morphological family. As shown in (48), the two meanings of the adjective can be used with the same head noun.

(47)	ÉLECTEUR 'elector' ÉLECTION 'election'	$\rightarrow \rightarrow$	ÉLECTORAL 'electoral'

(48) carte électorale 'voter card' / 'electoral map'

(49) Une carte d'électeur ou une carte électorale est une carte utilisée dans le cadre d'un vote ou de votation. 'A'

```
http://fr.wikipedia.org/wiki/Carte_d'\%C3\%A9lecteur
\ archive.html
```

As a result, denominal adjectives built on denominal nouns may have multiple meanings relating to distinct members of the morphological family, modulo some semantic and morphophonological conditions.

5 Conclusion

In this paper, we studied the impact of the derivation history on the possibility of adjectival derivation.

We observed some effects of the existing lexicon on the coining of new lexemes in terms of potentiation of affixes which create a strong link between the base noun and the derived adjective. Thus the attractiveness of some suffixal combinations, possibly together with the semantic type of the base, plays an essential role in multiple derivation.

On the one hand, morphological regularity emerges following a discontinuity in the meaning as we could see with denominal adjectives which are relating to their immediate base noun which is not a property noun. On the other hand, adjectives built on suffixed denominal nouns can have multiple meanings relating to distinct nouns in the derivational chain.

To sum up, the derivation history impacts on the possibility of adjectival derivation insofar as it constrains morphophonological and semantic properties of the bases.

References

- Bochner, H. 1993. *Simplicity in generative morphology*. Berlin; New York: Mouton de Gruyter.
- Dal, G., F. Namer. 2010. "Les noms en -ance/-ence du français: quel(s) patron(s) constructionnel(s)?" In Actes en ligne du 2e Congrès Mondial de Linguistique Française, 893–907.
- Fradin, B. 2007. "Three puzzles about denominal adjectives in *-eux*". Acta Linguistica Hungarica 54(1): 3–32.
- Fradin, B. 2008. "Les adjectifs relationnels et la morphologie". In Fradin, B. (ed.), *La raison morphologique Hommage à la mémoire de Danielle Corbin*, Amsterdam/Philadelphia: John Benjamins, 69–92.
- Koehl, A. 2009. "Are French -ité Suffixed Nouns Property Nouns?" In Montermini, F., G. Boyé, J. Tseng (eds.), Selected Proceedings of the 6th Décembrettes, Somerville, MA: Cascadilla Proceedings Project, 95–110.
- Lignon, S. 2000. *La suffixation en*-ien. *Aspects sémantiques et phonologiques*. Ph.D. thesis, Université de Toulouse le Mirail, Toulouse.
- Lindsay, M., M. Aronoff. 2013. "Natural Selection in Self-Organizing Morphological Systems". In Hathout, N., F. Montermini, J. Tseng (eds.), *Morphol-*

ogy in Toulouse. Selected Proceedings of Décembrettes 7, München: Lincom Europa, 133–153.

- Mélis-Puchulu, A. 1991. "Les adjectifs dénominaux: des adjectifs de relation". *Lexique* 10: 33-60.
- Namer, F. 2007. "Composition néoclassique: est-on dans l'hétéromorphosémie"?" In Hathout, N., F. Montermini (eds.), Morphologie à Toulouse. «Actes du colloque international de morphologie 4e Décembrettes», LINCOM Studies in Theoretical Linguistics 37, Munich: LINCOM Europa, 187–206.
- Namer, F. 2009. *Morphologie, Lexique et Traitement Automatique des Langues*. TIC et Sciences cognitives. London: Hermès Science Publishing.
- Namer, F. 2013. "Adjectival Bases of French -aliser and -ariser Verbs: Syncretism or Under-specification?" In Hathout, N., F. Montermini, J. Tseng (eds.), Morphology in Toulouse. Selected Proceedings of Décembrettes 7, München: Lincom Europa, 185–210.
- New, B. 2006. "Lexique 3: Une nouvelle base de données lexicales". In Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN), 892–900.
- Plénat, M. 1997. "Analyse morpho-phonologique d'un corpus d'adjectifs dérivés en *-esque*". *Journal of French language studies* 7: 163–180.
- Plénat, M., M. Roché. 2003. "Prosodic constraints on suffixation in French". In G. Booij, A. R., J. DeCesaris, G. Scalise (eds.), *Topics in Morphology. Selected Papers from the Third Meditteranian Meeting*, Barcelona, IULA-Universitat Pompeu Fabra, 285–289.
- Roché, M. 2004. "Mot construit ? Mot non construit ? Quelques réflexions à partir des dérivés en *-ier(e)*". *Verbum* 26(4): 459–480.
- Roché, M. 2006. "Comment les adjectifs sont sémantiquement construits". *Cahiers de grammaire* 30: 373–387.
- Roché, M. 2009a. "Pour une morphologie lexicale". Mémoires de la Sociétés de Linguistique de Paris XVII: 65–87.
- Roché, M. 2009b. "Un ou deux suffixes? Une ou deux suffixations". In Fradin,B., F. Kerleroux, M. Plénat (eds.), *Aperçus de morphologie du français*,Saint-Denis: Presses Universitaires de Vincennes, 143–173.
- Roché, M. 2011. "Quel traitement unifié pour les dérivations en -isme et en iste?" In Roché, M., G. Boyé, N. Hathout, S. Lignon, M. Plénat (eds.), Des unités morphologiques au lexique, Lavoisier, 69–144.
- Romary, L., S. Salmon-Alt, G. Francopoulo. 2004. "Standards going concrete : from LMF to Morphalou". In Zock, M. (ed.), Actes du Workshop on Electronic Dictionaries de Coling 2004, Genève, Suisse, 22–28.
- Sagot, B. 2010. "The Lefff, a freely available, accurate and large-coverage lex-

icon for French". In *Proceedings of the 7th Language Resource and Evaluation Conference*, La Valette, Malte, 2744–2751.

- Strnadová, J. to appear. "Les adjectifs dénominaux du français problèmes de base(s)". In Lemaréchal, A., P. Koch, P. Swiggers (eds.), Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013).
- Strnadová, J., B. Sagot. 2011. "Construction d'un lexique des adjectifs dé nominaux". In Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN), volume 2, 69–74.
- Tournier, M. 2001. "*Humanitaire* est-il apolitique de naissance?" *Mots* 65: 136–145.
- Tribout, D. 2010. *Les conversions de nom à verbe et de verbe à nom en français*. Ph.D. thesis, Université Paris 7.
- Williams, E. 1981. "On the Notions "Lexically Related" and "Head of a Word"". *Linguistic Inquiry* 12(2): 245–274.

EPISTEMOLOGICAL TENSIONS BETWEEN LINGUISTIC DESCRIPTION AND ORDINARY SPEAKERS' INTUITIVE KNOWLEDGE: EXAMPLES FROM FRENCH VERB MORPHOLOGY

Christian Surcouf

École de Français Langue Étrangère, Faculté de Lettres, University of Lausanne, Switzerland

Abstract

In this article, I address epistemological questions regarding the status of linguistic rules and the pervasive-though seldom discussed-tension that arises between theory-driven object perception by linguists on the one hand, and ordinary speakers' metalinguistic intuition on the other. Several issues will be discussed using examples from French verb morphology, based on the 6500 verbs from *Le Petit Robert* dictionary (2013).

1. Introduction

A journalist commenting on French actress Juliette Binoche's performance declared on the radio "elle est insupportable, elle ne joue pas elle *binoche*" (she is unbearable, she does not act, she "*binoches*"). Undoubtedly, any French native speaker can spontaneously produce the whole morphological paradigm of this brand new verb, and for instance add /ʁa/ to this Pr3¹ /binoʃ/ in order to form Fut2-3 /binoʃ/ʁa/. But what is the status of this "rule"?

In this article, I will mainly raise epistemological questions regarding the tension between scientific expectations while analyzing French verb inflectional morphological rules on the one hand and ordinary speakers' possible inflectional production rules on the other.

¹ Tenses are abbreviated as Pr(esent), Imp(erfect), Fut(ure), Inf(initive), P(assé) S(imple), P(ast) P(articiple). Persons follow the conventional I to *they* order from 1 to 6. Thus, Pr1-3 indicates Present singular.

2. Some epistemological issues about explanations and descriptions

2.1 Linguistics as a science: a brief overview of some of the constraints

As scientists, linguists aim at proposing *coherent*, *exhaustive*, *predictive* and ideally *thorough*, *clear* and *simple*² theories³ (Allan 2003; Lerot 1993: 22-23), while also mostly implicitly "put[ting] a high value on elegance and generality" (Wolpert 1993: 18) (see also Guillaume 1973: 84). According to Soutet (1995: 190) (see also Martin 2002: 68-69; Thagard 2008: 471-475), to be coherent, a theory may "not include two contradictory statements," if it does, then:

L'exigence – difficile à satisfaire – de cohérence conduit le linguiste à se confronter au couple de la règle et de l'exception. [...] De deux choses l'une alors: ou bien on estime que ce conflit est dans l'objet lui-même, ce qui revient à considérer que, dans le domaine considéré, coexistent des propriétés contradictoires; ou bien on postule l'objet homogène et l'on est alors conduit à considérer que la contradiction résulte d'une perception fautive ou, à tout le moins, lacunaire de l'objet. (Soutet 1995: 191)

Underlying Soutet's stance toward the object lays a strong implicit postulate, namely that language should be regarded as homogeneous (for a discussion about language homogeneity from two different perspectives, see Croft 2000: 90ff; Milner 1989: 639ff), as clearly stated by Saussure (de) (1916/1959: 15) "Whereas speech is heterogeneous, language [...] is homogeneous." Is language intrinsically homogeneous? Or does the very idea of homogeneity result from the linguists' endeavor to provide coherent and exhaustive descriptions⁴, hence highlighting the rules that work while downplaying the possible epistemological significance of exceptions? Do rules reflect actual regularities in the language?

 $^{^2}$ As Hurford (1977: 574) puts it "Science seeks to discover as much lawfulness as possible in the universe–but, paradoxically, to formulate as few laws as possible, since its lawmaking propensities are strictly curbed by the requirement that theories be maximally simple."

³ By *theory*, I mean: a provisional intellectual model of a humanly perceivable part of the universe, consisting of interrelated and partly conjectural propositions seeking to accurately describe, explain, and predict observed regularities in the part of the universe under investigation.

⁴ Biases might also arise from data collection "since the documentation does not repeat the documented reality itself, but only represents a sample of it, there is necessarily a process of selection, which in itself is not objective and which, in fact, can be highly tendentious." (Lehmann 2001: 87-88).

2.2 Do rules always exist?

'Well, I don't go all the way with the neuroscientists. OK, the mind is a machine, but a *virtual* machine. A system of systems.'

'Perhaps it isn't a system at all.'

'Oh, but it is. Everything in the universe is. If you are a scientist you have to start from that assumption'.

Lodge, David (2001), Thinks

If, as mentioned in the introduction, a French native speaker can effortlessly produce all the forms belonging to the verb paradigm of the absolute neologism /binoj/, some kind of "rule" leading to these new forms has to exist (see Morin 1987: 14). But what is a "rule"? According to Kiefer (2000: 297), "a grammatical rule is any statement expressing a linguistically significant generalization about the grammatical facts of a particular language" (see also Fradin 2003: 306). As the term "statement" clearly suggests, such "a given generalization [...] can only acquire significance in relation to a particular linguistic theory" (Berg 1998: 2). The problem is then to determine whether rules are mere "statements about actual linguistic behavior" (Trask 1999/2007: 248) proposed by linguists and inherently bound to specific-and often incompatible-theoretical models, or actually reflect a real "linguistic behavior", which ordinary speakers rely on to produce grammatical sentences. To what extent do rules actually organize language? Let us consider the tension between the theory ("meta-level") and the object ("object-level"⁵).

As a first approximation, I will consider the four possible configurations of Table 1, making the linguists' endeavor to formulate rules either a dead end or a possible achievement. *Intrinsic rules* (IR) represent the rules assumed to exist in the language itself–no matter whether linguists can find them or not–, and *proposed rules* (PR) the ones formulated by linguists. The following table presents the four possibilities of existing (1) or non-existing (0) rules on either side:

⁵ These two terms are borrowed from Lehmann (2001: 89, table 2). The word "object" is by itself already theory-dependent, since "it is the viewpoint that creates the object" (Saussure (de) 1916/1959: 8).



Table 1: Theoretical approaches to rules

Scientists would normally discard I and IV, since their responsibility and duty–which justifies their social status–is precisely to discover regularities and formulate rules. In I, denying the very existence of intrinsic rules in language would just make this scientific quest irrelevant (however, see 2.4). Moreover, as native speakers of at least one language, linguists–no matter their epistemological stance toward their object–have to spontaneously acknowledge that some "device" allows languages to be learned, and previously unknown grammatical forms to be accurately produced (e.g. Fut2-3 /bin5ʃʁa/, see above). These two reasons suffice to acknowledge the existence of rules, and hence regard IV as a scientific failure.

Now what about II and III? III obviously represents the ideal scientific situation. Nonetheless, there is no guarantee that the rules proposed by the linguist actually reflect the intrinsic rules of the language investigated. As to configuration II, one might want to definitely reject it on the ground that–as argued above–intrinsic rules must exist if language learnability and productivity are to be explained. However, claiming that rules exist in a language is not an all-or-nothing issue. There might be *some* parts of the language that obey rules while *others* do not, although this latter state of affairs does not preclude linguists from positing the existence of rules⁶. Indeed, scientists (see Lodge's epigraph), cannot help assuming that what they investigate works in a systemic way, and generations of linguistics students have been taught that language is a "system of signs⁷" (Saussure (de) 1916/1959: 15) (or a "system of systems" according to Guillaume 1973: 176), whose mechanisms need to be uncovered by linguists. However, does

⁶ Like any scientist, linguists might incorrectly posit the existence of an entity that complies with their theoretical model and justifies their observation (e.g. phlogiston before Lavoisier, ether before Einstein, see Chalmers 1999: 114). This issue is probably more difficult to tackle in linguistics than in natural sciences, since *rules* are at best neurological processes that so far cannot be observed.

⁷ Lass (1980: 89) criticizes the term *system* "used very loosely in describing various aspects of language [...]. The most this would be likely to do is to give us a pseudo-precision, i.e. create a false sense that there are algorithms at the bottom of the garden–when we get there."

language nicely meet the scientific expectations harbored by linguists? Let us consider an example.

2.3 The quest for the unruly rule...

While comparing the sentences "La pendule retarde [lapādylrətard]" (the clock is slow) and "La pendule retardait [lapādylrətardɛ]" (the clock was slow), Dubois (1967: 9) refers to a "system of marks" ("système de marques"), thus allowing the systemic discrimination between Pr3 /ʁətaʁd/ and Imp3 /ʁətaʁdɛ/. Hence, accordingly Dubois (1967: 61) argues that for the three idiosyncratic Pr5 forms from, respectively, be, say and do /ɛt/ (êtes), /dit/ (dites), /fɛt/ (faites), "a specific type of morphophoneme /t/" is added to Pr2 /ɛ/ (es), /di/ (dis), /fɛ/ (fais). From a scientific point of view, this statement undeniably constitutes an accurate description of the data. However, it raises some epistemological questions regarding its status. Does it reflect a cognitive reality in the speakers' brain? If, after Dubois, we grant the above description the status of a systemic rule, for French verbs' Pr5, there would be two "competing" rules, namely the one adding /t/ to Pr2 for être, dire and faire, and the one adding /e/ to Pr2, here represented by laver (wash), and courir (run):



Table 2: Two possible competing rules for Pr5 in French?

Among the 6470^8 verbs from the *Petit Robert* (2013) dictionary, rule 1 applies exclusively to these three verbs, whereas rule 2 concerns more than 5500 verbs (~86%), i.e. the ones with only one Pr radical, such as /lav/ (wash-Pr1-3.6), or /kus/ (run-Pr1-3.6). Could Dubois's description reflect what really takes place in French speakers' brain? Are ordinary speakers aware of the existence of this idiosyncratic morpheme /t/⁹? What could be the processing stages involved to become aware of this rule, and memorize it?

⁸ This corpus of 6470 verbs was collected from the electronic version of the dictionary *Le Petit Robert* (2013). Verbs have been manually organized by types following the classifications proposed by Pouradier Duteil (1997) and Séguin (1986).

⁹ Such a morpheme *theoretically* exists elsewhere in verb paradigms, namely in PS5. However, this tense is no longer used orally and hence makes the morpheme /t/ purely theoretical (see 2.5.4).

Here are some possible steps:

- a) Have access to at least two forms from one of these three verbs;
- b) Hypothesize rule 1; e.g. for $\hat{e}tre$, the observation of $\epsilon/$ for Pr2, and $\epsilon/\epsilon t$ for Pr5 leads to the following rule: Pr5=Pr2+/t/ and Pr2=Pr5-/t/;
- c) Memorize that rule 1 applies to *être*;
- d) Hypothesize the extension of this rule to other verbs;
- e) Realize that rule 1 does not work for thousands of French verbs, and hence has to be inhibited to avoid deviant forms such as /*kuʁt/ (run-Pr5), instead of /kuʁe/;
- f) Memorize the inhibition instruction e);
- g) Have access to the four other forms /di/-/dit/, and $/f\epsilon/-/f\epsilon t/$;
- h) Compare these forms as done in b) with *être*;
- i) Realize that the rules found in b) also apply to these forms;
- j) Memorize that these rules only apply to *dire* and *faire* (i.e. refine e) and f);

Apparently, since steps a) and g) are necessary to establish rule 1 (i.e. access the six forms governed by this rule), direct rote memorization of these forms looks more efficient and straightforward for the speaker¹⁰. Martinet, probably relying on his own native speaker's intuition, challenges Dubois's explanation:

on ne remarque pas que l'analyse /ê-t/, /fe-t/, /di-t/ que pourrait suggérer une comparaison avec les trois singuliers correspondants /il è/, /il fè/, /il di/ corresponde, dans l'usage, à un rapprochement analogique efficace: *vous êtes* est bien ancré chez les sujets qui, par millions, laisseront échapper *vous faisez* et *vous disez*. (Martinet 1974: 99)

Although oversimplified, this short presentation raises questions about the possible tensions between a scientifically coherent explanation and its actual

¹⁰ My point obviously echoes the debate in verb morphology as to the extent of the role of rote memorization versus rule implementation (see Baayen 2007; Bybee 1995; Clahsen 2006; Marcus 2000; Nakisha, Plunkett & Hahn 2000; Pinker 1999: 121ff). Langacker's (1987: 29) argues about this "rule/list fallacy" i.e. "the assumption, on grounds of simplicity, that particular statements (i.e. lists) must be excised from the grammar of a language if general statements (i.e. rules) that subsume them can be established. Given the general N + *-s* noun-pluralizing rule of English, for instance, specific plural forms following that rule (*beads, shoes, toes, walls*) would not be listed in an optimal grammar. [...] this is a specious kind of simplicity for anyone taking seriously the goal of 'psychological reality' in linguistic description. It is gratuitous to assume that mastery of a rule like N + *-s*, and mastery of forms like *beads* that accord with this rule, are mutually exclusive facets of a speaker's knowledge of his language; it is perfectly plausible that the two might sometimes coexist. We do not lose a generalization by including both the rule and specific plural forms in the grammar of English, since the rule itself expresses the generalization".
plausibility from the ordinary speakers' point of view¹¹. Can linguists posit the existence of rules on mere theoretical grounds without ever attempting to assess their actual plausibility in the speakers' brain? Should we endorse the "proponents of linguistic rules [who] do not necessarily view them as psychologically real" (Corrigan & Lima 1994: xv)? Although most of these questions have been addressed by (psycho)linguists (see note 10 for references), there remain some essential epistemological issues that I wish to partially discuss here.

2.4 Is language cloud- or clock-like?

In his Of clouds and clocks lecture presented in 1965, Popper explains his title as follows:

My clouds are intended to represent physical systems which, like gases, are highly irregular, disorderly, and more or less unpredictable. I shall assume that we have before us a schema or arrangement in which a very disturbed or disorderly cloud is placed on the left. On the other extreme of our arrangement, on its right, we may place a very reliable pendulum clock, a precision clock, intended to represent physical systems which are regular, orderly, and highly predictable in their behavior. (Popper 1972: 207)

Although Popper's subject is remote from our linguistic discussion, his metaphor raises an important epistemological issue concerning the nature of the object under investigation, and hence the type of description scientists might come up with¹². If languages were clock-like, then descriptive coherence would be easy to achieve: observed regularities could be formulated as rules by linguists and would reflect all the mechanisms involved in language. However, although language is not as unpredictable as a cloud, there might be some parts of it that *do not* follow rules¹³. Then the question arises as to whether linguists should keep trying to look for them. Let us take an example. While trying to formulate rules to describe French PS and PP within the once acclaimed generative phonology framework, Plénat (1987: 137-138) ends up proposing "around fifteen often very simple

¹¹ One may object that though predictive, Dubois's rule does not have a high enough predictivity rate. This example raises as well the question of the threshold to be reached for a rule to become productive for ordinary speakers: 10 items? 30? 100? (in his "minimal generalization learner model", this issue of "critical number" is also raised from a somewhat different perspective by Albright 2002: 41). ¹² Popper's lecture discusses "the problem of rationality and the freedom of man."

¹³ In his discussion about rules-from a very different perspective from mine-, Fradin (2003: 266) notes : "En morphologie [...] il est rare qu'on couvre l'ensemble des expressions relevant d'un phénomène au moyen d'une règle unique et nettement formulée. [...] Très souvent, la règle décrit une portion des faits, et laisse un résidu plus ou moins important."

rules." Nonetheless, albeit his scrupulous study, Plénat (1987: 138) wonders whether his endeavor was not in its very essence vain since "dans un tel ensemble, il est fatal qu'un linguiste découvre un certain nombre de régularités" (see also Morin 1987: 76). Besides this insightful epistemological self-criticism, Plénat raises what I hold to be a fundamental question about the status of the rule with respect to ordinary speakers (my emphasis):

[le linguiste] a-t-il [...] le droit de supposer que ces régularités sont repérées et apprises comme telles par les locuteurs, d'en faire des règles de la grammaire que chacun intériorise? *Certainement pas*. On pourrait soutenir avec autant de vraisemblance que les verbes irréguliers constituent en synchronie *un chaos de formes disparates* qui doivent être mémorisées [...]. (Plénat 1987: 138)

This "*chaos* of disparate forms that must to be memorized" would belong to the cloud-like part of language and impede the linguists' quest for rules, since, as Plénat argues, speakers "certainly" do not implement his proposed rules to produce irregular PP or PS forms, but–as advocated earlier in the case of Pr5 for *dire*, *faire* and *être*–just learn the whole forms by rote. This epistemological stance is summarized in the figure below.



Plénat's mentioning of speakers¹⁵ is of crucial importance here. Indeed, as a first approach, language could be regarded as any scientific object, and

¹⁴ Plénat (1987: 139) modestly acknowledges: "Les règles découvertes n'ont jamais l'occasion de s'appliquer au-delà du corpus qui a servi à les établir". This remark is particularly appropriate concerning PS, since most French native speakers ignore the standard forms from 2nd and 3rd groups provided by grammar books (examples of common deviant forms are given in Kilani-Schoch & Dressler 2005: 199ff).

¹⁵ This obviously depends on the epistemological status given to the role of the theory with respect to language. For instance, Chomsky's (1969: 25) highly influential position holds that "a child who has learned a language has developed an *internal representation of a system of rules*", and that the "long-range task for general linguistics [is to] set the problem of developing an account of this innate linguistic

treated by linguists, as astronomers would a planet, biologists a virus, etc. For instance, by bypassing the cardinal role speakers play in the existence of the language data. Chomsky somehow illustrates such a position:

A grammar of the language L is essentially a theory of L. Any scientific theory is based on a finite number of observations, and it seeks to relate the observed phenomena and to predict new phenomena by constructing general laws in terms of hypothetical constructs such as (in physics, for example) 'mass' and 'electron'.¹⁶ Chomsky (1957/2002: 49)

However, in linguistics such a "hard science" approach faces at least the following problems: 1) there is no such thing as "*a* language" per se (see Langacker 2008: 215ff)–any such entity is in fact a scientific and/or sociocultural reconstruction by linguists and/or native speakers; 2) native speakers: a) need to acquire their language; b) develop idiosyncratic ways of using language; c) need to use their brain and body in order to be able to speak; d) have access to what they say, but not to how they actually came up with the sentences they uttered; e) speak to other people to communicate; f) end up developing some metalinguistic representation of their own language. My main concern here relates to the status of the rule in relation to speakers once we acknowledge these facts. One of the first examples is alluded to by

Plénat in his conclusion (my emphasis):

Personne, sans doute, n'irait jusqu'à prétendre qu'elles [=les formes irrégulières] sont apprises une par une. Mais *elles pourraient l'être très diversement suivant les locuteurs*, et la description présentée ici n'aurait, sur telle ou telle autre présentation que le mince avantage – si c'en est un –, d'être plus compacte. *Elle ne révèlerait en rien la façon dont un locuteur organise ses connaissances*.

[...] Il n'y a naturellement pas lieu de croire que tous les locuteurs ont intériorisé exactement les mêmes règles de formation [...] (Plénat 1987: 138-139 & 141)

What do we know about the way individual speakers organize their own knowledge? The fact that for *scientific* reasons, scientists need to work inductively on large amounts of data raises epistemological questions regarding the observation of regularities and the resulting formulation of

theory that provides the basis for language learning. (Note that we are [...] using the term 'theory' [...] with a systematic ambiguity, to refer *both to the child's innate predisposition to learn a language* [...] *and to the linguist's account of this.*)" (my emphasis). Such a position allows linguists working within this paradigm to disregard the difference between meta-level and object-level, and ignore the speakers' point of view since "a generative grammar is not a model for a speaker or a hearer" (Chomsky 1969: 9).

¹⁶ Elsewhere Chomsky (1957/2002: 48) uses a revealing analogy concerning his position toward the object: "Perhaps the issue can be clarified by an analogy to a part of chemical theory concerned with the structurally possible compounds. This theory might be said to generate all physically possible compounds just as a grammar generates all grammatically 'possible' utterances."

rules¹⁷. Here the difference between say physics and linguistics is fundamental. A physicist studying the way stones fall does not need to consider that (2a) these stones had to learn how to fall, and above all (2b) that they all do it in a "personal" way (bound to a specific temporal and spatial context)¹⁸. In linguistics however, while studying language acquisition, linguists do notice individual differences (see also Kail 2012: 38-39):

One conclusion seems uncontroversial: the Average Child is a fiction, a descriptive convenience like the Average Man or the Average Woman. Theories of language development can no longer rely on this mythical being. Any theory worth the name will have to account for the variations that are reliably observed in early language learning. (Bates, Dale & Thal 1995: 151)¹⁹

One might argue that this is limited to "early language learning" and does not concern adults' speaking ability. But how do we know? How can we be sure that the variety of learning styles at early age fades out in mature daily language practice and never gives rise to different underlying speaking strategies? Besides, while analyzing data, whereas "deviant" forms can easily be detected, and interpreted as echoing underlying "defective" structures, a correct surface form never absolutely guarantees that a "correct" underlying rule²⁰ has been used. Let us examine some of Damourette & Pichon's collected mistakes for PC forms, found in children's, as well as adults' popular language ("le parler du peuple"):

Il est curieux de constater que, d'après l'analogie *vous répondez/répons* [kepõ], *il craint/craint*, le parler du peuple et celui des enfants reforment des sus [=PP] [...] par troncature du radical: *vous cous-ez* [kuze]/cous [ku(z)]; *vous vivez* [vive]/vi [vi]; *vous taisez* [tɛze]/*tais* [tɛ], etc. [I reformatted the phonetic transcription]

(Damourette & Pichon 1911-1930: 134)

There is no need to postulate a truncation from the Pr5 radical–which anyhow does not work for *craindre*, yielding /k $\mathfrak{k}\mathfrak{g}\mathfrak{p}$ / instead of /k $\mathfrak{k}\mathfrak{k}$ / (*craint*) as mentioned by the authors–, children and adult might as well be using Pr1-3 (i.e. as PP) to form their PC (see Morin 1987: 76), hence leading to /ilavi/– from, say, Pr3 /ilvi/ (*il vit*)–instead of /ilaveky/ (*il a vécu*) (an error, among others, also mentioned by Plénat 1987: 139-140)²¹. In cases such as /ilavi/, linguists cannot but identify these forms as errors, and then logically hypothesize a wrong underlying rule leading to these wrong surface forms.

¹⁷ (for other issues raised by inductive reasoning, see e.g. Hempel 1966: 200).

¹⁸ Let alone (2e): to communicate with other stones.

¹⁹ Child language being individually investigated naturally receives closer attention than adult language.

²⁰ I am aware that my use of "correct" here is problematic. We could equate it with "a standard adult grammatical rule hypothesized by linguists and possibly intuited by ordinary adult speakers" (somehow resembling the 'constructive rule' among the three types proposed by Besse 1991).

²¹ Plénat proposes an alternative explanation based on his theoretical model.

However, the child (or the adult), who constructs PC with the rule "use avoir-Pr+Pr1-3" will *also* produce "correct" surface-forms such as /ila**fini**/ (*il a fini*), /ila**f**ɛ/ (*il a fait*), /ila**di**/ (*il a dit*), /ila**ekʁi**/ (*il a écrit*), /ila**kõdųi**/ (*il a conduit*), etc. These "mistakes" (see infra) based on a "wrong underlying rule" will obviously pass unnoticed. In other words, what the grammatical and linguistic tradition regards as the *correct* formation of PC–involving a PP (i.e. a specific form, not to be equated with Pr1-3)–might not reflect what speakers actually do.

To further investigate the epistemological consequence of this issue, let us imagine that French PC is–for the sake of the demonstration–exclusively constructed with the Pr of the *avoir* auxiliary (and never with *être* auxiliary). Then an ordinary speaker (whether a child or an adult) might use the abovementioned rule complemented by Rule 2^{22} :

Rule 1: "For type X verbs, PC=avoir-Pr+Pr1-3": e.g. /ilaa3i/ (*il a agi*), from Pr1-3 /a3i/ (*agis/t*)

Rule 2: "For verbs with /e/-ending Inf, PC=avoir-Pr+Inf": e.g. /ilalave/ (*il a lavé*) from Inf /lave/ (*laver*)

These two rules would cover 96% of the 6500 verbs from the *Petit Robert*, respectively 6.5% and 89.5%. The remainder (250 verbs) would have to be memorized as "idiosyncratic". The deviant forms mentioned above by Damourette & Pichon (as well as the ones cited by Plénat 1987: 139-140) belong to the set of more than 400 verbs, which conform to Rule 1 and have homophonous forms for Pr1-3 and PP.

Even though probably most linguists would reject Rules 1 and 2, the fact that these two "wrong" underlying rules can produce 96% of *correct* surface forms (/ilaa3i/, /ilalave/ sound perfect!) raises serious epistemological questions. One could argue that so-called 1st group verbs are overrepresented compared to their actual share in daily conversation, thus giving a distorted view of the issue. But even within the 200 most frequent verbs (according to Gougenheim, Michéa, Rivenc & Sauvageot 1964), the production of *correct* surface forms with Rules 1 and 2, if implemented, still reaches 71% (6% for R1, 65% for R2, contra 29% for the rest²³). It could be further objected that, on linguistic ground, the semantic or functional nature of Inf does not allow its usage as PP to construct PC according to our speaker's Rule 2. If, again *from an ordinary speaker's point of view*, this difference were clear, then

²² Bassano *et al.*'s (2001: 143) French acquisition data corroborates Rule 2. Accordingly, as Kilani-Schoch (2003: 289) notes, "there should be analogical PP forms based on Inf"; she found "avait mettre" instead of "avait mis" in her corpus.

²³ For the remainder, which can be subdivided into less than twenty similar rime subsets such as Pr6 / $\tilde{a}t\tilde{a}d$ / > PP / $\tilde{a}t\tilde{a}dy$ / (entendu), Pr6 / $\kappa\tilde{a}d$ / > PP / $\kappa\tilde{a}dy$ / (rendu), Bybee's (1995: 428) network model suffices to explain how "morphological structure emerges from the connections [words] make with other words in the lexicon."

why would French grammar books need to give "tricks" such as the one below to avoid the spelling confusion between *fermer* (Inf) and *fermé* (PP)?

Pour distinguer les diverses terminaisons des verbes du 1^{er} groupe, on peut remplacer la forme pour laquelle on hésite par un verbe du 2^{e} ou du 3^{e} groupe; on entend alors la différence.

[...] Nous avons fermé la porte \rightarrow Nous avons ouvert la porte (Bled, Bled & Berlion 2010: 45)

Since such spelling confusions between homophonous PP and Inf persist despite years of explicit and implicit teaching (up to 20% of spelling errors for 15 year-old students, according to Brissaud 2002: 63), it seems legitimate to question whether the semantic or functional difference tacitly advocated by linguists and grammarians actually corresponds to ordinary speakers' use of underlying rules.

Let us synthesize the epistemological issue raised by these imagined "wrong underlying rules" yielding a vast amount of correct surface forms. In the table below, LA stands for "linguistically acceptable", PCSF for "production of correct surface forms" and DE for "detectable errors"; figures are rounded and based on the corpus of verbs from the *Petit Robert* (2013) and the *Français élémentaire* (Gougenheim *et al.* 1964)²⁴.



The problem arises from the tension between the ordinary speaker's perspective and the linguist's. On the one hand, the ordinary speaker uses rules that would be discarded as *wrong* by linguists (since neither Pr1-3 nor Inf *are* PPs), although the forms she produces sound objectively *right* (e.g. R1: /ila**fui**/, *il a fui*; R2: /ila**fate**/, *il a chanté*) in 96% (N=6500) or 71% (N=200) of the data. On the other hand, the linguist posits a scientifically acceptable rule Rx, say PC=avoir-Pr+PP (in which PP meets linguistic criteria acknowledged in the field), and sees that the data complies with Rx, *even though* this is *not* the rule implemented by the ordinary speaker imagined in my example. The matching of the posited scientific rule Rx with the underlying speaker's rules R1 & R2 responsible for a vast amount of the data is thus an illusion, due to identical resulting surface forms (e.g. R1's Pr1-3 surface form /fui/ is identical to Rx's PP /fui/). Hence configuration A

²⁴ Calculations are based on data manually tagged in Excel.

leads to an "incorrect"–albeit unavoidable–interpretation of the data by the linguist. The only correct interpretation would in fact come from B, which, in 4% (N=6500) and 29% (N=200) of the cases, allows the detection of errors in surface forms produced by R1–e.g. /ila**wepõ**/, from Pr1-3 /wepõ/, instead of /ila**wepõdy**/ (*il a répondu*) (listed by Damourette & Pichon 1911-1930: 134). Insofar as linguistics does not have any direct access to speakers' production processes, there seems to be no way to escape this linguistically embarrassing epistemological state of affairs (see however the tentative exploration of neural substrates underlying production processes in Sahin, Pinker, Cash, Schomer & Halgren 2009). In the end, this discussion boils down to the fundamental question (mostly irrelevant in other sciences): what are the possible arguments that would allow us to hypothesize that ordinary speakers think and speak in accordance with a specific scientifically coherent model?

2.5 Advocating an ordinary speaker's perspective in linguistics

In order to clarify the expression "ordinary speaker", let me first quote this highly instructive account from a bilingual native–albeit unique–speaker of French and English, raised in Louisiana, who never learned to write or read French. In this passage, he explains to the interviewer why it is difficult for him to translate a single English word into a single French word (here the translation for *tree*):

- Comment vous dites cold?

- [lœfrɛ]. C'est-à-dire, c'est selon l'histoire [...]. Tu vois, pour un n-exemple, t'as larbre, narbre, arbre ou zarbre: un narbre. Tu vois, t'as [...] des fois t'uses le mot larbre, narbre, arbre ou zarbre. Zarbre veut dire "plus [...] qu'un". En anglais, t'uses un mot. Ça me gêne pas si y en a un ou i n'n a dix, c'est toujours le même mot. Et en français, t'as quatre mots [...] (quoted in Morin 2005: 14)

In his metalinguistic discussion about what linguists would call the French "liaison", the speaker mentions the four surface forms *larbre*, *narbre*, *arbre zarbre*, which, for him represent "four words" ("en français, t'as quatre mots"), among which "zarbre" is the word "meaning more than one" ("plus qu'un"). This explanation is all the more interesting that it is unbiased by writing²⁵, thus providing a genuine metalinguistic account of the speaker's intuitive knowledge. Had he been literate, he would probably never have made such a claim. Though anecdotal, this quote gives us some insight into what a literate ordinary speaker–and not a theory-driven one–would be if not influenced by years of schooling, reading, writing and grammar teaching. Such an account is probably the closest we can get to self-introspection about

²⁵ It is hard to evaluate the impact of writing, reading and schooling on our metalinguistic knowledge. Moreover, introspective and retrospective efforts into personal preschool memories in order to shed light on the issue are doomed to fail.

underlying rules. It should therefore be considered as a possible perspective for linguistic investigation (this is done by Morin 2005).

Although speakers from compulsory schooling societies are probably highly influenced by writing and grammatical tradition in their metalinguistic analyses, it is nonetheless unlikely that core language mechanisms acquired during early childhood change significantly under the influence of schooling. And as the spelling confusions, mentioned earlier, between $-\acute{e}$ and -er in verb endings ("il a *tuer" for "il a tué"; see the extensive study on this issue by Brissaud, Chevrot & Lefrançois 2006) tend to show, it takes years of training to inculcate what might be perceived as counter-intuitive spelling to some speakers who, we could hypothesize, use different underlying rules.

In other words, an "ordinary speaker" is a normal language user who primarily *speaks* the language in her own way. Although trivial, this statement is of prime importance to avoid sacrificing ordinary speakers' point of view for the sake of preserving the coherence of the theoretical model proposed by linguists. This position echoes Saussure's difference between *objective* analysis, "based on history" and carried out by linguists, and *subjective* analysis, that "speakers constantly make of the units of language" (Saussure (de) 1916/1959: 183).

Thus, while dealing with analogy:

The grammarian is prone to think that spontaneous analyses of language are wrong; the truth is that subjective analysis is no more false than "false" analogy. [...] There is no common yardstick for both the analysis of speakers and the analysis of the historian although both use the same procedure: the confrontation of series that have a common element. Both analyses are justifiable, and each retains its value. In the last resort, however, only the speakers' analysis matters, for it is based directly on the facts of language.

(Saussure (de) 1916/1959: 183) (my emphasis)

The importance of ordinary speakers is even more explicitly stated in Saussure's manuscripts found in 1996: "avant de venir nous parler d'abstractions, il faut avoir un critérium fixe touchant ce qu'on peut appeler *réel* en morphologie. *Critérium*: Ce qui est réel, c'est ce dont les sujets parlants ont conscience à un degré quelconque; tout ce dont ils ont conscience et rien que ce dont ils peuvent avoir conscience" (Saussure (de) 2002: 183) (for a thorough argument, see also Komatsu & Wolf 1996: 70ff). This emphasis on speakers' awareness naturally leads to pay more attention to surface forms, i.e. "living units perceived by speakers" (Saussure (de) 1916/1959: 170):

In Modern French, *somnolent* 'sleepy' is analyzed *somnol-ent*, as if it were a present participle. Proof of this is the existence of the verb *somnoler* 'be sleepy.' But in Latin the division was *somno-lentus*, like *succu-lentus*, etc., and before that it was *somn-olentus* 'smelling of sleep,' from *olere*, as in *vin-olentus* 'smelling of wine'. The most obvious and important effect of analogy is thus the substituting

of more regular forms composed of living elements for older irregular and obsolescent forms. (Saussure (de) 1916/1959: 170-171)

Following Saussure's emphasis on the role of "subjective analysis" in language, I will advocate an "intuitive" linguistic analysis²⁶, to remain as close as possible to ordinary speakers' intuition. As linguists, we should endeavor to satisfy at least the following points²⁷:

- a) Refrain from projecting theoretical constraints onto the object
- b) Keep as close as possible to ordinary speakers' expressed metalinguistic intuition
- c) Be consistent with language acquisition data
- d) Be consistent with ordinary language use
- e) Beware of any biased analysis induced by writing
- f) Take existing surface forms as basis for the analysis and avoid postulating abstract underlying entities on theoretical ground
- g) Account for diachronic "residue" in the language
- h) Consider that inflected verbs are the result of a process in time and that an inflected verb is not an isolated item but *always* appears in an utterance

I will now discuss points a) to d) with examples taken from the literature about French morphology²⁸.

2.5.1 Refrain from projecting scientific constraints onto the object

As I have already partly discussed this issue in 2.1, I will only briefly recall two of the main scientific constraints a theoretical model faces when it strives to be *coherent* and *predictive*²⁹. The figure below schematically synthesizes the tension between the projected theory-driven object, resulting from these criteria, and the "real" object (see note 5).

²⁶ Even though Saussure (de) (1916/1959: 183) clearly states that "subjective" refers to speakers (i.e. *subjects*), this term is somehow misleading both in French and English, since it is usually opposed to "objective", which tends to be equated with "true" or "scientifically demonstrated". I will therefore use "intuitive" instead.

 $^{^{27}}$ This list is not meant to be exhaustive. For the time being, it should be considered as a general guideline. Points a) and b) are found in Allan (2003: 552).

 $^{^{28}}$ The original draft of this article dealt with each point, but has been shortened by half for editorial reasons.

²⁹ Obviously I am not claiming that these criteria are irrelevant. Nonetheless, in linguistics, confusion between meta-level and object-level may lead us astray. For instance, the word *rule* may refer to both levels, either the way the language is supposed to work intrinsically or the account that linguists give of such 'mechanisms' (I am thus opposed to Chomsky's deliberate ambiguous use of the word theory. For a discussion, see Langacker 1990).



Figure 2: Epistemological tensions between projected and "real" object

2.5.2. Keep as close as possible to ordinary speakers' expressed metalinguistic intuition

This first point is admittedly impressionistic and obviously leaves a lot of room to personal interpretation but relates to the other points. It is meant as a reminder to avoid treating language as an abstract object that transcends speakers' daily usage and intuition. After all, linguistics itself is a scientific offshoot of speakers' own intuition on language, and it seems paradoxical that some linguists propose models inaccessible to their own native speakers' intuition, despite years of training in the field.

Let me give one example of metalinguistic comment about a spelling confusion between homophonous /e/ endings verbs³⁰. While asked why she wrote "On avait crier tous en même temps" (orthographically: avoir-Imp3+Inf) (instead of orthographic PP crié), a primary school pupil answers: "Crier, er, parce que quand deux verbes se suivent, le deuxième est à l'infinitif" (Brissaud, Cogis, Jaffré, Pellat & Fayol 2011: 238). At this level of schooling, such a statement ("when two verbs follow one another, the second must be an Infinitive") clearly shows a good mastery of metalinguistic terms (verb, infinitive), but the pupil's spontaneous intuitive explanation would be regarded as wrong by linguistic standards, although while speaking nobody would have noticed that her rule "does not work". The question is now: is this pupil's intuition right or wrong? From a scientific point of view, her explanation is wrong. From an ordinary speaker's point of view her intuition is right. It yields correct surface forms with /e/-ending type verbs (such as *crier*) and allows to successfully produce an infinite number of correct sentences ("Il avait /dãse/"; "Elles auront /3we/"; "J'ai /mã3e/", etc.). Then are linguists entitled to claim that they are right against speakers' intuitions? I would opt for a negative answer. As mentioned earlier, linguistics as a science strives for coherence and exhaustiveness, and on this

³⁰ I have selected this comment to serve my epistemological discussion. I have no specific position regarding the spelling issue itself, and to what extent it actually reveals underlying rules used by speakers.

ground would discard this intuitive rule since 1) it is not specific enough³¹ and 2) it *does not* work for *all* types of verbs (e.g. "On avait *finir"), and science would thus require the formulation of a better and more general rule. Then, coming back to the essence of our discussion, if linguists are to describe language, is it epistemologically legitimate for them to disregard the way an ordinary speaker uses her own rules, which, in the end, along with other speakers' idiosyncratic rules provide the data, on which basis linguists propose "*a*" description, and formulate rules?

In other words, taking speakers' metalinguistic intuition seriously results in the following unsatisfactory epistemological dilemma:

- 1. Linguists acknowledge that the data collected jumbles up speakers' idiosyncratic sets of rules, but they still endeavor to formulate 'average rules' that do not represent any speaker³²;
- 2. Or they presuppose (or most likely believe) that all speakers use the same underlying rules, which "just" need to be uncovered and then scientifically formulated.

Once again, I have no solution to this puzzling state of affairs.

Let us now turn to the next point, which concerns theoretical consistency with language acquisition data.

2.5.3 Be consistent with language acquisition data

It might look unfair to evaluate the validity of a linguistic theory against language acquisition data, if it was initially meant to describe mature speakers' language. However, if we consider that the early years of language acquisition set the foundation for further developments and effortless language use in later life, then it seems reasonable to examine whether a linguistic theory is consistent with language acquisition data. Whenever it fails to do so, then possible reasons for this incompatibility should be discussed.

While describing French Pr, Morin (1987: 37) proposes–among others–the following "implication rule": "Pr2 serves as basis for Pr1". This rule does not accord with the child acquisition data provided by Bassano *et al.* (2001: 125), who mention that children use 1^{st} before 2^{nd} person. Morin anticipates this objection and clarifies the status of his implication rules:

³¹ Such a legitimate scientific criticism would be unfair to the pupil's metalinguistic self-justification, which, although general, was limited to the written sentence she was asked to comment.

³² Recall Bates *et al.*'s (1995: 151) discontent mentioned earlier: "the Average Child is a fiction, a descriptive convenience". Is the "average rule" a fiction as well?

ce ne sont pas des stratégies d'acquisition de la langue [...]. Elles appartiennent en propre à la grammaire de l'adulte et ne se sont probablement mises en place que progressivement pendant les premiers stades de l'acquisition. (Morin 1987: 38)

If these implication rules belong to adults' grammar proper, then 1) what kind of strategies do children use while acquiring language? 2) When, how and why does the shift from child grammar to adult grammar happen? Since some of Morin's implication rules may not match children's data, then his hypothesized "gradual rule setting" procedure would need to be tested.

Let us take another more undisputable example. In an earlier model of his, Boyé (2000: 397) derives all French simple tenses from Imp's theme³³, thus yielding, among others, the following rules: a) Fut=Imp+(\Rightarrow); b) Inf=Imp+(e). This, again, raises an order of acquisition issue. Indeed, Inf and Fut being acquired before Imp (Bassano *et al.* 2001; Kilani-Schoch 2003; Sabeau-Jouannet 1973), both a) and b) would be unavailable to the child, thus raising the same questions as above.

2.5.4 Be consistent with ordinary language use

In his description of Fut formation, Touratier compares Fut1-3 with PS1-3 inflections, that are indeed similar /e/, /a/, /a/ (see also Van Den Eynde & Blanche-Benveniste 1970: 417):

Elles sont comparables à celles du passé simple de la première conjugaison:

je chanterai, tu chanteras, il chantera; je chantai, tu chantas, il chanta"

et l'on peut les décrire en disant que le futur est marqué par un segment /(a)Ra/ (Touratier 1996: 38-39)

As stated decades ago by Benveniste (1966: 237ff), PS is no longer used in conversational interaction in contemporary French, thus precluding the use of *tu* and *vous*. Even the most frequent French verbs (*être, avoir, aller, vouloir,* etc.) *never* allow PS2 and PS5 (see Blanche-Benveniste 2002: 21), although they still appear, to my knowledge, in all contemporary conjugation textbooks. What is then the scientific status of Touratier's comparison once we acknowledge that 1) PS is almost exclusively used in written French, and 2a) does not allow *tu/vous*, and 2b) almost never *je/nous*, and that 3) Fut is still a commonly used tense that allows all persons? To be relevant, such an asymmetrical comparison would need to be scientifically justified since ordinary speakers probably do not make this comparison.

³³ Taking Pr4-5 as basis also goes against acquisition data.

3. Conclusion: "Wouldn't it be better to make the map conform to the yard?"

...In that Empire, the craft of Cartography attained such Perfection that the Map of a Single province covered the space of an entire City, and the Map of the Empire itself an entire Province. In the course of Time, these extensive maps were found somehow wanting, and so the College of Cartographers evolved a Map of the Empire that was of the same Scale as the Empire and that coincided with it point for point.

Of Exactitude in Science (Borges 1954/1972: 141)

While describing languages, linguists, like other scientists, run into Borges's cartographers' epistemological dilemma:

- i) be as accurate as possible
- ii) be as concise as $possible^{34}$

However this problematic situation is not the sole difficulty scientists run into. Language, as any scientific "object", is not directly observable, but needs to be theoretically reconstructed by linguists. A second major epistemological problem then arises in relation to i), and humorously summarized in Bill Watterson's comic strip:



Calvin and Hobbes (Watterson 2005: 407)

One of the greatest and most difficult challenges linguists have to face, as Hobbes (the tiger) wisely suggests to Calvin, is "to make the map conform to the yard" and not the other way round. Such an epistemological issue is mostly ignored in linguistics. Fradin (2003: 265), for instance, is one of the few linguists to tackle the problem, but he ends up stating an "epistemological disclaimer": "Le niveau conceptuel mis en jeu par les règles proposées ici est donc uniquement descriptif et ne prétend en aucun cas être causal". Why are linguists allowed to adopt such an epistemological stance toward their object? Could such disclaimers be used in mechanics, nuclear physics, medicine, etc.? Probably not. A "wrong" linguistic rule has no consequence whatsoever. It will never make a plane crash or kill human

³⁴ This tension is summarized in Van den Eynde & Blanche-Benveniste's (1970: 406) explicit goal: "Le but de toute analyse est d'arriver à décrire une multitude de formes en partant d'un nombre restreint d'éléments et de règles de combinaison entre ces éléments, qui permettent de présenter ces formes comme 'prédictibles'".

beings. Apart from conflicts resulting from differences in theoretical approaches, it is clear that nothing crucial is at stake in linguistics. But should this state of affairs allow us to proceed like Calvin and project our own theoretical conceptions onto language and ordinary speakers' expressed metalinguistic intuition without ever questioning the legitimacy of such an epistemological position? No.

No matter what, it would probably be better to make the map conform to the yard...

References

Albright, A. C. 2002. The Identification of Bases in Morphological Paradigms, PhD thesis under the supervision of B. Hayes, Los Angeles: UCLA.

Allan, K. 2003. "Linguistic metatheory", Language Sciences 25 - 6: 533-560.

- Baayen, R. H. 2007. "Storage and computation in the mental lexicon", in G. Jarema & G. Libben (eds), The mental lexicon: core perspectives, Amsterdam/London: Elsevier, 81-104.
- Bassano, D., Maillochon, I., Klampfer, S. & Dressler, W. U. 2001. "L'acquisition de la morphologie verbale en français et en allemand autrichien: II. L'épreuve des faits", *Enfance 53-2*: 117-148.
- Bates, E., Dale, P. S. & Thal, D. 1995. "Individual differences and their implications for theories of language development", *in* P. Fletcher & B. MacWhinney (eds), *The Handbook of Child Language*, Oxford/Cambridge: Blackwell, 96-151.
- Benveniste, É. 1966. Problèmes de linguistique générale, 1, Paris: Gallimard.
- Berg, T. 1998. *Linguistic structure and change : an explanation from language processing*, Oxford: Clarendon Press.
- Besse, H. 1991. "(Pré)conceptions et finalités des techniques d'enseignement de la grammaire d'une langue seconde ou étrangère", *Triangle 8*: 75-86.
- Blanche-Benveniste, C. 2002. "Structure et exploitation de la conjugaison des verbes en français contemporain", *Le français aujourd'hui 139*: 11-22.

Bled, É., Bled, O. & Berlion, D. 2010. Le Bled, Paris: Hachette.

- Borges, J. L. 1954/1972. *A Universal History of Infamy*, New York: Dutton, (translation of *Historia universal de la infamia* by N. T. di Giovanni).
- Boyé, G. 2000. *Problèmes de morpho-phonologie verbale en français, en espagnol et en italien*, PhD thesis under the supervision of J. Lowenstamm, Paris: Université Paris VII.
- Brissaud, C. 2002. "Travailler la morphologie écrite du verbe au collège", *Le français aujourd'hui 139*: 59-66.
- Brissaud, C., Chevrot, J.-P. & Lefrançois, P. 2006. "Les formes verbales homophones en /E/ entre 8 et 15 ans : contraintes et conflits dans la construction des savoirs sur une difficulté orthographique majeure du français", *Langue française 151*: 74-93.
- Brissaud, C., Cogis, D., Jaffré, J.-P., Pellat, J.-C. & Fayol, M. 2011. Comment enseigner l'orthographe aujourd'hui ?, Paris: Hatier.
- Bybee, J. 1995. "Regular morphology and the lexicon", *Language and Cognitive Processes 10-5*: 425-455.
- Chalmers, A. F. 1999. What is this thing called science?, Indianapolis: Hackett.
- Chomsky, N. 1969. Aspects of the theory of syntax, Cambridge: M.I.T. Press.
- 1957/2002. Syntactic structures, Berlin/New York: Mouton.
- Clahsen, H. 2006. "Dual-Mechanism Morphology", *in* K. Brown (ed.), *Encyclopedia of Language & Linguistics*, Oxford: Elsevier, 1-5.
- Corrigan, R. & Lima, S. D. 1994. "Introduction", in S. D. Lima, et al. (eds), The Reality of linguistic rules, Amsterdam/Philadelphia: John Benjamins, xiii-xxiii.
- Croft, W. 2000. *Explaining language change: an evolutionary approach*, Harlow/New York: Longman.

- Damourette, J. & Pichon, É. 1911-1930. Des Mots à la Pensée, Tome III, Paris: D'Artrey.
- Dubois, J. 1967. *Grammaire structurale du français. Le verbe*, Paris: Larousse.
- Fradin, B. 2003. *Nouvelles approches en morphologie*, Paris: Presses Universitaires de France.
- Gougenheim, G., Michéa, R., Rivenc, P. & Sauvageot, A. 1964. L'élaboration du français fondamental : étude sur l'établissement d'un vocabulaire et d'une grammaire de base, Paris: Didier.
- Guillaume, G. 1973. Principes de linguistique théorique de Gustave Guillaume, recueil de textes inédits préparé en collaboration sous la direction de Roch Valin, Québec/Paris: Presses de l'Université Laval/Klincksieck.
- Hempel, C. 1966. *Philosophy of Natural Science*, Upper Saddle River: Prentice Hall.
- Hurford, J. R. 1977. "The significance of linguistic generalizations", *Language* 53-3: 574-620.
- Kail, M. 2012. *L'acquisition du langage*, Paris: Presses universitaire de France.
- Kiefer, F. 2000. "Regularity", in G. Booij, et al. (eds), Morphologie: ein internationales Handbuch zur Flexion und Wortbildung. 1 Halbband, Berlin/New York: Walter de Gruyter, 296-302.
- Kilani-Schoch, M. 2003. "Early verb inflection in French: An investigation of two corpora", in D. Bittner, et al. (eds), Development of verb inflection in first language acquisition: a cross-linguistic perspective, Berlin/New York: Mouton de Gruyter, 269-295.
- Kilani-Schoch, M. & Dressler, W. U. 2005. *Morphologie naturelle et flexion du verbe français*, Tübingen: Gunter Narr.

- Komatsu, E. & Wolf, G. (eds) 1996. F. de Saussure: premier cours de linguistique générale d'après les cahiers d'Albert Riedlinger (1907), Oxford: Pergamon,
- Langacker, R. W. 1987. Foundations of cognitive grammar Volume I. Theoretical Prerequisites, Stanford: Stanford University Press.
- 1990. "The rule controversy: A cognitive grammar perspective", CRL Newsletter 4-3: 4-15.
- 2008. Cognitive grammar: a basic introduction, Oxford: Oxford University Press.
- Lass, R. 1980. *On Explaining Language Change*, Cambridge: Cambridge University Press.
- Lehmann, C. 2001. "Language Documentation: a program", *in* W. Bisang (ed.), *Aspects of Typology and Universals*, 83-97.
- Lerot, J. 1993. Précis de linguistique générale, Paris: Minuit.
- Marcus, G. F. 2000. "Children's Overrergularization and its Implications for Cognition", in P. Broeder & J. Murre (eds), *Models of Language Acquisition*, Oxford: Oxford University Press, 154-176.
- Martin, R. 2002. *Comprendre la linguistique*, Paris: Presses Universitaires de France.
- Martinet, A. 1974. *Le français sans fard*, Paris: Presses universitaires de France.
- Milner, J.-C. 1989. Introduction à une science du langage, Paris: Seuil.
- Morin, Y. C. 1987. "Remarques sur l'organisation de la flexion des verbes français", *International Review of Applied Linguistics* 77-78: 13-91.
- 2005. "La liaison relève-t-elle d'une tendance à éviter les hiatus?", Langages 158: 8-23.
- Nakisha, R., Plunkett, K. & Hahn, U. 2000. "Single- and Dual-route Models of Inflectional Morphology ", *in* P. Broeder & J. Murre (eds), *Models of*

language acquisition: inductive and deductive approaches, Oxford/New York: Oxford University Press, 201-222.

Petit Robert électronique 2013. L. Catach (ed), Paris: Robert.

- Pinker, S. 1999. Words and rules: the ingredients of language, New York: Basic Books.
- Plénat, M. 1987. "Morphologie du passé simple et du passé composé des verbes de l'« autre » conjugaison", *ITL Review of Applied Linguistics* 77: 93–150.
- Popper, K. R. 1972. *Objective knowledge: an evolutionary approach*, Oxford: Clarendon Press.
- Pouradier Duteil, F. 1997. Le verbe français en conjugaison orale, Francfort/Main: Peter Lang.
- Sabeau-Jouannet, É. 1973. "L'expression des modalités aspectivo-temporelles et son évolution", *Études de linguistique appliquée* 9: 91-100.
- Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. 2009. "Sequential processing of lexical, grammatical, and phonological information within Broca's area", *Science 326-5951*: 445-449.
- Saussure (de), F. 1916/1959. Course in General Linguistics, New York: Philosophical Library, (translation of Cours de linguistique générale by W. Baskin).
- 2002. Écrits de linguistique générale, Paris: Gallimard, (edited by S. Bouquet & R. Engler).
- Séguin, H. 1986. *Tous les verbes conjugués*, Montréal: Centre éducatif et culturel.
- Soutet, O. 1995. Linguistique, Paris: Presses Universitaires de France.
- Thagard, P. 2008. "Explanatory Coherence", in J. E. Adler & L. J. Rips (eds), *Reasoning: studies of human inference and its foundations*, Cambridge/New York: Cambridge University Press, 471-513.

Touratier, C. 1996. Le système verbal français, Paris: Armand Colin.

- Trask, R. L. 1999/2007. *Language and Linguistics*. *The key concepts*, London/New York: Routledge, (edited by P. Stockwell).
- Van den Eynde, K. & Blanche-Benveniste, C. 1970. "Essai d'analyse de la morphologie du verbe français", *Orbis 19*: 404-429.
- Watterson, B. 2005. *The complete Calvin and Hobbes. Book Three*, Kansas City: Andrews McMeel.
- Wolpert, L. 1993. *The unnatural nature of science*, Cambridge: Harvard University Press.

THE MORPHOLOGY OF SPATIAL P: A LOOK INSIDE ADPOSITIONS AND CASE

Francesco-Alessio Ursini Stockholms Universitet, English Department

Abstract

The goal of this paper is to offer a unified analysis of the morphological structure of spatial adpositions and spatial case markers in three languages: English, Spanish, and Finnish. This analysis combines Distributed Morphology assumptions with a Type-Logical formal treatment. Two key results emerge from this unified morpho-syntactic analysis. First, spatial adpositions (*behind, encima, laelta*) can be accounted as the result of merging different "types" of spatial morphemes, including spatial case markers and particles. Second, cross-linguistic syntactic phenomena involving these categories (e.g. argument demotion) can also be accounted for straightforwardly, via our analysis. We suggest that these results also support a "morphology all the way up" view of Distributed Morphology.

1 Introduction: an Overview on Spatial Adpositions and Cases*

In recent years, several works have thoroughly investigated the category known as "spatial P" (henceforth: SPs: Levinson & Wilkins 2006; Cinque & Rizzi 2010). Works belonging to distinct generative research programs have offered different fine-grained analyses of the syntactic properties of SPs. Examples include Head Phrase Structure Grammar (HPSG: Tseng 2000, 2004, 2005) and the Minimalist Program (Asbury *et al.* 2008). To see why this is the case, consider examples (1a)-(1e):

- (1) a. The boy has gone to in front of the table
 - b. The boy has arrived from on top of the hill
 - c. [to in front of [the table]]
 - d. [to [in front of [the table]]]
 - e. [from [on [top [of [the hill]]]]]

^{*}I would like to thank the participants of *Les Decembrettes 8* and an anonymous reviewer who gave very useful feedback on a preliminary version of this work. Thanks to my princess for the support, as always. The usual disclaimers apply.

Examples (1a)-(1b) contain the (complex) SPs to in front of and from on top of. In both examples, the NP the boy denotes a located entity or figure; the NPs the table and the hill denote the landmark objects or grounds of the underlying spatial relations (Talmy 2000: ch.1). Although all analyses agree on these key assumptions, they differ on the morphological structure they assign to SPs. Strongly lexicalist analyses suggest that SPs project a single syntactic head. Examples include early Government and Binding analyses (GB: van Riemsdijk 1978; Emonds 1985) and HSPG (Tseng 2000, 2004). Instead, later GB-based decompositional analyses suggest that SPs involve at least two hierarchically ordered heads. One head denotes a "directional" component of meaning (here: to, from). A second is a lower head denoting a locative component (in front of, on top of) (Jackendoff 1983, 1990; Wunderlich 1991; van Riemsdijk 1990; van Riemsdijk & Huysbregts 2007). The structures in (1c) and (1d) respectively illustrate lexicalist and decompositional analyses in a theory-neutral (i.e. label-free) format. Importantly, both structures show that these approaches tend to fall silent on one aspect of SPs: their morphological structure. One version of the minimalist program (Chomsky 1995) that partly addresses this problem is the so-called "cartographic approach". Cartographic approaches assume that SPs, qua functional elements, can project a complex sequence of heads, one per identifiable morpheme (Koopman 2000; Asbury 2008; den Dikken 2010; Svenonius 2010). For instance, to, in, front and of form a sequence of heads that in turn form a so-called "SP field"; an example is also on top of in (1e). Cartographic approaches, however, fall silent on certain sub-sets of SPs data, of which we offer preliminary English and Spanish examples:¹

- (2) a. The boys sit a-round the chairs
 - b. The shop lies a-cross the road
 - c. The chair is out-side the room
 - d. *El niño esta en-cima de la casa* The child is-S in-top of the house 'The child is on top of the house'
 - e. *El niño esta de-lante de la casa* The child is-S of-front of the house 'The child is in front of the house'

¹In these examples I mark putative segments of SPs via hyphenation, as an orthographic norm used to *emphasize* word structure (Oshima & Hogue 2006: 170).

Examples (2a)-(2c) include the English SPs *a-round*, *a-cross*, *out-side*; similar SPs are *a-mong*, *be-hind*, *a-long*, and so on. Examples (2d)-(2e) include some Spanish counterparts of these SPs: *en-cima* 'on top', and *de-lante* 'in front'. Differently from English, these SPs must always combine with the relational (S)P *de* 'of' (Fábregas 2007). Descriptive and theoretical analyses often acknowledge that their morphological structure seems to include a prefix, and a noun-like element (English: Huddlestone & Pullum 2002: ch.7; Svenonius 2010: fn. 1; Spanish: Butt & Benjamin 2004: ch.4; Fábregas 2007: §2). Nevertheless, both classes of SPs are often labeled as *simple* SPs, their morphological structure being still unaccounted for.

A similar problem emerges when one looks at *spatial case markers* (henceforth: SCMs). Cartographic approaches build on Fillmore (1968), and capture the similarities between these spatial parts of speech by treating SCMs as part of the SP field. We illustrate this assumption via Finnish, a language with a wealth of descriptive and theoretical analyses (Karlsson 1999; Kracht 2004; Asbury 2008; Caha 2009). Standard minimalist analyses suggest that SCMs act as suffixes on ground NPs, and project a Kase head (Svenonius 2007; Asbury 2008). We show this analysis in (3)-(4):²

- (3) a. *Mario on huon-essa* Mario is house-INESS. 'Mario is in the house'
 - b. [KaseP[NP huon-]-essa]
- (4) a. Mario on auton ed-essä Mario is car-GEN front-INESS.
 'Mario is in front of the car'
 - b. [_{KaseP}[_{Kase'P} auton] ed-essä]

Example (3) shows that a Kase Phrase corresponds to a Kase head, here the Inessive SCM *–essa* (locative 'in'), combined with the ground NP *huon*-'house'. Example (12) shows that, when *postpositional* SPs such as *ed-essä* 'in front' occur in a sentence, a non-spatial case marker (the Genitive), combines with a ground NP. The SP *ed-essä* bears the Inessive SCM *-essä*, as a suffix to the SP morpheme *ede-* 'front'. Although the structures in (3b)-(4b) show that SCMs can be analyzed as part of the SP field, they do not

²We employ these glosses for Case morphemes: ESS.=essive, ACC.=Accusative, LAT.=Lative, GEN=Genitve, INS=Instructive, PAR=Partitive. We will use more fine-graned glosses for the Essive and Lative paradigms, in section 2.3.

explicitly state their status within the structure. As for English and Spanish SPs, their morphological properties seem still unaccounted for.

Overall, the data and analyses discussed in examples (1)-(4) seem to rise at least three empirical questions, with respect to SPs and SCMs. A first question is whether we can offer a fully derivational account of SPs for the morphological structure of SPs such as *across* or *encima*. A second is whether this treatment can be extended to SCMs, as in the case of Finnish *edessä*, hence offering a unified account for SPs and SCMs. A third is whether this account can also capture the *syntactic*, sentence-level distributional properties of prepositions and postpositions (*qua* SPs) and SCMs.

The goal of this paper is to answer these three questions via a unified theoretical perspective. Hence, we organize our paper as follows. Section 2 outlines a broader set of *explananda*. Section 3 presents the formal background: a combination of *Distributed Morphology* (DM: Embick & Noyer 2001, 2006) and the *Type Logical* (TL) calculus formal apparatus (Moortgat 2010). Section 4 offers our analysis; section 5, the conclusions.

2 The Data: A Broader Picture on SPs and SCMs

2.1 The Data: The Morphological Structure of English SPs

The goal of this section is to offer a broader overview of our SPs and SCMs, starting from an analysis of simple SPs in English. We maintain this label, even if it is in part a misnomer, for mere descriptive reasons. Simple SPs can actually be conceived as the combination of at least two morphemes. One is often a noun-like element referring to a body part or axis (e.g. *side*, *head*), the other a prefix that seems to lack a specific semantics (*be-*, *a-*). Recent works have labeled these noun-like morphemes "Axpart SPs", to highlight their ability to denote the specific "axis" of a spatial relation (Svenonius 2006, 2010; Asbury 2008). We implement this label for expository purposes, too.

The fact that simple SPs appear to be bi-morphemic has not gone unnoticed in the literature. It has been observed that the prefix *a*- seems to originate from the Middle English counterpart of the free SP *on*, similarly to the morpheme *be*- (Svenonius 2006: 79-84, 2010: fn.1). In this regard, English (spatial) prefixes share both syntactic and phonological properties with spatial(-like) prefixes in Slavic and Romance languages (*cf*. Matushansky 2002; Svenonius 2004; Mateu 2008). However, etymological arguments and cross-linguistic parallels, do not offer evidence for SPs being involved in synchronically *active* morphological processes. In order to solve this problem, we propose two more compelling arguments based on productivity. A first more compelling argument involves three sets of simple SPs that seem to involve active but seldom discussed processes of word formation. These three apparently heterogeneous sets consist of SPs that include various "combinations" of free spatial morphemes and affixes. A first set includes SPs that are the combination of some *particles (up, down)* or other simple SPs (*in, out, on, off*) with the Axpart morphemes *-side* and *-wards*. Recent works have suggested that (spatial) particles can be also considered as part of the SP field (Hale & Keyser 2002; Svenonius 2003, 2007; Cappelle 2004). Although these proposals differ with respect to their analysis on the syntactic status of particles, they converge on treating particles (e.g. *upwards*) as types of SPs, too. Thus, we include particles as part of our discussion of SPs.

We move to the other understudied SPs sets. A second set includes SPs that are the combination of particles and spatial nouns, which may denote specific locations or orientations (e.g. *uphill, downstairs*). A third set of data includes "cardinal" simple SPs, Axpart SPs that denote cardinal coordinates, such as *North, South, North-West* and so on (Levinson 1994; Levinson & Wilkins 2006; Svenonius 2006). We present two lists of examples per sub-set in (5), with the *proviso* that our lists can be non-exhaustive (i.e. open):

- (5) a. -wards type={back-wards, in-wards, to-wards, up-wards,...} (1^{st})
 - b. -side type={a-side, be-side, down-side, in-side, up-side, ...} (1st)
 - c. Particle type={up-front, down-hill, up-stairs, up-wards,..} (2nd)
 - d. Cardinal type={North, North-East, North-North-East,...} (3nd)
 - e. a-type={a-bove, a-cross, a-far, a-head, a-mong, a-round, a-top}
 - f. be-type={be-hind, be-low, be-neath, be-side, be-tween, be-yond}

The *-wards* and *-side* "types" in (5a)-(5b) present two connected but nonoverlapping sub-sets of the first sub-set, although these sets seem to draw morphemes form common inventories of particles and prefixes (e.g. *up*, *in*). The *Particle* type set in (5c) is our second sub-set, and includes SPs that carry a particle as a prefix, and Axpart or a similar other noun-like element as a basic SP (e.g. *hill*). The *Cardinal* type in (5d) is the third sub-set of understudied simple SPs. The two sets in (5e)-(5f), labelled as the *a*- and *be*types, include amply discussed SPs that can be treated as morphologically complex, although this complexity is not synchronically active.

The key unifying aspect of these data is that these three understudied sub-sets seem to be the result of simple word-formation processes. For instance, the SPs *upfront, downhill* and *Southwards* have emerged during the last two centuries, in standard British and American English (BNC, 2007; COHA, Davies, 2008). These processes apply to sub-sets of SPs, a category with a

small set of lexical items. Thus, they seem to belong to the lower end of the productivity spectrum, as they produce few but stable forms over time (Bauer 2005; Hay & Baayen 2002). For these reasons, though, they seem to meet *criteria* of potential and expanding productivity (Baayen 1994, 2009; Plag 2006). These processes seem active even if infrequent (potential productivity), resulting in the slow growth of our sub-sets, over the decades (expanding productivity). Two examples are *upfront* and *downhill*. Their frequencies (per million words) went from 0.08 (1960's) to 1.05 (2000's), and from 0.22 (19830's) to 8.76 (2000's), respectively. Our other examples follow a similar tack, so we do not report a full frequencies' list. Thus, we can conclude that if simple SPs are the result of a productive morphological process, then an account of their structure seems motivated.

A second more compelling argument is based on a *syntactic* property of SPs that, however, plays a role in their morphological structure as well. All the SPs we discussed so far can (or must) undergo *argument demotion*, hence occur without a ground NP. This usually occurs when the interpretation of the ground NP can be recovered from the previous context (Merchant 2001: ch.2; den Dikken 2010; Svenonius 2010). Thus, (simple) SPs can be treated as complement phrases of the verb they combine with, as we show in (6):

- (6) Mario goes behind/inside/in(to)/in front/in/North/North-West (the car)
- (7) Mario walks back/backwards//up/uphill (*the hill)

Treatments of argument demotion commonly treat the remnant SP (e.g. *in front, back*) as a phrasal complement of the verb, regardless of its exact categorical status. Hence, all the SPs in (6) are treated as complement phrase of the verb *goes*. A *proviso* is that SPs such as *in front of*, *North/North-West of* involve the demotion of the relational morpheme *of*, as well, only leaving Axpart or simple SPs as a result. Furthermore, *Particle* type SPs such as those in (7) cannot usually combine with ground NPs, as shown in the example. These facts show that our sub-types of SPs share two key properties, *qua* SPs. A first is their underlying morphological structure, and the second is their syntactic status as complement phrases of the verb, when argument demotion occurs. The second property, in turn, can and should be seen as the result of combining different types of spatial morphemes together, in a principled manner. Thus, simple SPs seem to be the result of different and yet related morphological processes, but also share the same syntactic properties. With these arguments in mind, we turn to Spanish.

2.2 The Data: The Morphological Structure of Spanish SPs

Our goal in this section is to provide two arguments for the morphological analysis of Spanish SPs. We focus on two known and two novel sets of data. The first argument is based on known Spanish SPs data, divided in two subsets. First, Spanish has two sub-sets of simple SPs that stand in complementary distribution, with respect to their *syntactic* properties. A first sub-set of SPs includes the prefixes *de-* and *en-*, which must combine with the relational P *de* 'of', and *cannot* undergo argument demotion. A second sub-set includes the prefix *a-*, and *must* involve argument demotion (Pavón 1999; Fábregas 2007; Ursini 2013a). Second, Spanish lacks an equivalent of the *-side-* and *-wards* English type of SPs. However, other simple SPs such as *hacia* 'towards' or *desde* 'from' can take simple SP phrases as their complement, to convey the same meaning. These sequences of SPs, known as *preposición trás preposición* 'preposition after preposition', in the literature (Bosque 1997; Pavón 1999), are shown in (8)-(9):

- (8) El niño esta de-lante/en-frente *(de la casa) The child is-S of-front/in-front *(of the house)
 'The child is in front of the house'
- (9) El niño esta a-lante (*de la casa) The child is-S of-front (*of the house)
 'The child is in front (of the house)'
- (10) El niño ha ido hacia ar-riba/a-trás (*de la casa) The child has gone towards up-part/behind (of the house)
 'The child has gone upwards/backwards (of the house)'
- (11) El niño ha llegado desde delante/enfrente *(de la casa) The child has arrived from ahead/in-front (of the house)
 'The child has arrived from in front of the house'

The pair (8)-(9) shows the distribution of the first sub-set of Spanish simple SPs, and how the de_{-} , en_{-} sub-sets stand in complementary distribution to the a_{-} set. In cases such as *delante* and *alante*, the two SPs seem to form a minimal pair with respect to their syntactic distribution, at least in Iberian Spanish (Fábregas 2007: 1-10). While *delante* cannot undergo argument demotion, *alante* must do so, lest (9) be ungrammatical. Examples (10)-(11) show, instead, how *hacia* 'towards' and *desde* 'from' can take another SP phrase as a complement. Demotion must target the ground NP of *delante* and *enfrente* 'in front', as (10)-(11) show. Therefore, if the morphological

structure of Spanish simple SPs affects their syntactic distribution via a form of (feature) percolation, then a morphological analysis seems to be necessary. Aside these two better understood sub-sets of simple SPs, two sub-sets are still in need of an account. A first novel sub-set includes Spanish SPs that have relatively heterogeneous meanings, but share the property of being indeed simple, or mono-morphemic SPs. Examples include SPs such as *fuera* 'out', *junto* 'close', *cerca* 'near' and *frente* 'ahead'. A second novel sub-set includes that of the understudied Spanish counterparts of the *Cardinal* type SPs (e.g. *Norte* 'North', *Oeste* 'West'). We present the key lists in (12):

(12) a. de- type={de-bajo, de-trás, d-entro, de-lante, en-cima} (1st set)

b. en- ype={en-cima, en-frente,en-tre} (1st set)

c. a- type= $\{a-lante, a-trás, a-bajo, a-rriba, a-fuera, a-dentro\}$ (1st set)

d. *PtP* type={*hacia SP, desde SP, a SP, de SP, ...*} $(2^{nd} set)$

e. Bare type={bajo,cerca,contra,frente,fuera,hasta,junto,lejos}(3rd set)

f. Cardinal type={Norte, Oeste, Este, Sur, Nor-Oeste,...} (4th set)

g. *El niño ha ido bajo/cerca/Norte *(de la casa)* The child has gone down/near/North of the house 'The child has arrived from in front of the house'

The lists in (12a)-(12c) are near-exhaustive, while those in (12d)-(12f) are not, since they include SPs that are the output of synchronically active processes. We label the second sub-set the *PtP* type, in reference to their label in the literature. We then label the third sub-set as the *Bare* type, to outline their lack of attached prefixes. We note that the *a*-type seems in part to correspond to the particle (*up/down*)-type in English. For instance, *arriba* and *abajo* roughly correspond to the English SPs 'upstairs' or 'uphill', and 'downhill', respectively. Aside these parallels, the distribution of these SPs is akin to the *de*- and *en*- types of SPs, since they cannot undergo demotion unless they occur within a *PtP* context, and combine with *de* (*cf.* (12g)). Hence, our novel data support an argument for their morphological analysis.

2.3 The Data: The Morphological Structure of Finnish SPs and SCMs

The goal of this section is to discuss the Finnish data, and outline which parallels we can find with the English and Spanish data. We first discuss some well-known facts, and then we move to two sets of understudied facts.

Most works on Finnish SCMs observe that there are least six spatial markers. Polysemy is a common trait: the Adessive case, for instance, can translate English SPs *at*, *around*, *on* (Nikanne, 1993; Kracht 2004). Other markers that seem to have spatial interpretations are the Genitive, Partitive, Instructive and Translative (Fong 1997; Asbury, 2008). Some authors suggest that SCMs involve "affixhaume" or "case-stacking" processes, with a "directional" marker stacked onto a "locative" marker (Kracht 2008, Svenonius 2008). The table in (32), based on Kracht (2004: 177), shows the six "pure" spatial cases and their stacked structure (*cf.* Kracht 2002, Asbury 2008). We will discuss this aspect in more detail in section 4.3, in which we offer our analysis to the distribution of these morphemes. The table is as follows:

(13)	Adessive	Ablative	Allative
	Tallo-l-la	Talo-1-ta	Talo-1-le
	'at the house'	' from the house'	'to the house'
	Inessive	Elative	Illative
	Talo-s-sa'	Talo-s-ta	Talo-n

Table 1: Finnish spatial case markers system.

(14) $\left[\text{DirP}\left[\text{LocP}\left[\text{NP} \text{talo-} \right] - 1 - \right] - 1a \right]$

The structure in (14) shows how the two layered SCMs can be hierarchically organized. The two markers *-l-* and *-la*, respectively instantiating a "locative" and a "directional head (*cf.* Svenonius 2010), are stacked on a "root" NP, *talo-* 'sea'. Aside these facts about SCMs and their paradigmatic structure, some works shed some light on the distribution of Finnish simple SPs. For instance, Kracht (2004: 177-178) and Svenonius (2007) observe that Finnish Axpart SPs occur in postpositional position, and combine with SCMs (*edessä* 'in front'). When these SPs occur, the ground NP is usually marked with Genitive case. Furthermore, Kracht (2004) observes that particles can also occur as SPs (e.g. *alas* 'down') that must undergo argument demotion. Overall, Finnish "old" data parallel English and Spanish "old" data.

Matters become even more complex when we look at two sub-sets of understudied Finnish SPs, each respectively offering a further compelling argument for a morphological (and unified) analysis of SPs and SCMs. A first set includes a set of SPs that corresponds to the union of *-side* and *Cardinal* types in English and Spanish. Note that the infix *-puole-* (roughly, 'side') can combine with other SCMs, by occurring between the locative and the directional markers. A second set includes a set of SPs derived from particles, via the suffixation of one of three case markers: Ablative,

Instructive and Partitive. Different particles can combine with one of these cases, to form an intransitive-like SP, like English *upwards* or Spanish *alante*. We offer a set of relevant examples in (15):

- (15) a. Mario istuu (auto-n) ed-es-sä Mario sits (car-GEN.) front-in-ESS.
 'Mario sits in front of the car'
 - b. *Mario on mennyt (auto-n) et-ee-n* Mario is gone (car-GEN.) front-in-LAT. 'Mario has gone in front (of the car)
 - c. *Mario on (huoneen) ulko-puole-l-la/sisa-puole-l-la* Mario is (room-GEN.) out-side-At-ESS./in-side-at-ESS. 'Mario is outside/inside the room'
 - d. *Lampu on (pöydän) ala-puole-l-la/ylä-puole-l-la* Lamp is (table-GEN.) down-side-at-ESS./up-side-at-ESS. 'The lamp is below/above the table'
 - e. *Tukholma on (Göteborgin) pohjo-puole-l-la/etelä-puole-l-la* Stockholm is (Gotheburg-GEN.) North/South-side-at-ESS 'Stockholm is North/South of Gothenburg'
 - f. *Mario on (*auto-n) kävellyt taka-perin/ylä-mäkeä/pohjo-see-n* Mario is (*car-GEN.) walked back-INS./up-PAR./North-in-LAT. 'Mario has walked backwards/upstairs'
 - g. Axpart type={ede-, jalj-,pai-,yla-,ympär-,a-, kautta-,..} (1st set)
 - h. Puole- type={pohjo-,et/ed-,ulko-,sisa, etelä-...} (1st set)
 - i. Particle type={taka-,ala-, alla-, ylä,läpi,...} (1st set)
 - j. Suffix type={X-PART, Y-TRANS, Z-ESS} (2^{nd} set)

The examples in (15a)-(15b) show that SPs such *ed-es-sä* and *eteen* can be treated as different realizations of one underlying SP, translatable as 'front'. While *ed-es-sä* corresponds to the locative (Adessive) form, *eteen* is its directional (Illative) counterpart (*cf. to in front*). The examples in (15c)-(15e) show some SPs that involve a particle, on which SCMs and the *-puole-* infix are stacked on (e.g. *ulko-puole-l-la* 'outside', lit. 'out-side-at-ESS'). These

examples show that not only *Cardinal* and *Particle* types SPs can combine with SCMs, but also that certain Axpart SPs have an equivalent structure (e.g. *ala-puole-l-la* 'above'). Example (15f), then, shows that particle-like SPs, akin to English *-wards* type SPs, can be further accessed via the occurrence of one of three possible case markers (Instructive, Partitive, Illative), only one being an SCM proper (the Illative). In all of these cases, demotion is either possible or obligatory, when intransitive SPs are involved. Also, in all of these cases the ground NP combines with the Genitive case, a case that can hardly be seen as purely spatial in nature.

The non-exhaustive lists in (15g)-(15j) offer a more compact overview of the two sets of understudied Finnish simple SPs that justify two arguments in favour of a morphological analysis. First, our SCM-based data show that Finnish counterparts of English and Spanish simple SPs involve a fairly complex, but relatively transparent morphological structure. As all of these SPs can or must undergo argument demotion, they also share the phrasal status of their cross-linguistic counterparts, and the fact that morphological and syntactic properties interact. Second, SPs and SCMs clearly interact in a regular way, with the further *proviso* that SPs seem to include different types, such as particles, but also locative, cardinal and Axpart SPs. Therefore, they offer us a further pair of compelling arguments for offering a unified analysis of SPs and SCMs, and answer our three empirical questions. For this purpose, we present our formal proposal and analysis in sections 3 and 4.

3 The Proposal: DM meets TL calculi

The goal of this section is to present the formal tools that we employ to tackle our three problems. We combine two distinct frameworks: *Distributed Morphology* (henceforth: DM: Embick & Noyer 2001, 2006; Harbour 2007; Harley 2010a, 2010b, 2012) and *Type-Logical calculi* (TL calculi: Jäger 2005; Moortgat 2010, 2011; Morryll 2011). A more thorough discussion of this integrated framework can be found in Ursini (2013a,b, 2015); Ursini & Akagi (2013a,b). Our choice is based on two reasons. First, DM is perhaps the only minimalist framework that addresses productive morphological processes, as the one discussed so far, while other approaches (e.g. Cartography) only address morpheme orders. Second, we also wish to extend this framework's treatment of SPs, as few works have investigated a small sub-set of these data (Thomas 2001, 2004; Ursini & Akagi 2013a, b).

In order to further motivate our choice, we present the three key assumptions that underpin DM. First, morphology and syntax are taken to be a single derivational system, which recursively combines morphemes into larger structures (words, phrases, sentences). Differently from syntax-centric views of the framework (e.g. Halle & Marantz 1993), we like to conceive this

approach as "morphology all the way up". Second, morphemes correspond to clusters/sets of features, with different combinations corresponding to categories such as SPs, NPs or other Phrases. Third, the output of the morphological system maps onto the semantic and phonological components of grammar. In this paper we mostly focus on some phonological phenomena (vocabulary insertion and fusion), leaving semantic matters.

In order to explicitly represent these assumptions, we import some core aspects of TL calculi. Since we want to give a unified account of SPs and spatial SCMs, we need a precise, formal account of their morphological properties and syntactic distribution. For this purpose, we show that some key assumptions found in TL calculi suffice to formally account the data at hand. Two such basic assumptions play a key role, which we define below.

First, in TL calculi parts of speech are mapped onto *types*, which can be considered as either being "complete" or "incomplete" bits of morphological information. Complete types represent derivational units (morphemes, phrases) that can stand as distinct, independent units (e.g. *np* for NPs as *the girl*). Incomplete types are units that must combine with other units, to form a complete type. For instance, an intransitive verb such as *runs* can be assigned type *s/np*. So, if it combines with an *np* item, *the girl*, then the result is the sentence *the girl runs*, which is assigned the type *s* of sentences.

Second, in TL calculi types can be combined in a principled way, via a small set of operations. We use the connectives "/" and "•" to represent the Merge and the Product operations, respectively (Moortgat 2010: § 2; Morrill 2011: ch. 1). We define Merge as a binary, associative operation, and Product as also a (strictly) non-commutative operation. We only implement the rightassociative version of merge, which we label Merge right, leaving aside the use of other possible "slash" connectives (e.g. merge left "\", Jäger 2005's connective "i" for anaphors), We then assume that derivations compute information about types in a top-down manner, following psychological models of word production (Levelt 1989; Hay & Baayen 2002; Phillips 2006; Jarema & Libben 2007). For compound types, we use the Product "•" operator. This operator represents types that act as "compound" units in derivation, taken in a pair-wise order. Thus, while Product explicitly represents DM's second assumption about morphemes' structure, Merge represents DM's first assumption about morphemes' combinatoric principles. We then add a novel assumption about the basic set of atomic types in our lexicon. Standard definitions of atomic types in type-logical calculi take a perhaps naïve view of parts of speech, representing them via types such as *np* and s (cf. Jäger 2005; Morrill 2011: ch.1). Current minimalist approaches, including DM, suggest that such traditional categories can be reconstructed as instances of more abstract categories. For instance, Hale & Keyser (2002) analyse any "concrete" morpho-syntactic category (SPs or NPs) as abstract heads, which can also vary in valence. For instance, bare NPs (*girl*) are treated as heads with 0-valence: they cannot take any argument phrases. Transitive verbs (copula *is*), instead, are treated as heads with 2-valence. In our framework, we capture these ideas by implementing a universal type p (for "phrase"), and a set of recursive rules to derive complex types:

(16)	a. p is a morphological type	(Lexical type)
	b. If x is a type and y is a type, then x/y is a type	(F. type: merge)
	c. If \mathbf{x} is a type and \mathbf{y} is a type, then $\mathbf{x} \bullet \mathbf{y}$ is a type (F. type: product)
	d. If d. x/y is a type and y is a type, then $(x/y)\bullet y \vdash x$, $y \bullet y$	(x / y)⊬ x (FT: :MI)
	e. If x/y is a type and y/z is a type, then $(x/y) \bullet (y/z) \vdash x/z$: (Cut rule)
	f. Nothing else is a type	(Closure rule)

The rules read as follows. Rule a. introduces our basic type; rules b. and c. define how heads or complex morphemes are formed via Merge and Product, respectively. Rule d. defines the principle of "forward application", which defines how Merge combines units into larger constituents (Moortgat 2010, 2011). Rule e., known as the "cut rule", defines a special case of Merge, by which two incomplete units can be conflated/fused in a principled way. We discuss its exact import when we will discuss the data in more detail. Rule f., then, says that no other rules are found that can derive type structures.

Via this set of assumptions, we can generate various type sets. For our purposes, the set $TYPE' = \{p \bullet p, p \bullet p/p \bullet p, p \bullet p/p \bullet p/p \bullet p\}$ will suffice. This is that smallest type set that can represent phrasal elements, prefixes and relational heads, respectively, hence the smallest set that allows us to capture our data. The product type $p \bullet p$ can involve *n* basic types (e.g. $p \bullet p \bullet p \bullet p)$, but for the sake of clarity we only represent the basic binary type. We will motivate the full import of these assumptions in the next section, as their application to the data at hand will enlighten their usefulness. We will also make a brief clarification about the possible *values* that features can take in our system, and how this aspect is related to our data.

We add one final aspect of our formalism. We define a simple pre-order as the pair of an interval set *I*, and an addition operation "+", i.e. $\langle I, + \rangle$. This pre-order represents an *index set*, which in turn allows to represents the steps in a derivation as ordered elements. With these formal tools at our disposal, we turn to our derivational analysis of our data.

4. The Analysis: The Data

4.1. The Analysis: The English Data

Our goal in this section is to account the data via our proposal. We start from English SPs. For reasons that will become clear in a few paragraphs, we start from *Cardinal* and *Particle*-type SPs, respectively our second and third subset of understudied simple SPs. Recall that SPs such as *North* mostly are noun-like elements that have become part of the SP field, and can undergo argument demotion (e.g. *North (of the car))*. The same reasoning can be extended to all other Axpart morphemes, too (cf. *the North, the front, the cross, etc.*). We take this fact as evidence that the "remaining" SPs act as full phrases, hence complements of a verb they merge with, like their non-demoted SP phrase counterpart. Therefore, we assign the type $p \bullet p$, the type of bundled morphemes/features *and* of complete phrases, to these SPs.

We then suggest that argument demotion supports a parallel analysis for both particles (e.g. up, down) and locative SPs (*in*, out). Recall that when the SPs *into* or *onto* undergo demotion, the morpheme to (and the ground NP) is demoted, and only *in* and *on* are spelt out. Hence, these locative SPs can also be assigned type $p \cdot p$. Recall that the first sub-set of SPs (*-wards* and *-side* types) follow a similar pattern: *inside* and *upwards* can and must, respectively, act as complements of a verb after demotion. Therefore, we can also assign the type $p \cdot p$ to these SPs, and hence to all SPs morphemes.

We turn to the set of "old" simple SPs, those carrying the *a*- and *be*- prefixes. We assume that these prefixes act as zero-derivational morphemes that turn spatial nouns into SPs (*cf.* Svenonius 2006, 2010). Hence, they are assigned type $p \cdot p/p \cdot p$. Our assumption captures the idea that "bare" Axpart SPs (e.g. *front, cross*) change category, from bare NPs to SPs, when they are merged in an SP morpho-syntactic context. We represent this fact by assuming that they must change feature value, type-wise (Adger 2010; Morryll 2011). Thus, while *cross* is an NP denoting an object with a given spatial structure, *across* is the SP counterpart denoting a cross-like path. In order to keep our notation readable, we represent different feature values via indexes, in derivations (i.e. *p1*, *p2*). With these simple assumptions about our morphemes in hand, we can offer a compact type assignment in (17), and show how our SPs in lists (8) are derived. We mark the "Merge Introduction" operation as MI, and the selection of a lexical item in a derivation as LS (Lexical selection). We present our initial derivations in (18)-(19):

(17) a. *p*•*p*={*up*, -*side*, *upside*, *North*, *West*, *in*, *front*, *in front*, *cross*, ... }

b. *p•p/p•p* ={*a*-,*be*-}

(18)	a. t. [North $_{p \bullet pI}$] t+1. [West $_{pI \bullet p}$]	(LS) (LS)
	<i>t+2.</i> [North _{<i>p</i>•<i>p</i>]}],[West _{<i>p</i>1•<i>p</i>}] \vdash [North-West _{<i>p</i>•<i>p</i>}]	(MI: cut rule)
	b. t. $[a_{-p \bullet p/p \bullet p2}]$	(LS)
	$t+1. \left[-cross_{p2 \bullet p}\right]$	(LS)
	$t+2. [a_{pop/pop2}], [cross_{p2op}] \vdash [pop[a_{pop/pop2}[cross_{p2op}]]$	(MI)
(19)	t. $[in_{p \bullet p 3}]$	(LS)
	$t+1. [\operatorname{front}_{p3 \bullet p}]$	(LS)
	$t+2. [in_{p \bullet p 3}], [front_{p 3 \bullet p}] \vdash [in front_{p \bullet p}]$	(MI: cut rule)

The type assignment in (17) partitions all the different descriptive SP types in two sets: one for morphemes as sets of features, and one for prefix elements. In our derivational and more "dynamic" approach, prefixes are those elements that take a free morpheme (to their right) as in input, and return a phrasal element as an output. Phrasal elements, in turn, can be "simple" or complex elements, respectively particles and SPs, that can act as arguments of some other element. The derivation in (18a) says that the Merge of SPs North and West derives the SP North-West, which has the same type **p**•**p** of its constituting morphemes. Hence, we account that North, West and North-West belong to same underlying type of SPs, the Cardinal-type. Via the cut rule instance of Merge, we furthermore show that one "shared" type is removed, and the resulting type is obtained by combining the remaining types together. The same analysis can be extended to SPs belonging to the Particle type, such as uphill, and -wards types of SPs (e.g. backwards). In other words, this analysis can be applied to all three subsets of "new" SPs. The derivation in (18b), then, shows that almost the same type of analysis can be extended to "old" simple SPs. When the prefixes a- and be- merge with Axpart SPs such as -cross, the resulting SP is shown to be another SP, in this case *across*. This element is also of type $p \bullet p$, like the other simple SPs. The derivation in (19), instead, shows that we can derive the structure of "old" SPs, such as in front or on top, via the same set of assumptions that adopt for our novel data. Via the use of feature values, we can also account the ungrammaticality of unattested combinations. Although both West and front can be represented as compound types, their feature values differ, so their merging will be ungrammatical: *Cardinal* types of SPs cannot combine with

other sub-types of SPs (*cf.* **in-West*). Overall, the data in examples (2)-(7) can now find a unified, principled account.

Before we discuss the other data, however, we wish to make a comment on morpho-phonological matters. We suggest that the occurrence of the cut rule in morphological derivations can determine the assimilation ("fusion", in DM) of the two inserted exponents (Embick & Nover 2001, 2006). Since the morphological component "fuses" structures, the phonological component can mirror this process by producing North-West/Northwest or uphill, via fusion. We observe that there seems to be some idiosyncratic variation as to which SPs can undergo fusion: we have pairs such as up to vs. into, but also SPs such as in front and ahead. We conjecture that such cases can be seen as idiosyncrasies in orthographic norm, since there seems to lack any relevant difference among the prosodic properties of these vocabulary items. Before we continue, note that we can also block the formation of SPs such as *bemong, *be-cross, *a-hind in our analysis. This is the case, as we can assume that -mong and be- have non-matching feature values. However, a complete account would require of the semantic effects of this analysis. We need to defer it to another occasion, as we need to move to our Spanish data.

4.2. The Analysis: Spanish Data

Since we now have an analysis of the English data, our analysis of the Spanish data can follow a more compact format. Our sub-sets of novel data, involving *Cardinal* and *Bare* types of SPs (*Norte*, *cerca* respectively), can be assigned the type $p \bullet p$ as their English counterparts. For our sub-set of old data, involving SPs including the prefixes a-, de-, en-, we can also extend the corresponding English analysis. We assign the type $p \bullet p/p \bullet p$ to prefixes, and the type **p**•**p** to their Axpart morphemes (e.g. -frente, -lante, and so on). So, the resulting SPs (e.g. enfrente, alante and delante) are now accounted to be of type *p*•*p*, much like their English counterparts (e.g. *be-hind*). However, to capture the Spanish argument demotion data, as well as the occurrence of the relational SP de, we need a further step. We need to project our morphological analysis "all the way up", to a syntactic level of analysis. To achieve this result, we proceed in three steps. First, we assume that a minimal difference between a- and de- prefixes lies in the value attributed to the output type, the type $p \bullet p$ obtained after merging with an Axpart SP. Second, we assume that de, as a 2-valence head, can be assigned the (relational) type $p3 \bullet p/p \bullet p/p \bullet p$, a type that takes SPs with certain feature values on its "left". While SPs such as delante match the specific (left-)type and value of de, SPs such as *alante* do not. The converse holds for a- type prefixes, such as *alante*. We then make two supplementary assumptions, in

order to derive sentences and place ourselves in a position to account

argument demotion data. We assign the type $p \circ p$ to figure NPs such as *el* niño, and the type $p \circ p/p \circ p/p \circ p$ to the copula *estar*, as a transitive verb and head. We offer a type assignment in (20), and key derivations in (21)-(23):

(20) a. p•p={sobre, Norte, Oeste, enfrente, cerca, el niño, ar-riba, riba,...}
b. p•p/p•p={a-,de-,en-,...}

c. *p*•*p*/*p*•*p*/*p*•*p=*{*esta*, *de*, *hacia*, (*P*), ... }

(21)	t. $[el niño_{p \bullet p}]$	(LS)
	$t+1. [esta_{p \bullet p/p \bullet p/p \bullet p}]$	(LS)
	$t+2. [el nino_{p \circ p}], [esta_{p \circ p/p \circ p/p \circ p}] \vdash [p \circ p7p \circ p] [el niño_{p \circ p}] esta_{p \circ p/p \circ p/p \circ p}]$	(MI)
	$t+3.$ [delante _{$p \bullet p3$}]	(LS)
	<i>t+4.</i> $[p \bullet p p \bullet p$ [el niño $p \bullet p$] esta $p \bullet p / p \bullet p / p \bullet p$], [delante $p \bullet p a$] \vdash	
	$[p \circ p[el nino_{p \circ p}] esta_{p \circ p/p \circ p/p \circ p} [delante_{p \circ p 3}]]$	(MI)
	$t+5. \left[\operatorname{de}_{p3 \bullet p/p \bullet p/p \bullet p} \right]$	(LS)
	<i>t+6.</i> $[p \circ p[el niño_{p \circ p}] esta_{p \circ p/p \circ p/p \circ p} [delante_{p \circ p3}]], [de_{p3 \circ p/p \circ p/p \circ p}] \vdash$	
	$[p \bullet_p [el nino_{p \bullet_p}] esta_{p \bullet_p / p \bullet_p} [p \bullet_p / p \bullet_p [delante_{p \bullet_p 3}] de_{p 3 \bullet_p / p \bullet_p / p \bullet_p}]]$	(MI)
	$t+7. [la casa_{p \circ p}]$	(LS)
	<i>t+8.</i> $[p \circ p[eln_{p \circ p}]esta_{p \circ p/p \circ p/p \circ p}[p \circ p/p \circ p[delante_{p \circ p 3}]de_{p 3 \circ p/p \circ p/p \circ p}]], [lc_{p \circ p \circ p}]$	_p]⊢
	$[p_{op}[eln_{pop}]esta_{pop/pop/pop}[p_{op}[delante_{pop3}]de_{p3op/pop/pop}[lc_{pop}]]]$	(MI)
(22)	$t+k. [p_{op}[el nino_{p \circ p}] esta_{p \circ p/p \circ p/p \circ p} [alante_{p \circ p 4}]]/[de_{p3 \circ p/p \circ p/p \circ p}] \vdash^*$	(D.Cr.)
(23)	$t+k. \begin{bmatrix} p \bullet p/p \bullet p \end{bmatrix} \begin{bmatrix} hacia_{p \bullet p} \end{bmatrix} \begin{pmatrix} P \end{pmatrix}_{p \bullet p/p \bullet p/p \bullet p/p} \end{bmatrix}$	(LS)
	t+k+1. [arriba _{p40p}]	(LS)
	$t+k+2$. $\begin{bmatrix} 1 & 1 & 1 \\ p \bullet p/p \bullet p \end{bmatrix} \begin{bmatrix} hacia_{p \bullet p} \end{bmatrix} (P)_{p \bullet p/p \bullet p/p \bullet p/p \bullet p/q} \end{bmatrix}$, $\begin{bmatrix} arriba_{p4 \bullet p} \end{bmatrix} \vdash$	
	$\begin{bmatrix} p \bullet p \\ hacia_{p \bullet p} \end{bmatrix} (P)_{p \bullet p/p \bullet p/p \bullet p/q} \begin{bmatrix} arriba_{p4 \bullet p} \end{bmatrix}$	(MI)

Note that we have used use abbreviated forms (e.g. *eln* for *el niño*), for simple reasons of space. The derivation in (21), which in turn is based on (8), reads as follows. The figure NP *el niño* is first merged with the verb *esta*; the new-formed constituent *el niño esta* is then merged with *delante*. When *de* is merged, *delante* becomes the specifier of this newly merged head, since *de* matches the features of *delante*, and must attach to it as a consequence (Phillips 2006; Adger 2010).³ The ground NP and *de* are merged accordingly.

³ Here we simplify derivational matters: the operation *co-application* would apply, and derive the structure we obtain in steps t+6 (see Ursini & Akagi 2013b: fn.16).
Some further observations are in order, before we continue. We gloss over the structure of our verbs and figure NPs, as they are not crucial to our analysis. Also, we follow analyses of SPs' structure such as Hale & Keyser (2002)'s "P-within-P hypothesis", rather than cartographic-bound analyses. This is the case, since we treat our set of SPs (*enfrente, alante*) as elements that are merged in the specifier of a relational head, the SP *de*. We then retroactively extend this assumption to English SPs. Hence, *in front of the hill* and similar other SPs receive the same analysis as *enfrente de*. Other simple English SPs (*behind, uphill*) can receive an equivalent analysis. For instance, the underlying structure of *behind the car* would approximately be *behind (P) the car*. This analysis also entails that the merge of prefixes such as *be*- and Axpart elements such as *-hind* forms a phrasal element. Thus, we offer an analysis of this category similar to previous analyses (e.g. Svenonius 2004, 2010), although via a different theoretical trajectory.

We turn to the simplified derivations in (22)-(23). In (22) we show that when *alante* merges with *de*, their features do not match, hence causing the derivation to crash, unlike in the *delante* case. Hence, we can now account the distribution of the different types of argument demotion in English and Spanish that we discussed in examples (6)-(11), via this analysis. This analysis can also be easily extended to the *PtP* types of SPs (e.g. *hacia arriba*), our second sub-set of Spanish data, as the compressed derivation in (23) shows. The SP *hacia* merges with a silent head (i.e. "(P)") and with *arriba*, forming an SP phrase that, in turn, cannot merge with a ground NP, because of its feature values. This is the case, as *arriba* lacks the features that allow it to successfully merge with a ground NP and the relational *de*. Thus, we can now also account the structure of the *PtP* sub-types of SPs.

Overall, our Spanish data in (8)-(12) can now find a unified account that can be retroactively extended to the English SP data in (1)-(7). SPs such as *in front of, on top of, North of* and several others can receive the same analysis, with respect to morphological structure and distribution. Furthermore, our examples (1a)-(1b) can now find an account, as *to in front of* and *from on top of* can be seen as English counterparts of the Spanish *hacia SP* types. Hence, we can now claim that we have offered a positive answer to our first empirical question: whether English and Spanish SPs can receive a unified answer. We can now move to the Finnish data, and the other questions.

4.3 The Analysis: Finnish Data

Our goal in this section is to offer an account of the Finnish data. Since we now have an account of the structure of all types of SPs, of SP phrases and the sentences they are part of, we can directly focus on offering a type assignment for SCMs. As we concluded in section 2.3, SCMs and other case markers invariably act as suffixes on either Axpart "root" SPs, or on ground

NPs. In our type assignment, summarized in (24), the simplest type assignment for suffixes is $p \bullet p/p \bullet p$, a type that has interesting derivational properties when it is iterated, as in our case-stacking SPs. We then offer a derivation of its SP (*ulkopuollea* 'outside'), based on (15c), in (25):

(24) a. *p*•*p*={ede-, jalj-,pai-,yli-,ympär-,a-, kautta-, Mario, edessa,...}
b. *p*•*p*/*p*•*p*={-s-,-l-,-0-,-puole-,-sa,-la,-on,-ta,...}

c. *p*•*p*/*p*•*p*/*p*•*p*^{*i*={on,(P),...}}

(25)	<i>t.</i> [ulko- _{p•p/p•p}]	(LS)
	<i>t+1.</i> [-puole- _{<i>p</i>•<i>p</i>/<i>p</i>•<i>p</i>]}	(LS)
	<i>t+2.</i> [ulko- $_{p \circ p}$],[-puole- $_{p/p \circ p}$] \vdash [ulkopuole- $_{p \circ p}$]	(MI: cut rule)
	$t+3. \left[-1 - p \cdot p \cdot p \cdot p \cdot p \right]$	(LS)
	<i>t+4.</i> [ulkopuole- _{p•p}],[-l _{p•p/p•p}]⊢[ulkopuolel- _{p•p}]	(MI: cut rule)
	$t+5. [-la_{p \bullet p/p \bullet p}]$	(LS)
	<i>t+6.</i> [ulkopuolel- $_{p \circ p}$],[-la $_{p \circ p / p \circ p}$] \vdash [ulkopuolella $_{p \circ p}$]	(MI: cut rule)
(26)	<i>t+k.</i> [$_{p \bullet p}$ [Mario $_{p \bullet p/p \bullet p/p \bullet p}$] on $_{p \bullet p/p \bullet p/p \bullet p}$ [huoneen $_{p \bullet p}$]]	(LS)
	$k+1. [(P)_{p \bullet p/p \bullet p/p \bullet p}]$	(LS)
	<i>k+2.</i> $[p_{\bullet p}[Mario_{p \bullet p/p \bullet p/p \bullet p}] on_{p \bullet p/p \bullet p/p \bullet p} [huoneen_{p \bullet p}]], [(P)_{p \bullet p/p \bullet p/p \bullet p}] \vdash$	
	$[p \bullet p[Mario_{p \bullet p/p \bullet p/p \bullet p}] on_{p \bullet p/p \bullet p/p \bullet p} [p \bullet p/p \bullet p[huoneen_{p \bullet p}][(P)_{p \bullet p/p \bullet p/p \bullet p}]$	
	<i>k+3.</i> [ulkopuolella _{p•p}]	(LS)
	$k+4. \left[{}_{p \bullet p} \left[\text{Mario}_{p \bullet p/p \bullet p/p \bullet p} \right] \text{on}_{p \bullet p/p \bullet p/p \bullet p} \left[{}_{p \bullet p/p \bullet p} \left[h_{p \bullet p} \right] \left[\left(P \right)_{p \bullet p/p \bullet p/p \bullet p} \right], \left[u_{p \bullet p} \right] \vdash$	
	$\left[p \bullet_p \left[\text{Mario}_{p \bullet_p / p \bullet_p / p \bullet_p} \right] \text{on}_{p \bullet_p / p \bullet_p / p \bullet_p} \left[p \bullet_p \left[h_{p \bullet_p}\right] \left[(P)_{p \bullet_p / p \bullet_p / p \bullet_p}\right] \left[u_{p \bullet_p}\right] \right] \right] (MI)$	

The type assignment in (24) should be straightforward to read, by this point. The derivation in (25) shows how an SP marked for Inessive case (*ulkopuolella* 'outside') can be derived. An Axpart morpheme (*ulko-juole-la*). In other words, case-stacking can be simply treated as a derivational sequence of morphemes being merged together, in an orderly way. Ground NPs can be derived accordingly, although via only one affixal cycle (e.g. *huoneen* 'house-GEN'). Thus, we can account case stacking as a simple but principled instance of iterated merging of case suffixes.

With this result in hand, we show how our Finnish sentences emerge via the partial derivation in (26). Note that we show how an explicit ground NP is merged, whereas a derivation that would roughly "skip" steps t+k to k+2 would derive a sentence involving argument demotion (e.g. (15f)), instead. In (26), we would obtain *Mario on ulkopuolella* ('Mario is outside') as a result.

Note, furthermore, that we follow the previous literature on Finnish in also assuming a silent (P) head to take ground NP and SP as its arguments (Nikanne 1993; Kracht 2002, 2004). By this point, then, we can claim that we can offer a general account of Finnish SPs and sentences, covering examples (13)-(15). We have a full account of our data. Before we move to conclusions, however, we wish to discuss four consequences of our analysis. First, we can assign the same type to SCMs and prefixes, qua members of the class of affixes. Our rule of forward application blurs the distinction between the two directions of affixation. Second, we consequently can capture different linear orders of morphemes without any additional assumptions. Intuitively, in English and Spanish the locative morphemes in and en precede their Axpart matches, front and frente. In Finnish, the (compound) Inessive case *puole-l-la* follows the Axpart morpheme *ulko-*. In our approach, the input types and the derived type are the same across the three languages, although the linear order in English/Spanish is the mirror of the linear order in Finnish. Indeed, we can actually predict that two symmetrical linear realizations of the same underlying derivation (affix plus argument) may be realized as forms of cross-linguistic variation (Harbour 2007; Harley 2010b). Third, both optional and obligatory argument demotion patterns can be now explained, although we only discussed in detail the first case, for Finnish. For the second case, we suggest that the precise Case marker that occurs on an SP, when a Suffix type is involved, may be ultimately determined on semantic grounds (Kracht 2002, 2004). Fourth, we compress the derivation in (26) for pure reasons of space, but we assume that the intermediate steps producing sentence fragments, e.g. Mario on huoneen ulko-... are part of the derivation. Works on sentence production suggest that sentences including incomplete words not only can occur, but give important cues as to "where" speech production errors can occur (Levelt 1989: ch.10; Jarema and Libben 2007: ch.3; Pfau 2009). Our "morphology all the way up" motto seems more justified, given these facts. We now have a unified account of SPs and SCMs, and an analysis of key syntactic properties (demotion); hence, we have a positive answer to the second and third question, respectively.

5 Conclusions

In this paper we have presented a unified analysis of the morphological properties of English and Spanish SPs (*in front, encima*) and Finnish SCMs (*ulkopuolella*). Our analysis is couched in a combination of DM and TL calculi, thus being able to treat morphological (SP/SCM derivation) and syntactic (argument demotion) phenomena under one system. Thus, the analysis offers three positive answers to our empirical questions: first, whether English and Spanish SPs can receive a uniform analysis; second,

whether SCMs fall under this analysis; third, whether the analysis can cover syntactic (demotion) phenomena. However, some problems are still outstanding: for instance, our current analysis cannot (yet) rule out unattested SPs such as **be-mong*, **front in*, **lante-a*, and **puole-lla-l*. A more thorough analysis of feature values that we have left aside, in this paper, would perhaps solve such problems; we leave such analysis for future works, however.

References

- Adger, D 2010. "A minimalist theory of feature structure" in A.Kibort & G.Corbett (eds.), *Features: Perspectives on a Key Notion in Linguistics*. Oxford: Oxford University Press, 185-218.
- Asbury, A 2008. *The morphosyntax of case and adpositions*. Doctoral Dissertation, Utrecht University.
- Asbury, A. J. Dotlačil, B. Gehrke, Ø. Nilsen, & R. Nouwen eds. 2008. *Syntax* and Semantics of Spatial P. Amsterdam: John Benjamins.
- Bauer, L 2005. "Productivity: theories" in P. Stekauer & R. Lieber eds., *Handbook of Word-formation*. Dordrecht: Springer, 315-334.
- Baayen, R. H 1994 Productivity in production. *Language and Cognitive Processes*, 9, 447-469.
- Baayen, R. H 2009 "Corpus linguistics in morphology: morphological productivity" *in* Luedeling, A. Luedeling and M. Kyto eds., *Corpus Linguistics. An international handbook*. Mouton De Gruyter, Berlin, 900-919.
- Bosque, I. 1997 "Preposición tras preposición" in M. Almeida y J. Dorta (eds.), Contribuciones al estudio de la lingüística hispánica, Homenaje al profesor Ramón Trujillo. Tenerife, Editorial Montesinos, 133-155.
- The British National Corpus, version 3 (BNC XML Edition) 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/
- Butt, J. & C. Benjamin 2004. A New Reference Grammar of Modern Spanish, 4th Edition. Hodder Education.

- Caha, P 2009. *The Nanosyntax of Case*. Doctoral Dissertation, University of Tromsø.
- Cappelle, B 2004. "The particularity of particles, or why they are not just 'intransitive' prepositions" in H. Cuyckens, W. De Mulder & T. Mortelmans (Eds.), Adpositions of movement. Belgian Journal of Linguistics (18): 29-57.
- Chomsky, N 1995. *The Minimalist Programme*. Cambridge, MA: the MIT press.
- Cinque, G, & L. Rizzi. eds. 2010. *The Cartography of Syntactic Structures: vol.* 6. Oxford: Oxford University Press.
- Comrie, B 1999. "Spatial cases in Daghestanian languages". *Sprachtypologie und Universalienforschung* 52: 108–117.
- Davies, M 2008. The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at <u>http://corpus.byu.edu/coca/</u>.
- Dikken, M. den 2010. "On the functional structure of locative and directional PPs" in G. Cinque & L. Rizzi. (eds.), *The Cartography of Syntactic Structures: vol. 6.* Oxford: Oxford University Press, 74-126.
 - Emonds, J 1985. *A Unified Theory of Syntactic Categories*. Dordrecht: Foris Publications.
- Embick, D. & R. Noyer 2001. "Movement Operations after Syntax". Linguistic Inquiry 32(4): 555-595.
- Embick, D. & R. Noyer 2006. "Distributed Morphology and the syntaxmorphology interface" in G. Ramchand & C. Reiss (eds.), *The Oxford handbook of linguistic interfaces*. Oxford: Oxford University Press, 289-324.

- Fábregas. A. 2007 "An exhaustive Lexicalisational Account of Directional Resultatives" in M. Basic, M. Pantcheva, M. Son, & P. Svenonius (eds.), *Tromsø Working Papers on Language & Linguistics* 34 (2): 1–32.
- Fillmore, C. 1968 "The Case for Case" in E. Bach & C. Harms (Ed.): Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1-88.
- Fong, V. 1997 *The Order of Things: What Directional Locatives Denote.* Doctoral Dissertation, Stanford.
- Hale, K. & Keyser S. J. 2002. *Prolegomena to a theory of argument structure*. Cambridge, MS: The MIT Press.
- Halle, M. & A. Marantz 1993. "Distributed morphology and the pieces of Inflection" in K. Hale & S. J. Keyser (Eds.), *The View from Building 20*. Cambridge, MA: The MIT Press, 111-176.
- Harbour, D 2007. Morphosemantic Number: From Kiowa Noun Classes to UG Number Features. Dordrecht: Springer.
- Harley, H 2010a. "A Minimalist Approach to Argument Structure". In C. Boeckx, ed., *The Oxford Handbook of Linguistic Minimalism*, 426-447. Oxford: Oxford University Press, 426-447.
- Harley, H 2010b. "Affixation and the mirror principle" in R. Folli & C. Ullbricht eds., *Interfaces in Linguistics*. Oxford: Oxford University Press, 166-186.
- Harley, H 2012. "Semantics in Distributed Morphology" in Maierborn, C. K. von Heusinger, & P. Portner eds., Semantics: An International Handbook of Natural Language Meaning. Amsterdam: De Gruyter, 688-709.
- Hay, J. and Baayen, R. H. (2002) Parsing and productivity. In Booij, G. E. and van Marle, J. (eds), *Yearbook of Morphology 2001*, Kluwer Academic Publishers, Dordrecht, 203-235.

- Huddlestone, R. & G. Pullum 2002. *The Cambridge Encyclopedia of English Grammar*. Cambridge: Cambridge University Press.
- Jackendoff, R 1983. *Semantics and Cognition*. Cambridge, MA: The MIT Press.
- Jackendoff, R 1990. Semantic structures. Cambridge, MA: The MIT Press.
- Jäger, G 2005. *Anaphora And Type Logical Grammar*. Springer: Dordrecht.
- Jarema, G. & G. Libben eds. 2007. The mental lexicon: core perspectives. Amsterdam, Elsevier.
- Karlsson, F 1999. *Finnish: an essential Grammar, 2nd Edition*. New York: Routledge.
- Koopman, H 2000. "Prepositions, postpositions, circumpositions and particles: The structure of Dutch PPs" in H. Koopman (ed.), *The syntax* of specifiers and heads. London: Routledge, 204-260.
- Kracht, M 2002. "On the Semantics of Locatives". *Linguistics & Philosophy* 25 (1): 157-232.
- Kracht, M 2004. "Against the Feature Bundle Theory of Case" in E. Brandner & H. Zinsmeister (eds.): New Perspectives on Case Theory, CSLI, 165 -190.
- Kracht, M 2008. "The Fine Structure of Spatial Expressions" in A. Asbury, J. Dotlačil, B. Gehrke, Ø. Nilsen, & R. Nouwen (eds.): *The Structure of Local P*, John Benjamin's. Amsterdam, 35-62.
- Levelt William J. M 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levinson, S. C 1994. "Vision, shape, and linguistic description: Tzeltal bodypart terminology and object description". *Linguistics* 32(5): 791–855.
- Levinson, S. C. & D. Wilkins eds. 2006. *Grammars of space: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

- Mateu, J 2008. "On the l-syntx of directionality/resultativity: The case of Germanic preverbs" in Asbury, A. J. Dotlačil, B. Gehrke, Ø. Nilsen, & R. Nouwen eds., Syntax and Semantics of Spatial P. Amsterdam: John Benjamins, 221-250.
- Matushansky, O 2002. "On formal identity of Russian prefixes and prepositions" MITWPL 42: 217-253.
- Merchant, J 2001. The syntax of silence. Oxford: Oxford University Press.
- Moortgat, M. J 2010. "Typelogical grammar". In E. N. Zalta (ed.,) *The Stanford Encyclopedia of Philosophy* (Winter 2010 Edition). Stanford.
- Moortgat, M. J 2011. "Categorial Type Logics" *in* J. van Benthem & A. ter Meulen (eds.), *Handbook of Logic and Language. Second Edition*. Amsterdam: Elsevier, 10-97.
- Morrill, G 2011. Categorial Grammar: Logical Syntax, Semantics, Processing. Oxford University Press.
- Niikanne, U 1993. "On Assigning Semantic Cases in Finnish" in A. Holmberg & U. Niikanne eds., *Case and Other Functional Categories in Finnish Syntax*. Mouton de Gruyter, Berlin, 75-87.
- Noonan, M 2008. "Case compounding in the Bodic languages" in G. Corbett & M. Noonan (eds.), *Case and Grammatical Relations. Studies in honor of Bernard Comrie*. John Benjamins, Amsterdam, 127-148.
- Oshima, A. & Hogue, A 2006. *Writing academic English 4th Edition*. New York: Addison-Wesley Long Publishing Company, Inc.
- Pavón, M. V 1999. "Clases de partículas: preposición, conjunción adverbio" in I. Bosque & V. Demonte (eds.), *Gramática descriptiva de la lengua española*. Espasa, Madrid, 565-655.
- Pfau, R 2009. Grammar as processor: A Distributed Morphology account of spontaneous speech errors. Amsterdam: Benjamins.

- Plag, I. 2006. "Productivity" in B. Aarts & A. McMahon (eds.), *Handbook* of English Linguistics. Oxford, Blackwell, 537-556.
- Phillips, C 2006. "The real-time status of island phenomena" *Language* 82 (5): 795-823.
- Riemsdijk, H. van 1978. *A Case Study in Syntactic Markedness*. Dordrecht: Foris Publications.
- Riemsdijk, H. van 1990. "Functional Prepositions" in H. Pinkster & I. Genée (eds.), Unity in Diversity. Festschrift for Simon Dik. Dordrecht: Foris Publications, 229-241.
- Riemsdijk, H. van & R. Huybregts 2007. "Location and locality" in M. van Oostendorp & E. Anagnostopoulou, (eds.), Progress in Grammar: Articles at the 20th Anniversary of the Comparison of Grammatical Models Group in Tilburg. Meertens Instituut, Amsterdam, 123-140.
- Stabler, E 2013. "Two models of minimalist, incremental syntactic analysis". *Topics in Cognitive Science* 5(3): 611-633.
- Svenonius, P 2003. "Limits on P: Falling in holes vs. falling in holes" in Peter Svenonius (ed.), Tromsø Working Papers on Language & Linguistics: Proceedings of the 19th Scandinavian Conference of Linguistics 31(2): 431-445.
- Svenonius, Peter. 2004b. Slavic prefixes inside and outside VP. In *Nordlyd*, *Tromsø Working Papers on Language and Linguistics* 32(2): 205-253. Available at www.ub.uit.no/munin/nordlyd/.
- Svenonius, P 2006. "The emergence of axial parts". Nordlyd, Tromsø University Working Papers in Language and Linguistics 33(1): 49-77.
- Svenonius, P 2007. "Adpositions, Particles, and the Arguments they Introduce" in E. Reuland, T. Bhattacharya, & G. Spathas (eds.), Argument Structure. John Benjamins, Amsterdam, 63-103.
- Svenonius, P. 2008. "Projections of P" in A. Asbury, J. Dotlačil, B. Gehrke,

Ø. Nilsen, & R. Nouwen (eds.), *The Syntax and Semantics of Spatial P.* John Benjamins, Amsterdam, 63-84.

- Svenonius, P 2010. "Spatial P in English" in G. Cinque & L. Rizzi. eds., The Cartography of Syntactic Structures: vol. 6. Oxford: Oxford University Press, 161-194.
- Talmy, L 2000. *Towards a Cognitive Semantics*. Cambridge, MA: the MIT Press.
- Thomas, E 2001. "On the expression of directional movement in English". *Essex Graduate Papers in Language and Linguistics* 4: 87-124.
- Thomas, E 2004. "Syntactic' vs. 'Semantic' Telicity: IN and ON' in H. Cuyckens, W. De Mulder & T. Mortelmans (eds.), Adpositions of movement. Belgian Journal of Linguistics, 18: 145-166.
- Tseng, J 2000. *The representation and selection of prepositions*. Edinburgh: Ph.D. dissertation.
- Tseng, J 2004. "Directionality and the complementation of Dutch prepositions" in H. Cuyckens, W. De Mulder & T. Mortelmans (eds.), Adpositions of movement. Belgian Journal of Linguistics, 18: 167-194.
- Tseng, J 2005. "Prepositions and Complement Selection" in A. Villavicencio & V. Kordoni (eds.), Proceedings of the 2nd ACL-SIGSEM Workshop on the linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications. Colchester: Essex University, 11-19.
- Ursini, F-A 2013a. "On the Syntax and Semantics of Spatial Ps in Spanish." *Borealis: An international journal about Hispanic Linguistics* 2(1): 117-166.
- Ursini, F-A. 2013b. "On The Syntax and Semantics of "Tener" and "Haber." *Lingue & Linguaggio* 11 (1): 89-120.

- Ursini, F.-A. 2015. "On The Syntax and Semantics of Italian Spatial Prepositions." *Acta Linguistica Hungarica* 62(1): 3-57.
- Ursini, F.-A. & Akagi, N. 2013a. "On the distributed morphology and semantics of spatial Ps." In I.-J. Lee and U. Dolgormaa (eds.), Proceedings of the 15th Seoul International Conference on Generative Grammar (SICOGG 15). Hankuk University Press, Seoul, 447-468.
- Ursini, F-A. & Akagi. N 2013b. "Another look at modification in Spatial Prepositions." *Iberia* 5(2): 37-84.
- Wunderlich, D 1991. "How do prepositional phrases fit into compositional syntax and semantics?". *Linguistics* 29 (5): 591-621.