University of Patras

Univeriteit Leiden

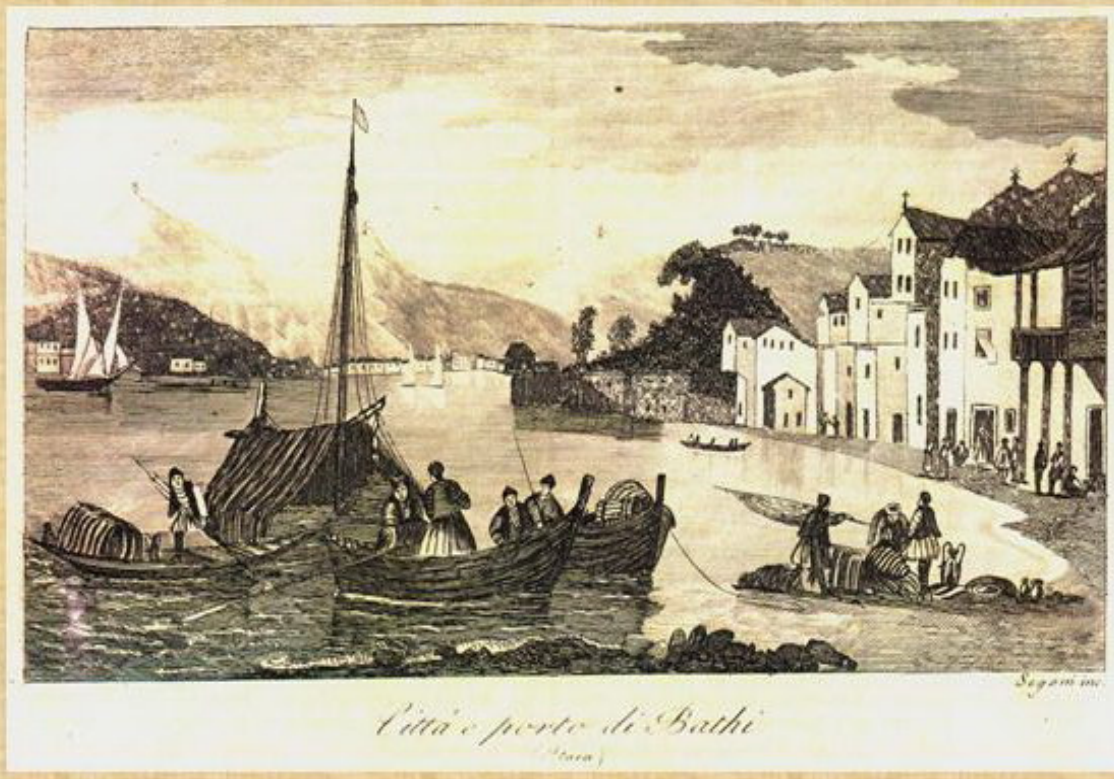
Universita degli Studi di Bologna

PROCEEDINGS

6th Mediterranean Morphology Meeting

Ithaca, Greece

27– 30 September 2007



Editors:

Geert Booij (Universiteit Leiden)

Angela Ralli (University of Patras)

Sergio Scalise (Università degli Studi di Bologna)

Morphology and Dialectology

On-line Proceedings of the
Sixth Mediterranean Morphology Meeting (MMM6)
Ithaca, 27-30 September 2007

Edited by:
Angela Ralli, Geert Booij, Sergio Scalise and Athanasios Karasimos

URL: http://morbo.lingue.unibo.it/mmm

2008 University of Patras, Greece
On-line Proceedings of the Sixth Mediterranean Morphology Meeting (MMM6)
ISSN: 1826-7491

Contents

Foreword

The Mediterranean Morphology Meetings (MMM) are organized by Prof. Geert Booij (Leiden), Prof. Angela Ralli (Patras), and Prof. Sergio Scalise (Bologna). For each meeting they cooperate with a local organizer.

The aim of MMM is to bring together linguists who work on morphology in an informal setting, which guarantees maximal interaction between researchers, and gives young linguists the chance to present their work at a conference of moderate size, where fruitful contacts with senior linguists can be established.

The first six meetings, in 1997 (Mytilene, Greece), 1999 (Lija, Malta), 2001 (Barcelona), 2003 (Catania, Sicily), 2005 (Fréjus, France), 2007 (Ithaca, Greece) have proven the success of this formula. The attendance was high, many abstracts were submitted, and a number of leading morphologists were invited.

Each MMM has two sessions, one with a specific topic, and another with no topic. The specific topic of the Ithaca meeting was 'Morphology versus Dialectology'. Many of the papers of this session are under review for publication in the Journal *Morphology* (Springer). These electronic proceedings (ISSN: 1826-7491) include papers from the free-topic session.

The permanent scientific committee
Geert Booij
Angela Ralli
Sergio Scalise

Morphology meets Dialectology: insights from Modern Greek dialects[1]

Angela Ralli
University of Patras
ralli@upatras.gr

0. Abstract

The focus of this paper is to address a fundamental question regarding the relation between morphology and dialectology. On the one hand, it shows that the study of dialects offers new challenges to morphology, since dialects represent an important source of morphological phenomena. Dialectal research allows us to throw light on morphological theoretical issues, and establishes robust theoretical proposals. On the other hand, from a theoretical morphological perspective, it argues that morphological theory may provide accurate and interesting descriptions to dialectal phenomena because theoretical analyses can offer a repertoire of tools and concepts, which can be adapted within a dialectological framework, help systematize the research object, and refine the traditional description of dialects.

To these two goals, one should not neglect the typological and historical dimensions: a closer look at dialectal morphology can be profitable from the point of view of typology and historical morphology, since the study of dialects may offer additional insights to the discussion on possible language structures and language change. For instance, **a** number of grammatical features and patterns of variation that can be found in non-standard varieties are not part of the relevant standard linguistic systems. These patterns may be new, or old, as features of previous language periods may still exist alongside with new ones. In this sense, dialects portray ongoing linguistic change.

The data which illustrate views and proposals include examples from several Modern Greek dialects, and are extracted from grammars of individual dialects, as well as from the oral corpora of the *Laboratory of Modern Greek Dialects* of the University of Patras.

1. Morphology meets dialectology: general remarks

In morphological research, dialects have been accounted for only sporadically and rather unsystematically, since modern morphological theory is largely oriented towards the standard form of languages. As a result, a considerable number of interesting phenomena in spoken dialects of various languages are left without any consideration, and thus overlooked in morphological studies. As Anderwald and Kortmann (2002: 160) point out, the limitation to standard varieties is problematic, especially in languages with a long literary tradition, where the setting of norms has always played an important role, and certain features do not reflect the natural change, but rather more or less arbitrary changes, which are imposed by various prescriptivists. This view mirrors the situation in Modern Greek (hereafter MG), as the standard language (hereafter SMG) sometimes gives a false picture of what the grammar of the language is like.

In this paper, I show that a closer look at dialectal morphology is profitable from several points of view.

First, research from the dialectal domain is crucial from the theoretical point of view: it shows that dialectology helps refine morphological approaches, since dialectal phenomena offer a rich testing ground for morphological theoretical claims and proposals. In this study, I add to the discussion about crucial morphological issues, such as the following:

[1] I wish to thank Mark Janse for his precious help on the Cappadocian data, and the audience of MMM6 for their most helpful comments.

- The inventory of morphological entities/primitives,
- The set of morphological phenomena.
- The set of constraints.
- The key notions of productivity and morphological creativity.

Second, the application of theoretical morphological approaches to dialectal phenomena may also provide accurate and interesting analyses to these phenomena, since theoretical proposals add a new dimension to the old description of dialects.

Third, dialectal evidence may offer additional insights to the discussion about linguistic change and typology, i.e. it can shed light on how a grammar of a particular language may look like, and what are its structural limits. For instance, a number of grammatical features and morphological structures that are detected in several MG dialects are not part of SMG. In this sense, dialects constitute a rich source of information on the grammatical possibilities of MG as well as on its historical evolution.

Research from the domain of MG dialects offers several interesting cases that could illustrate the points above. Most particularly, I deal, first, with the issues of morphological phenomena and constraints. By showing the important role of dialectal evidence, I aim to contribute to the debate about the structure of grammar, and morphology being an independent grammatical module. To this purpose, I examine the rise of an innovative pattern of compounds during the late medieval period (14th c. AD), namely the set of coordinative verbal compounds, as well as the crucial role of stem allomorphy in the organization of inflectional paradigms. Second, I add to the discussion about the demarcation of morphological entities and word-formation processes by providing evidence about a borderline case, which is situated between prefixation and compounding. I show that this case could also contribute to the discussion about the notions of morphological productivity and creativity. Finally, with the use of evidence drawn from Cappadocian, a peripheral MG dialect, I question the structural limits of a linguistic system with respect to change; namely, I demonstrate how a linguistic system may shift from one typological pattern to another, as a result of language-contact situations.

2. Morphologically proper constraints

If morphology is different from syntax or phonology, and forms a module of its own, it should display entities and phenomena that are different from those of the other modules. In this section, I examine the postulation of a constraint, which seems to be proper to morphology, since it applies to stems, that is to constituents playing a significant role in morphological structures, especially in fusional languages, such as Greek, where inflected words are made up of stems and inflectional endings. To this purpose, I use evidence from the set of coordinative verbal compounds, which are particularly developed in MG dialects.

As pointed out by Ralli (2007, to appear, in preparation.a), compounding is a rich word-formation process, and very productive in MG. There are compounds of all types and categories, the most peculiar of which are those of the coordinative verbal type, as these formations are not usually found in the other Indo-European languages, and did not characterize Classical Greek (5th-4th c. BC).[2] Like other Greek compounds, they constitute phonological words (i.e. they bear one stress), have a stem as their first member, and a compound marker/linking element *–o-* between the first and the second constituents.[3] Semantically, they combine two verbal stems of compatible or opposite meanings. Typical examples of these compounds are the following:

(1)a. aniγoklino < aniγ- klino
to open and close open close

[2] However, according to Andriotis (1957), there is a single Ancient Greek (hereafter AG) example: *strephedineomai* 'twist' attested in the Homeric poems.

[3] Note that according to Ralli (2008a), the linking element *–o-* between the first and the second constituent denotes the process of compounding, and may be considered as a 'compound marker. See Ralli (2008a) for additional details.

b. benovjeno < ben- vjeno
to enter and go out enter go out
c. pijenoerxome < pijen- erxome
to come and go go come
d. troγopino < troγ- pino
to eat and drink eat drink
etc.
(from Ralli 2007)

Following the Greek historical grammars (Jannaris 1897, Hatzidakis 1905-7), verbal dvandva compounds can be traced back at the end of the Hellenistic period (around the 2nd c. AD), with the occurrence of the verb *afksomio* (Ptolem. Synt. Math. 6,7, as cited by Andriotis 1957: 44), which contains a portion of the AG verb *auksano:* 'increase, raise' and the verb *meio:* 'reduce'.[4] The number of verbal coordinative compounds has increased in subsequent periods, especially during the Medieval period, as attested by several vernacular texts (see Manolessou & Tsolakidis 2007, Joseph & Nichols 2007, Ralli 2008b). Today, a small number of verbal coordinative compounds can be found in SMG (as shown in (1)), but they do not belong to the most productive compound patterns. On the contrary, they are extremely productive in several MG dialects, where they occur in massive numbers. Andriotis (1957) offers numerous examples taken from the range of MG dialects, even from the most peripheral ones, such as Pontic, Cappadocian, and Cypriot:[5]

(2)a. Pontic: lambovrexi < lamb- vrexi
shines and rains shines rains
c. Cappadocian: maramuðjazu < mara- muðjazo
to fade and become numb fade become numb
d. Cypriot: skalopotizo < skal(iz) potizo
to grub and water grub water

Dialectal evidence is precious with respect to the analysis of these compounds: it helps us to draw conclusions about their structure, namely significant insights about the presence of derivational suffixes within compounds, and consequently, about the interaction between compounding and derivation.

On the basis of SMG compounding, Karasimos & Ralli (2007) have observed that, with some exceptions, the first constituent of the constructions under consideration is not a derived item. The MG dialects do not only corroborate this observation, but help us to formulate a plausible hypothesis about a possible account of it. A considerable number of dialectal examples show that derivative verbal stems participating as first constituents of coordinative compound structures are stripped of their derivational suffixes. As an illustration, consider the following cases that are taken from several geographical areas, as cited by Andriotis (1957).[6]

[4] As is widely known, the vocalic and consonantal systems of Greek have undergone a significant change during the Hellenistic period. Therefore, examples of Ancient Greek will be transcribed according to the Classical Greek pronunciation, while examples of the Hellenistic period (approximately 3rd c. BC – 3rd c. AD) as well as those of Medieval and Modern Greek will be given a Modern Greek transcription.

[5] All three dialects were (or are) spoken outside the Greek mainland and the islands. Cypriot is found in Cyprus. Pontic was spoken in the area of Pontus, in North-East Turkey. Today, it is still spoken by an unknown number of Pontic Muslims who still live in this area (see MacKridge 1987), as well as by Pontic refugees who settled in Greece after the end of the war between Greece and Turkey in 1922.

[6] According to Ralli (2008a), the linking element *–o-* which appears between the first and the second constituent is not a structural part of any of the two constituents. As such, it is not a derivational suffix, and appears in this position in order to denote the process of compounding.

(3)a. Crete:	alon-o-θerizo[7]	< alon-iz-	θerizo	
	to thresh and reap	thresh	reap	
b. Symi:	abar-o-kliδono	< abar-on-	kliδono	
	to padlock	bar	lock	
c. Karpathos:	aθθ-o-luluδizo	< aθθ-iz-	luluδizo	
	to blossom and bloom	blossom	bloom	
d. Rhodes:	imer-o-γalinizo	< imer-on-	γalinizo	
	to tame and calm down	tame	calm down	
e. Kefalonia:	klaδ-o-kaθarizo	< klaδev-	kaθarizo	
	to prune and clean	prune	clean	
f. Euboea:	vraδjaz-o-ksimeronome	< vraδj-az-		ksimeronome
	to spend all time	be overtaken by night		spend all night
g. Epeiros :	zim-o-majirevo	< zim-on-	majirevo	
	to knead and cook	knead	cook	
h. Lesbos:	kukl-u-stsipazo	< kukl-on-	stsipazu	
	to wrap up and cover	wrap up	cover	
i. Imbros:	majir-u-kinonu	< majir-ev-	kinonu	
	to cook and pour	cook	pour	

Note that in (3), the first constituent behaves like a verb: it belongs to the verbal category and has the meaning and the argument structure of a verb, although it is deprived of its derivational suffix, and superficially looks like a noun stem, where it derives from. On the basis of this evidence, and in order to account for the impossibility of derivational suffixes to surface within compounds, Ralli & Karasimos (2007) have proposed that Greek compounds undergo a morphological constraint, which is responsible for the deletion of the derivational suffix, the so-called *Bare stem constraint*, according to which stems appearing as fist constituents of compound words must be as bare as possible. This constraint is motivated by the fact that Greek compounds are basically [stem stem] or [stem word] structures,[8] where the first constituent, being a stem, owes to be in a tied relation with the second constituent, i.e. with another stem or a word, depending on the case. Ralli & Karasimos have assumed that the strong bond between the two constituents is better ensured if the first stem is bare, in other words if it is deprived of any additive elements, i.e. suffixes. Note that evidence for suffix deletion from compound-internal stems should be stronger in coordinative compounds than in any other compounds, which bear a dependency relation (subordinative or attributive) between their constituent parts: by nature, coordinative structures display a weaker cohesion between their members, than the ones which are not coordinative. As a result, the fact that a rather loose structure does not allow the presence of compound-internal derivational suffixes adds to the postulation of the bare-stem constraint a particularly robust support.

In addition to the importance of dialectal coordinative verbal compounds for the existence of morphologically proper constraints, this evidence can also be useful as an illustration on how change and innovative structures may occur and affect a linguistic system. In fact, while SMG has been developed mostly in the last two centuries, the MG dialects constitute variants which arose from the Hellenistic Koine (approximately 3rd c. BC – 3rd c. AD). Therefore, it is only through dialects that one could detect the rise of the pattern of coordinative verbal compounds in MG, which was rather absent in AG.

3. Stem allomorphy in the organization of paradigms

In what follows, I try to provide substantial proof for the existence of a morphologically proper phenomenon and its contribution to the organization of grammar, as already pointed

[7] See footnote 3 about the compound marker *–o-*. Due to mid-vowel raising in unstressed position this *–o-* becomes /u/ in the Northern Dialects, among which, the ones of Lesbos and Imbros. For clarity reasons the constituents are separated by a hyphen, which does not appear in written Greek.

[8] For details on the structure of Greek compounds, see Ralli (1988, 1992, 2005, 2007, in preparation.a), Drachman & Malikouti-Drachman (1994), Nespor & Ralli (1996), and Malikouti-Drachman (1997).

out by Booij (1997ab). Most particularly, on the basis of evidence from SMG and three dialectal varieties, namely Lesbian, Aivaliot and Moschonisiot (hereafter LAM), I deal with stem allomorphy, and its role in the determination of inflectional paradigms.[9] I show that allomorphy is not a simple synchronic residue of historical processes, but may assume a crucial classificatory role, which leads to the distinction of inflection classes. In this respect, allomorphy cannot be seen as a simple deviation from form uniformity, but as a central morphological property which constraints paradigms, paradigm organization, and paradigm restructuring.[10]

It is well known that nouns and verbs of fusional languages are distributed in inflection classes, the classification of which is based on certain specific criteria. In SMG, verbs are inflected according to two major inflection classes, each class bearing its own inflectional endings in at least two paradigms, those of the present and the imperfect tense. According to Ralli (1988, 2005, 2006), this classification is based on the systematic presence, or absence, of a specific allomorphy pattern, which characterizes the stems, and has assumed the role of an inflection-class demarcator, in the sense that verbs that do not adapt to it are predicted to inflect differently from verbs that have it.[11] This pattern is described as X(a) ~ Xi, where the X(a) form characterizes paradigms of an imperfective aspect (present, imperfect and future continuous), while the Xi form is typical of paradigms of the perfective value (aorist and simple future). Verbs whose paradigms are submitted to this pattern belong to inflection-class II. Verbs which do not undergo this pattern inflect according to inflection-class I. As an illustration, compare the inflection of the SMG class-I verb *γrafo* 'to write' (5) with that of the SMG class-II verb *aγapo* 'to love'(4):

(4) SMG: Stem allomorphs: aγapa ~ aγapi

	a. Present	b. Imperfect	c. Aorist
SG 1P	aγa'p(a)-o	a'γapa-γ-a / aγa'p-us-a[12]	a'γapi-s-a[13]
2P	aγa'pa-s	a'γapa-j-es / aγa'p-us-es	a'γapi-s-es
3P	aγa'pa-i	a'γapa-j-e / aγa'p-us-e	a'γapi-s-e
PL 1P	aγa'pa-me	aγa'pa-γ-ame / aγa'p-us-ame	aγa'pi-s-ame
2P	aγa'pa-te	aγa'pa-γ-ate / aγa'p-us-ate	aγa'pi-s-ate
3P	aγa'pa-ne	a'γapa-γ-an / aγa'p-us-an	a'γapi-s-an

(5) SMG: absence of stem allomorphs

	a. Present	b. Imperfect	c. Aorist
SG 1P	'γraf-o	'e-γraf-a[14]	'e-γrap-s-a

[9] Lesbian is spoken on the island of Lesbos, while the Asia-Minor dialectal varieties Aivaliot and Moschonisiot were spoken once (before 1922) in the North-west coast of Turkey (former Asia Minor), namely in the areas of Aivali (today Ayvalik) and Moschonisi (today Cunda). The last two varieties are still spoken by second and third generation refugees in certain Asia-Minor dialectal enclaves of the island of Lesbos (cf. Ralli in preparationb).

[10] For the close relation between allomorphy and paradigmatic morphology, see also Booij (1997ab), and Maiden (1992, 2003).

[11] This suggestion is in accordance with Maiden (1992) who has shown that allomorphy patterns are very robust in paradigms, on the basis of evidence drawn from Italian. Moreover, as already noted by Ralli (2006), the X(a)~Xi pattern functions like a 'schema', in a broader sense of what is defined as a schema by Bybee & Slobin (1982), since it determines the paradigmatic behavior of a class of verbs, the members of which form a series of 'family' inflectional resemblances.

[12] *-γ-* or *–us-* are markers of an imperfective aspectual value, which are rather free variants in SMG. They characterize only verbs belonging to Class-II, since verbs of Class-I do not display this marker. In 2P and 3P of the singular, - *γ* – is palatalized before a front vowel.

[13] *-s-* marks the perfective aspectual value in most verbs of both inflection classes.

	2P	'γraf-is	'e-γraf-es	'e-γrap-s-es
	3P	'γraf-i	'e-γraf-e	'e-γrap-s-e
PL	1P	'γraf-ume	'γraf-ame	'γrap-s-ame
	2P	'γraf-ete	'γraf-ate	'γrap-s-ate
	3P	'γraf-un	'e-γraf-an	'e-γrap-s-an

Ralli's proposal for the role of allomorphy as an inflection-class demarcator finds crucial support in the dialectal domain. Consider the evidence from LAM in (6) and (7) below, where the paradigms of present, imperfect, and aorist, of the verbs *'γrafo* and *aγa'p(a)o,* show the presence of the same systematic allomorphy pattern X(a) ~ Xi, as in SMG**.**[15]

(6) LAM underlying stem forms: X(a) ~ Xi (aγap(a) ~ aγapi)				
		a. Present	b. Imperfect	c. Aorist (underlying aγapi-)
SG	1P	aγa'p-o	a'γap-um/a'γap-umna[16]	a'γap-s-a
	2P	aγa'pa-s	a'γapa-s	a'γap-s-is
	3P	aγa'pa	a'γapa	a'γap-s-i
PL	1P	aγa'p-umi	aγa'p-us-ami	aγa'pi-s-ami
	2P	aγa'p-uti	aγa'p-us-ati	aγa'pi-s-ati
	3P	aγa'p-un	aγa'p-us-an	aγa'pi-s-an

(7)		a. Present	b. Imperfect	c. Aorist
SG	1P	'γraf-u	'e-γraf-a	'e-γrap-s-a
	2P	'γraf-s	'e-γraf-is	'e-γrap-s-is
	3P	'γraf	'e-γraf-i	'e-γrap-s-i
PL	1P	'γraf-umi	'γraf-ami	'γrap-s-ami
	2P	'γraf-iti	'γraf-ati	'γrap-s-ati
	3P	'γraf-in	'γraf-an	'γrap-s-an

It is important to note that the dialects, LAM in particular, provide additional and substantial proof to the general classificatory role of stem allomorphy: they show that the class-II pattern has been spread to a number of irregular verbs as well, the old stems of which displayed an allomorphic variation, but did not conform to the systematic X(a) ~ Xi pattern. As an example, consider the AG verb *sbennymi* 'to extinguish', which in SMG appears as *zvino* (8), while in LAM as *zvo* (9):

(8) SMG: stem allomorphs zvin ~zv(i)				
		a. Present	b. Imperfect	c. Aorist
SG	1P	'zvin-o	'e-zvin-a[17]	'e-zvi-s-a
	2P	'zvin-is	'e-zvin-es	'e-zvi-s-es
	3P	'zvin-i	'e-zvin-e	'e-zvi-s-e

[14] The *e-* preceding the verbal stem is the augment. It appears in front of stems beginning by a consonant and is a stress carrier in the past tenses (imperfect and aorist). See Babiniotis (1972), Kaisse (1982), Ralli (1988), and Drachman & Malikouti-Drachman (2001) for more details on this.

[15] According to the LAM phonology unstressed high vowels /u/ and /i/ are deleted, and mid-vowels /e/ and /o/ become /i/ and /u/ respectively (see also footnote 7). For instance, underlying *'γraf-i* and *'eγrap-s-es* become *'γraf* and *'eγrap-s-is.* These phonological phenomena are present in the larger group of Northern Greek Dialects, members of which are Lesbian, Aivaliot and Moschonisiot.

[16] *A'γapum* is the form used in Lesvos, while *a'γapumna* is the one used in Aivaliot and Moschonisiot.

[17] The *e-* preceding the verbal stem is the augment. It appears in front of stems beginning by a consonant and is a stress carrier in the past tenses (imperfect and aorist). See Babiniotis (1972), Kaisse (1982), Ralli (1988), and Drachman & Malikouti-Drachman (2001) for more details.

PL	1P	'zvin-ume	'zvin-ame	'zvi-s-ame
	2P	'zvin-ete	'zvin-ate	'zvi-s-ate
	3P	'zvin-un	'e-zvin-an	'e-zvi-s-an
(9) LAM underlying stem forms: X(a) ~ Xi (zv(a) ~ zvi)				
		a. Present	b. Imperfect	c. Aorist (underlying zvi-)
SG	1P	zv-o	'e-zv-um/'e-zv-umna[18]	'zuf-s-a[19]
	2P	zva-s	'e-zva-s	'zuf-s-is
	3P	zva	'e-zva	'zuf-s-i
PL	1P	zv-umi	'zv-us-ami	'zvi-s-ami
	2P	zv-uti	'zv-us-ati	'zvi-s-ati
	3P	zv-un	'zv-us-an	'zvi-s-an

We see in (8) that the SMG verb belongs to class-I, because it does not display the systematic X(a) ~ Xi pattern. Its unsystematic stem allomorphy *zvin ~ zv(i)* is proper to this verb, and does not characterize a larger group of verbs. On the contrary, the LAM verb stem in (9) has undergone restructuring, according to the pattern X(a) ~ Xi (*zv(a) ~ zvi*), which made it conform to the class-II paradigms. Following Kuryłowicz (1949), we could claim that a process of analogy has occurred in LAM in order to establish a central contrast of the language, i.e. the presence or absence of the X(a) ~ Xi allomorphy pattern, which is used as an inflection-class demarcator, and replaces a more marginal allomorphy pattern. In fact, LAM shows an extensive use of this pattern for a considerable number of verbs, which, in their SMG instances, display various unsystematic stem-allomorphy patterns, and as such belong to class-I.[20] The dialectal levelling of various irregular verb stems according to the X(a) ~ Xi pattern may be considered as an optimization of the verb system on the level of lexical representations. In Kiparsky's (2003) terms, this levelling removes the irregular allomorphic variants from certain verbs, establishes a uniform stem-allomorphy pattern, and optimizes lexical representations by increasing their conformity to the system.

Finally, the dialectal evidence provided above confirms Booij's (1997ab) hypothesis about the significant contribution of allomorphy to inflectional paradigmatic structure, and provides support to the thesis for the autonomy of morphology, as a grammatical domain with its own phenomena, non-phonologically conditioned stem allomorphy being one of them.

4. On a borderline case between morphological categories

In this part of the paper, I deal with items, the structural status of which is unclear, and as such have always been a problem for morphological theory, in synchronic terms, since they cannot be classified into one particular category, and the processes into which they participate cannot be adequately delimited. Namely, I look at a dialectal phenomenon, which is at the border between prefixation and compounding, and can be detected in a small number of LAM adverbial formations beginning by *sa-* (10). Beside the useful conclusions which we

[18] *A'γapum* is the form used in Lesvos, while *a'γapumna* is the one used in Aivaliot and Moschonisiot.
[19] /u/ in (9c) is an epenthetic vowel which appears between /z/ and /v/ (becoming /f/ in front of /s/) in order to make the three-consonant cluster /zfs/ easier to pronounce.
[20] Some of these examples are the following:

(i)	SMG	LAM
Verb	Stem allomorphs	Stem allomorphs
klino 'to close'	klin ~ kli	kl(a) ~ kli
ftino 'to spit'	ftin ~ fti	ft(a) ~ fti
arosteno 'to make/become sick'	arosten ~ arosti	arust(a) ~ arusti

can draw about the borderline cases between morphological categories and word-formation processes, this phenomenon portrays a typical example of an ongoing change, and provides significant insights about the distinction between two important notions in morphology, productivity and creativity.

Consider the following examples, which are taken from Ralli & Dimela (2007):[21]

(10)a. sapéra 'far away' < sa- péra 'away'
b. saδó 'over here' < sa- iδó 'here'
c. sáδju 'over here' < sa- éδju 'here'
d. sáδuna 'over here' < sa- éδuna 'here'
e. sací 'over there' < sa- icí 'there'
f. sácina 'over there' < sa- écina 'there'
g. sáftu 'ofter there' < sa- éftu 'there'
h. sáfna 'over there' < sa- éfna 'there'
j. sakátu 'straight down there' < sa- kátu 'down'
i. sapánu 'straight up there' < sa- apánu 'above'
k. samésa 'more inside' < sa- mésa 'inside'
but
l. *sáksu 'more outside' < sa- óksu 'outside'

These adverbs contain an adverbial word, which is preceded by a bound element *sa-*. The latter originates from the autonomous directional adverb *ísa* 'straight' (*ísja* in SMG), and functions as an intensifier of the locative meaning of the base.

Like in SMG (11a), *isa* in LAM can also be used as an autonomous directional adverb, modifying verbs (11b), but not locative adverbs, since this function has been replaced by its reduced form:

(11)a. SMG
kops-to isja
cut it straight
pijene isja kato
go straight down there

b. LAM
kops-tu isa

pani sakatu vs. *pani isa katu

-sa has undergone a phonological attrition with an initial /i/ deletion and the internal loss of the semi-vowel /j/. As shown by Ralli & Dimela (2007), in the case of *sa*-adverbs, phonological attrition cannot constitute a safe criterion for assigning to *–sa* a prefixal status, since both phonological changes are due to general phonological laws, which apply to LAM independently of the particular morphological environment of *sa*-formations: unstressed /i/ is deleted at the beginning of words, and /i/ is reduced into /j/ in word-internal contexts, between a /s/ and a vowel (see Newton 1972). More importantly, the appearance of *–sa* in morphologically complex adverbs is of limited productivity, since it is restricted to a handful of examples containing a locative adverb, and it does not combine with all locative adverbs, as shown by the ungrammatical example of **saksu* in (10.l). This particular selection opposes *sa-* to Greek prefixes, which are more or less category neutral, in that they do not impose any special selectional restrictions on the base. Thus, the status of *sa-* as a prefix is doubtful.

On the basis of these properties, Ralli & Dimela (2007) have proposed that an element such as *sa-* still keeps its lexeme status. As a result, its combining with locative adverbs could be seen as an instance of compounding. In fact, *sa-* with its full adverbial form *isa,* also appears at the right-hand position of adverbial compounds, as for instance, in the

[21] Words in (10) are given in their dialectal form. For the deletion of unstressed /i/ and /u/ and the mid-vowel raising, also in unstressed position, see footnote 15.

formation of (12a) or can accept the verbal derivational suffix *–az(u)*[22] in order to form the verb *sazu* 'to put order':

(12)a. uloisa < ulu isa[23]
all straight all straight
b. sazu < isa -az- -u
to put order straight DER INFL

However, the 'compounding' hypothesis runs against the fact that *sa-* in (10) selects locative adverbs, because selectional restrictions do not usually characterize compounds (but see Bisetto, Guevara & Scalise for an opposite view). Moreover, the meaning of *isa* as a second member of compounds (like in the example of 12a), or as the base of a derived word (12b) is not reduced into the general intensifying function displayed by *sa-* in (10).

Since there is no sufficient semantic or formal justification of the hypothesis that *sa-* is a lexeme, or of the one that it has been morphologized into a prefix, Ralli & Dimela (2007) have proposed that *sa-* is under the process of becoming a prefix, but has not acquired the full prefixal status yet. In other words, they have argued that although *sa-* does not have all the properties of a real prefix, and there is no guarantee that it will result into being one, there are serious indications (e.g. form reduction and reduced meaning) of a morphologization in progress.[24]

In languages, it is easy to find examples that appear to be at various points of a potential diachronic development, that is items that are in the process of losing their word independence (see also Bauer 2005: 98). Since the categorial status of these items is not clear, and the processes into which they participate are not well delimited, I propose to appeal to the notion of *morphological creativity* in an effort to provide an analysis to the problems raised by the peculiar behavior of *sa-*. According to Schultink (1961) and Lieber (1992), morphological creativity is the process under which there is a conscious coinage of a new word, as opposed to *morphological productivity*, which involves words that are unintentionally created (Bauer 1983, 2001, Plag 1999). Extending the notion of morphological creativity, Baeskow (2004: 78) assumes that it can also include phenomena involving a superficial reinterpretation of items, which may be done for specific purposes, but without any real change of their inherent categorial status. Adopting this broadened view of morphological creativity, I would like to suggest that it can account for the properties of the LAM *sa*-formations, from the synchronic point of view: As already pointed out, there are properties which list them as prefixation, and properties that make them similar to compounding. In the examples of (10), *sa-* seems to behave like a prefix in a specific context, i.e. when it is combined with the majority of locative adverbs, and may be reinterpreted as such. However, I suggest that this reinterpretation is only superficial, since the full form *isa,* from which *sa-* is derived for independent phonological reasons, still keeps its lexeme status as far as its lexical entry is concerned. In other words, *isa-* functions as a prefix in the particular context of its combination with the locative adverbs (see examples in (10)), but does not have undergone a radical category change from lexeme into prefix.

Finally, elaborating on the phenomenon of the ambiguous status, i.e. prefixal or lexematic, of *sa-*, and by taking into consideration data from other MG dialectal systems, I would like to suggest the importance of dialectal evidence into showing ongoing linguistic changes. We have already noticed that the directional adverb *isja* in SMG is not affected by any change with respect to its word status. The same seems to apply to the dialectal systems of the island of Corfu (Corfiot) and Peloponnese. On the contrary, *isja* seems to have become a full prefix in the dialect spoken on the island of Crete. As noted by Dimela (2005), in

[22] The *-u* in parenthesis is the inflectional ending of 1P singular.
[23] In this example, there is no need for /i/ deletion, since /i/ is not in initial position.
[24] For details about morphologization and its difference from grammaticalization, see Joseph (2003).

Cretan, a corresponding to *sa-* element, *s(j)o-*, is found prefixed to several categories, i.e. to verbs (13a), adjectives (13b), and adverbs (13c).[25]

(13) Cretan

a. sojerno	<	so-	jerno
'become very old'			'to become old'
b. soaspros	<	so-	aspros
'very white'			'white'
c. sodreta	<	so-	dreta
very straight			straight

As seen in (13), and as opposed to the LAM *sa-*, the Cretan *–s(j)o* is attested in a wider context, where it is extremely productive. As a proof of its productivity, we find *–s(j)o-* to be used in the creation of neologisms which cannot be detected in the most updated Cretan dictionaries (e.g. Idomeneas 2006, Garefalakis 2002, and Ksanthinakis 2000), while neology with the use of *sa-* is not generally possible in LAM. An additional argument in favor of the Cretan *s(j)o-* being a prefix comes also from the fact that, on synchronic grounds, native speakers make no link between its initial lexical meaning and the actual intensifying function. On the contrary, they often mix up *s(j)o* originating from 'straight', with *so-* which comes from the preposition *sin*. It is important to stress that such a confusion in form and meaning does not occur in LAM, where *sa-*, beside the phonological transparency, still keeps a certain degree of semantic transparency with the original *isja*.

In conclusion, variation in both the status and the form of *isja*, depending on the dialect, is a good illustration of an ongoing linguistic change within the same language, in our case, MG, and could further motivate the existence of a dialectal continuum with a graded hierarchy between its two poles. One of the poles of the particular continuum should contain the dialects of Corfu (Corfiot) and Peloponnese, and also by SMG, where a full adverbial word *isja* is present without being reduced into *sa-*. LAM are situated in the middle of the continuum, since in these dialectal varieties, there is no proof that *sa-* has become a full prefix yet. The other pole contains dialects like Cretan with a fully morphologized *sa-* into a prefix.

5. Dialects and morphological typology

In this last section, I examine the issue of change in morphological typology. I show how informative dialectal information can be into determining the extent of influence of one particular linguistic system onto another in language-contact situations, especially with respect to isolated speech communities which are unlikely to reflect the type-conforming regularities according to Trudgill (2004).

To this purpose, I investigate the nominal inflection of one of the Greek peripheral dialects, namely the Asia-Minor dialect of Cappadocian, which underwent a Turkish influence following the Seljuk invasion in the 11th century, and the subsequent conquest of Asia Minor by the Ottoman Turks in the 14th century.[26]

As already known, Greek is a typical fusional language, whose nominal inflectional endings are portmanteau morphemes, combining the features of number, case and inflection class, while gender, with its three values, i.e. masculine, feminine and neuter, belongs to the

[25] *Sjo-* appears in Western Crete and *so* in Eastern Crete. See Charalambakis (2001), Pangalos (1955) and Ksanthinakis (1996, 2000), for a detailed discussion about the origin and the formation of *s(j)o-*.

[26] Cappadocian was spoken in about 32 Greek-speaking settlements in central Asia Minor (today's Turkey), before 1923, when the exchange of populations between Greece and Turkey took place, following the Lausanne treaty. Today, there are few remaining native speakers, in certain areas of Greece (in the areas of Karditsa, Chalkidiki and Kavala), all of them descendants from Cappadocian refugees.[26]According to scholars who have described Cappadocian (e.g. Dawkins 1916, Janse forthcoming), this dialect shares a lot of similarities with an old form of Greek, namely the one of the Byzantine period.

features of the stem.[27] For instance, in the SMG nominal paradigms of (14) below, the nominative plural of the SMG masculine noun *milos* 'mill' ends in *–i*, and the one of the neuter noun *fito* 'plant' in *–a*.

As opposed to SMG, and according to the grammatical descriptions by Dawkins (1916) and Janse (forthcoming), the Cappadocian nominal inflection system appears to be simplified, in that there are less case forms (e.g. no vocative case) and a smaller variety of inflectional endings. For example, as depicted in (15), the Cappadocian endings corresponding to the SMG nouns of nominative plural are levelled into the single form *–ja*. Furthermore, under the influence of Turkish, which has no gender distinctions, the typical Greek three gender values (masculine, feminine and neuter) have been levelled into one, which has the form of the neuter gender:

(14) SMG		(15) Cappadocian[28]	
a. mat.NEU 'eye'		mat.NEU	
Singular	Plural	Singular	Plural
Nom mati	matj**a**	mat[29]	mat**ja**
Acc mati	matj**a**	mat	mat**ja**
Gen matj**u**	matj**on**	mat**ju**	mat**ju**
Voc mati	matj**a**		
b. fito.NEU 'plant'		fito.NEU	
Singular	Plural	Singular	Plural
Nom fit**o**	fit**a**	fit**o**	fit**a**/fit**ja**/fito**ja**
Acc fit**o**	fit**a**	fit**o**	fit**a**/fit**ja**/fito**ja**
Gen fit**u**	fit**on**	fitu/fit**ju**/fito**ju**	fit**u**/fit**ju**/fito**ju**
Voc fit**o**	fit**a**		
c. milos.MASC 'mill'		milo(s).NEU	
Singular	Plural	Singular	Plural
Nom mil**os**	mil**i**	milo(**s**)[30]	mil**us**/miloz**ja**/milo**ja**
Acc mil**o**	mil**us**	milo(**s**)	mil**us**/miloz**ja**/milo**ja**
Gen mil**u**	mil**on**	mil/mil**ju**/miloz**ju**/milo**ju**	mil/mil**ju**/miloz**ju**/milo**ju**
Voc mil**e**	mil**i**		
d. aδelfos.MASC 'brother'		aδelfo(s).NEU	
Singular	Plural	Singular	Plural
Nom aδelf**os**	aδelf**i**	adelfo(**s**)	adelfoz**ja**/ adelfo**ja**
Acc aδelf**o**	aδelf**us**	adelfo(**s**)	adelfoz**ja**/ adelfo**ja**

[27] See Ralli (1999, 2002, 2005) for details on Greek nominal inflectional features.
[28] The inflectional endings of the nouns below are given in bold characters.
[29] Like the Northern Greek dialects, Cappadocian undergoes high vowel deletion in unstressed position. For the same phonological phenomenon, see also footnotes 7, 15, and 18.
[30] The accusative definite appears without *–s,* while the accusative indefinite usually preserves the *–s*. Occasionally, there is a conflation between the nominative and the asigmatic accusative form (Janse, p.c.).

Gen	aδelf**u**	aδelf**on**	adelfo**ju**/adelfoz**ju**	adelfoz**jaju**
Voc	aδelf**e**	aδelf**i**		

However, the most striking innovation of Cappadocian, as far as the nominal inflection is concerned, is probably an indication for the emergence of an agglutinative pattern, which is also due to Turkish influence. This pattern is particularly evident in the southern varieties of Cappadocian, spoken in the towns of Ulaghats, Gurzono, Fertek, Aravan, and Semendere (see Janse forthcoming). As shown in Janse's grammatical description, for a small number of nouns of these varieties, the plural number and the genitive case are not expressed by the usual portmanteau morphemes, as in SMG and the rest of Greek dialects, but they are realized by distinct markers, which, in some cases, are added to the base, one after the other (see, for instance, (15d)). Consider again the inflection of the nouns 'eye', 'plant', 'mill', and 'brother' in both SMG (14) and Cappadocian (15). In (15) we observe that Cappadocian shows a plural nominative/accusative marker *–ja* and a distinct singular/plural genitive one *–ju*, in all nouns and inflectional paradigms. If we compare the Cappadocian nominal inflection in (15) with that of SMG in (14), we see that *–ja* and *–ju* result from a reanalysis, which has affected the Greek endings *–a* and *–u* and the stem-final vowel /i/ of the most productive paradigm of neuter nouns in *–i* (see *mati* in (14a)).[31] The spread of this reanalysis can be explained by the fact that in Cappadocian the neuter gender has prevailed over the masculine and feminine values, as already stated above.

Crucially, *–ja* and *–ju* are still used as fusional morphemes in some Cappadocian inflected nouns, that is as portmanteau morphemes added to a stem base, as shown by the inflection of *mat-* (15a) and the one of *fit-* (15b). However, it is important to point out that contrary to SMG, where the inflectional endings are usually combined with the stem of the base (see 14), Cappadocian *–ja* and *–ju* can be added to a full word form (see *milos* in (15c) and *adelfos* in (15d)), representing the singular nominative/accusative case, which is taken to be as a default basic form. This phenomenon reminds of the Turkish nominal inflectional paradigms, where the inflected forms are shaped on the basis of a nominative singular word form and not according to a stem form. For instance, the Turkish word for 'plants' is *bitkiler*. It is created by adding the plural ending *–ler* to the word *bitki* 'plant', which is also the form that we find in the nominative singular. On the contrary, the corresponding Greek word is *fita*, which combines the bound stem form *fit-* with the ending *–a*.

Most importantly though, there are hints of an agglutination pattern, which is shown by the example of *adelfo(s)* in (15d). In this example, the plural marker *–ja* and the singular marker *-ju* are added to the base *adelfo(s)*, one after the other (*adelfoz**jaju***), suggesting that *–ju* has lost its original number value (singular), since it appears preceded by a plural marker (*-ja*), and that *–ja* has been deprived from its original nominative/accusative syncretic case values, since it is followed by the genitive marker (*-ju*). This inflectional pattern also reminds the agglutinative Turkish nominal inflection, and a simple comparison of the paradigms of both the Cappadocian and the Turkish inflectional forms for 'brother' could prove this last observation:

(16)

	a. Turkish	b. Cappadocian
	Singular	
Nom	kardeş	adelfo(**s**)
Gen	kardeş**in**	adelfo**ju**/adelfoz**ju**
	Plural	
Nom	kardeş**ler**	adelfo**ja**/adelfoz**ja**
Gen	kardeş**lerin**	adelfoz**jaju**

[31] According to Christophidou (2003), there is evidence from the domain of language acquisition, according to which the paradigm of neuter nouns in *–i* is the most productive in MG.

Note that the same agglutination pattern does not occur in SMG and the other Greek dialects, where the feature of genitive is included either in the morpheme which expresses the singular or in the one expressing the plural, as shown in (14)). Nevertheless, in spite of the fact that (15d) proves that Cappadocian has undergone a Turkish influence, we cannot generalize that the Cappadocian nominal system has been turned into agglutinative, since the *–jaju* combination is not spread to the other paradigms. Besides, this example constitutes a very rare form, which, according to Janse (p.c.) has been detected by Sasse in his field work in early 60's. It is worth adding that a possible change of the Cappadocian nominal typology from fusion into agglutination will probably remain uncompleted, since Cappadocian speakers were forced to abandon Turkey and move to Greece following the exchange of populations between Greece and Turkey in 1923. Today, most descendants of Cappadocian refugees live in mainland Greece, and are in contact with SMG on a daily basis. Very few of them use Cappadocian in family, and Cappadocian is on the way to extinction.

6. Summary

In this paper, I have addressed a fundamental question regarding the relation between dialectology and morphology by showing that the study of dialects can offer a repertoire of entities, phenomena and concepts, which enrich our knowledge of morphology, and are easily adapted within a theoretical morphological framework. On the one hand, morphological theories and approaches systematize the research object, allowing us to code and analyze cross-dialectal data in a transparent way. On the other hand, dialectal research helps us to test and throw light on morphological theoretical issues. In addition, morphological research on dialectal data may determine the limits within which morphology may vary, and assist us into judging the range of cross-dialectal variation against the range of cross-linguistic variation with respect to morphology. Finally, I have contributed to the discussion that the study of dialects is intimately connected to the study of language variation and language change, since a number of grammatical features and patterns of variation found in non-standard varieties are not part of the relevant standard varieties.

Generally speaking, there are many fascinating things that can be detected in the dialects of languages, which are unknown in the relevant standard varieties, and thus overlooked in linguistic studies. The focus of the present study was on Modern Greek dialects, but there is no doubt that morphological studies could be improved if the scope of investigation was broadened to cover dialectal accounts of other languages as well. Thus, there is a rich territory for future work and collaboration between morphologists and dialectologists.

References

Anderwald, Lieselotte & Berndt Kortmann. 2002. Typology and Dialectology: a programmatic sketch. In J. Berns and J. van Marle (eds.) *Present-day dialectology*. Berlin/New York: Mouton De Gruyter. 160-171.

Andriotis, Nikos 1957. *Simvoli sti neoelliniki sinthesi* [A contribution to Modern Greek compounding]. Athens: National Library, vol. 7.

Babiniotis, George 1972. *To Rima tis Ellinikis*. Athens: Idrima Sofia Saripolou.

Baeskow, Heike 2004. *Lexical Properties of Selected Non-native Morphemes in English.* Tübingen: Gunter Narr Verlag.

Bauer, Laurie 1983. *English Word Formation*. Cambridge: Cambridge University Press.

Bauer, Laurie 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie 2005. The Borderline between Derivation and Compounding. In W. Dressler, D. Kastovsky, O. Pfeiffer and F. Rainer (eds.) *Morphology and Its Demarcation.* Amsterdam: John Benjamins. 97-108.

Booij, Geert. 1997a. Autonomous Morphology and Paradigmatic Relations. In G. Booij & J. van Marle (eds.) *Yearbook of Morphology 1997*. Dordrecht: Kluwer. 35-54.

Booij, Geert. 1997b. Allomorphy and the Autonomy of Morphology. *Folia Linguistica* XXXI (1-2): 25-56.

Booij, Geert 2002. Language variation and phonological theory: inflected adjectives in Dutch and related adjectives. In J. Berns and J. van Marle (eds.) *Present-day dialectology*. Berlin/New York: Mouton De Gruyter, 35-56.

Charalambakis, Christophoros 2001. Prolegomena se ena etymologiko leksico tis Kritikis Dialektou. *Kritologika Meletimata : glossa, logotexnia, politismos.* Herakleio: Panepistimiakes Ekdoseis Kritis. 97-109.

Christophidou, Anastasia 2003. Genos ke Klisi stin Elliniki (Mia Fisiki Prosengisi) [Gender and Inflection in Greek (A Natural Approach)]. In A. Anastasiadis-Symeonidis, A. Ralli & D. Cheila Markopoulou (eds.) *To Genos* [Gender]. Athens: Patakis.

Dimela, Eleonora 2005. *I prothimatopoiisi stin Kritiki dialecto. I periptosi ton sun-, s(j)o-, sw-* [Prefixation in Cretan Dialect. The case of sun-, s(j)o-, sw-]. Master Thesis. University of Athens.

Dawkins, Richard. 1916. *Modern Greek in Asia Minor*. Cambridge: Cambridge University Press.

Drachman, Gaberell & Malikouti-Drachman, Angeliki 1994. Stress and Greek Compounding. *Phonologica* 1992: 55-64.

Drachman, Gaberell & Malikouti-Drachman, Angeliki 2001. Concrete Morphology, Affix Typology and Concord Chains. In A. Ralli, B.D. Joseph & M. Janse (eds.) *Proceedings of the 1st International Conference of Modern Greek Dialects and Linguistic Theory.* Patras: University of Patras. 51-65.

Hatzidakis, Georgios. 1905-1907. *Meseonika ke Nea Ellinika* [Medieval and Modern Greek]. Athens: Sakellariou.

Jannaris, Anthony. 1897. *An Historical Greek Grammar*. London: McMillan.

Janse, Mark. Forthcoming. Cappadocian. In Ch. Tzitzilis (ed.) *Modern Greek Dialects*. Thessaloniki.

Joseph, Brian D. 2003. Morphologisation from Syntax. In B.D. Joseph & R. Janda. *Handbook of Historical Linguistics*. Oxford: Blackwell. 472-492.

Joseph, Brian D. & Nick Nicholas 2007. Coordinative Verbal Compounds in Modern Greek. *8th International Conference of Greek Linguistics, Workshop on Compounding*. Ioannina.

Idomeneas, Marios 2006. *Kritiko Glossari*. [Cretan Glossary].

Kaisse, Helen 1982. On the Preservation of Stress in Modern Greek. *Linguistics* 20: 59-82.

Kiparsky, Paul 2003. Analogy as Optimization: Exceptions to Siever's Law in Gothic. In A. Lahiri (ed.) Analogy, Levelling, Markedness. Berlin/New York: Mouton de Gruyter. 15-46.

Ksanthinakis, A. B. 1996. To glossiko idioma tis D. Kritis. *Leksilogio me ermineftika kai etumologika sxolia.* [Lexicon of the dialect of W. Crete]. Chania: Dimotiki Politistiki Epixeirisi Chanion.

Kury owicz, Jercy. 1949. *La nature des procès dits analogiques*. *Acta Linguistica* 5: 15-37.

Lieber, Rochelle 1992. *Deconstructing Morphology.* Chicago: The University of Chicago Press.

Mackridge, Peter. 1987. Greek Speaking Moslems of North-East Turkey: Prolegomena to a Study of the Ophitic sub-dialect of Pontic. *Byzantine and Modern Greek Studies* 11: 115-137.

Maiden, Martin. 1992. Irregularity as a Determinant of Morphological Change. *Journal of Linguistics* 28: 285-312.

Maiden, Martin. 2003. Morphology as a 'Disease. Is the Typological Distinction Healthy? *Paper read at the IV Mediterranean Morphology Meeting*. Catania.

Malikouti-Drachman, Angeliki 1997. Prosodic Domains in Greek Compounding. In G. Drachman, A. Malikouti-Drachman, C. Klidi & J. Fykias (eds.) *Proceedings of the 2nd International Conference of Greek Linguistics*. Graz: Neugebauer Verlag. 87-96.

Manolessou, Io & Symeon Tsolakidis 2007. Ta parataktika Sintheta tis Neas Ellinikis: Diachronia ke Synchronia [Coordinative Compounds of Modern Greek: Diachrony and Synchrony]. In G. Giannakis et als (eds.) *Electronic Proceedings of the 8th International Conference of Greek Linguistics* (Workshop on Compounding). Ioannina.

Nespor, Marina & Angela Ralli 1996. Morphology-Phonology Interface: Stress Domains in Greek Compounds. *The Linguistic Review* 16*:* 357-382.

Newton, Brian. 1972. *The generative Interpretation of a Dialect. A study of Modern Greek Phonology*. Cambridge: Cambridge University Press.

Ralli, Angela 1988. *Eléments de la morphologie du grec moderne: La structure du verbe*. Ph. D. Diss., Université de Montréal.

Ralli, Angela 1992. Compounds in Modern Greek. *Rivista di Linguistica* 4/1: 143-174.

Ralli, Angela. 1999. Inflectional Features and the Morphological Module Hypothesis. *Thessaloniki Working Papers in Theoretical and Applied Linguistics* 6: 111-141.

Ralli, Angela. 2000. A Feature-based Analysis of Greek Nominal Inflection. 2000. *Glossolojia* 11-12: 201-227.

Ralli, Angela 2004. Stem-based vs. word-based morphological configurations: the case of Modern Greek Preverbs. *Lingue e Linguaggio* 2004, 2: 241-275.

Ralli, Angela. 2005. *Morfolojia* [Morphology]. Athens: Patakis.

Ralli, Angela 2007. *I sinthesi Lekseon: Diaglossiki Morfolojiki Prosengisi* [Greek Compounding: A cross-linguistic morphological approach]. Athens: Patakis.

Ralli, Angela 2008a. Compound Markers and Parametric Variation. *Language Typology and Universals (STUF) 61: 19-38.*

Ralli, Angela 2008b. Dvandva [V V] Compounds: a Linguistic Link between Greece and East/South East Asia. Paper read at the

Ralli, Angela to appear. Hellenic Compounds. In R. Lieber & P. Stekauer (eds.) *The Oxford Handbook of Compounding.* Oxford/New York: Oxford University Press.

Ralli, Angela in preparation.a. *Greek Compounds*. To appear in Springer.

Ralli, Angela in preparation.b. *The Asia Minor Dialect of Aivali and Moschonisia*.

Ralli, Angela & Eleonora Dimela. 2007. On the borderline between prefixation and compounding: The *sa-adverbs* in the dialectal varieties of Lesbos, Kydonies (Aivali) and Moschonisia. In P. Pavlou et als (eds.) *Proceedings of the 3rd International Conference of Modern Greek Dialects and Linguistic Theory.* Nicosia.

Ralli, Angela & Athanasios Karasimos. 2007. Compound Constituency and Compound-internal Derivation. In G. Giannakis et als. (eds.) *Electronic Proceedings of the 7th International Conference of Greek Linguistics* (Workshop on Compounding). Ioannina.

Trudgill, Peter. 2004. The impact of language contact and social structure on linguistic structure: Focus on the dialects of Modern Greek. In B. Kortmann (ed.) *Dialectology meets Typology*. Berlin/New York: Mouton de Gruyter. 435-434.

Scalise, Sergio, Antonietta Bisetto & E. Guevara 2004. Selection in Compounding and Derivation. In W. Dressler et als. (eds.) *Morphology and its Demarcations*. 133-150.

Schultink, Hans 1961. Produktiviteit als Morfologisch Fenomeen. *Forum der Letteren* 2: 110-125.

Impoverishment in Dutch dialects

Suzanne Aalberse, *Jan Don*
ACLC, University of Amsterdam

0. Introduction

Dutch dialects show an enormous amount of variation with respect to the verbal inflectional paradigm. To wit, some dialects only have a single form in the present tense indicative to express all persons in singular and plural, whereas other dialects use two, three or even four different forms to do so. However, despite the variation, it is clear that not simply everything is possible; there are logically possible patterns that are found nowhere, whereas other patterns are geographically widespread and stable over time.

Dutch, like any other Germanic language, underwent a process of deflection showing loss of case endings and a loss of verbal inflections. Traces of this diachronic change can still be found in dialectal variation. The loss of these inflectional markings is probably due to dialect contact (Trudgill 1986), leading to a more or less classical spreading of the patterns in which certain distinctions were lost. That is, in more isolated areas where dialect contact is more limited, the process of deflection is less far under way than in those areas where different dialects mix.

Recently, Bennis and MacLean (2006) have tried to explain the variation in Dutch verbal inflection, and the limits thereof, in terms of a general economy principle that minimizes the amount of featural information per affix. We will argue that this purely affix-based approach is in a way too crude to describe the variation found. In this paper we would like to approach the variation by starting from two patterns that seem independent of the affixes that are used to express the inflection. For example, in Standard Dutch, person marking is absent in the past tense. We will argue that this cannot be a property of the past tense suffix, but is a more general 'paradigmatic' property of Dutch. Such patterns of impoverishment, or neutralization, to use a more neutral term, cannot be viewed as resulting from an underlying underspecification of affixes, but should result from a rule or property of the paradigm as such. Within Distributed Morphology (Halle & Marantz 1993) it would be expressed as an Impoverishment Rule, whereas in paradigmatic approaches, it would count as a case of paradigmatic neutralization. We do not take any particular theoretical position with respect to this issue in this paper, but simply use the Impoverishment rules as a descriptive device to express the relevant neutralization patterns.

Our inquiry starts from the assumption that such 'paradigmatic' patterns are stable across dialects, and that variation is basically limited to the affixes used to express these patterns. We will show that starting from this hypothesis, indeed a more or less classic pattern of distribution of variation becomes visible. At the center of the language area (roughly Holland) we find the neutralization patterns *in optima forma*, whereas more at the borders of the area where we can safely assume that less dialect contact occurs, we find the retention of older stages of the language.

1. 'Meta-paradigmatic' structure

Williams (1994) (see also Baerman 2000, and Bobalyik 2003 for discussion) argues that languages seem to display 'meta-paradigmatic' structure; i.e. separate from the individual morphemes, and across different paradigms, the same pattern may recur. To give an example, have a look at the following Dutch data:

(1) a. ***de*** *jongen* 'the boy' *het kind* 'the child'
plur.:***de*** jongen-s **de** kinderen 'the children'
b. ***deze*** *jongen* 'this boy' *dit kind* 'this child'
plur.: ***deze*** jongen-s 'these boys' **deze** kinderen 'these children'
c. ***die*** *jongen* 'that boy' *dat kind* 'that child'
plur.:***die*** jongens 'those boys' **die** kinderen 'those children'
d. een ***mooi-e*** jongen 'a beautiful boy' *een mooi kind* 'a beautiful child'
plur.: ***mooi-e*** jongens ***mooi-e*** kinderen 'beautiful children'

In the left-hand column in (1) we see that nouns with common gender in Dutch show the same choice of demonstrative and adjectival inflection in the singular and in the plural. However, in the right-hand column we see that neutral nouns pattern differently in the plural and the singular. Moreover, we can see that the plural pattern of neuter nouns exactly mirrors the pattern found in common nouns. This could be easily expressed by assuming that there is neutralization of gender in the plural. As a descriptive device to account for such neutralization, we could assume a rule such as (2):

(2) [gender] → ∅ / [plural]

Given this kind of 'meta-paradigmatic' structure, we may expect that such patterns, since they do not depend on individual morphemes and/or phonological patterns, but depend on the feature-structure (cf. Baerman, Brown & Corbett 2005), are stable across dialectal variation. In this way, we hypothesize that dialectal variation is limited to the affixes used in the paradigm, without impinging on the paradigmatic structure, as defined by the morphosyntactic features that are expressed. Put differently, we could also say that independent motivation for the rules or mechanisms with which we want to express this meta-paradigmatic structure, may come from the study of dialectal variation.

This leads us to the following hypothesis:

(3) (Meta-)paradigmatic structure is stable across dialects.

Given this hypothesis, before we turn to an investigation of the dialect-data, let us first consider which patterns in the standard language are of this meta-paradigmatic nature.

2. Standard Dutch

The inflectonal pattern for regular verbs in Dutch (which is an open class and comprises 99% of the verbs in Dutch referentie?) is as follows[1]:

(4) Standard Dutch verbal inflection (regular verbs):

	Present		Past	
	sing.	plur	sing.	plur.
1st	-∅	-en	-te /-de	-ten /-den
2nd	-t	-en	-te /-de	-ten /-den
3rd	-t	-en	-te /-de	-ten /-den

If we were to give a purely affix-based description of the syncretisms in this inflectional paradigm, allowing ourselves underspecification of features, we would arrive at the description in (5)(see also Bennis and MacLean 2006):

(5) Underspecification account:
-te /-de ↔ [past]
-en ↔ [plur]

However, this description misses a generalization, which becomes clear once we take a look at the inflectional paradigm of irregular verbs:

(6) Dutch verbal inflection (irregular verbs):

	Present		Past	
	sing.	plur	sing.	plur.
1st	-∅	-en	-∅	-en

[1] spelled 'e' is [ə]

2nd	-t	-en	-∅	-en
3rd	-t	-en	-∅	-en

Although the past tense of these irregular verbs is not expressed by means of the affix *–de*/*-te*, it still holds true that there is no person distinction in the past. Therefore, the underspecification of the past tense suffix *–te* /*-de* in (5) does not help us to cover this particular syncretism. Moreover, a similar pattern of syncretism is found in the highly irregular verb *zijn* 'to be'. Here, a completely different form expresses the past tense, but again this form is identical for all persons, as can be seen from (7):

(7) Standard Dutch *zijn* ('to be')

	Present		Past	
	sing.	plur	sing.	plur.
1st	ben	zijn	was	waren
2nd	bent	zijn	was	waren
3rd	is	zijn	was	waren

The conclusion is that the syncretisms in the plural and the past tense do not result from the underspecification of affixes, but is a property of the verbal inflectional paradigm of Dutch. In order to describe these syncretisms, we may make use of the following two 'Impoverishment' rules[2].

(8) Impoverishment rules for Standard Dutch:

a. [α person] → ∅ / [plural]

b. [α person] → ∅ / [past]

Given the hypothesis in (3), we expect that the inflectional patterns found in the dialectal variants of Dutch will reflect these impoverishment rules.

3. Dialectal variation in Dutch

The inflectional paradigms of many Dutch dialects have been recorded in a project often referred to as the GTPR-project[3]. These data are available through an electronic version of the Morphological Atlas of the Dutch Dialects (MAND), including no less than 613 measure points spread over the Netherlands, Belgium (Flanders) and the outermost Northern part of France (Van den Berg 2003). The subjects were native speakers of the dialect that were between 50 and 75 years of age and preference given to lower-educated people. Data were gathered through a questionnaire (about 1900 items per questionnaire). We have investigated the full inflectional paradigm of the verb *kloppen* 'to knock'.

3.1 Person marking in the plural

If we take a look at the forms in present tense plural, we find that there is no person marking in 362 (=59%) of the 613 dialects. To compare, 538 dialects (=95%) of the dialects have person markings in the singular present tense. Apparently, the impoverishment rule here describes a pattern that is indeed frequently attested in Dutch dialects, although not in all.

Two groups of dialects do not conform to the generalization that person marking is absent in the plural. First, there is a relatively small group of dialects (20) spoken in the Eastern part of the Netherlands (see map 1) that do not show the impoverished pattern.

This, however, is only an apparent exception. To this, consider first the plural present tense of the verb *kloppen* in the dialect of Smilde:

(9) a. 1st klop-m plural paradigm *kloppen* 'to knock' (Smilde)
2nd klop-t

[2] The term 'impoverishment' does not imply any judgment of whatever nature w.r.t to the language system; it is simply a term which adequatley describes that some feature ([person] in the present case) is not marked in a particular morphosyntactic environment.

[3] Goeman, Taeldeman & Van Reenen project, named after the three persons responsible for this large collection of dialect data.

3rd klop-t

From these data it seems that first person is marked with respect to the other two forms in the plural, which is not expected given the stability of the meta-paradigmatic forms. However, consider next the present plural paradigm from the same dialect of the verb *leven* 'to live':

4 Impoverishment in Dutch dialects[4]

(10) b. 1st leve-t plural paradigm *leven* 'to live' (Smilde)
2nd leve-t
3rd leve-n

In this paradigm, the nasal marks third person distinguishing it from the *–t* in first and second person. If we compare the same paradigms in these 20 dialects, we find that there is free variation between *–t* and a nasal in the plural. So, the marking of person here is only apparent, but underlyingly the system adheres to the impoverishment rule in (8a)[4].4

map 1: dialects with free variation between –n and –t in present tense plural

A second, much larger group of dialects (231 = 38%), does have a person marking in the present plural. A typical paradigm is the one of Nistelrode in (11):

(11) 1st klope plural paradigm of *kloppen* 'to knock' (Nistelrode)
2nd klopt
3rd klope

This pattern is found in the whole of Flanders and the South Eastern part of Dutch Brabant and Dutch Limburg (see map 2).

[4] For more details see Aalberse (2007)

map 2: dialects with –t in second person plural present tense.

De Vogelaer (2005) observes that almost all dialects (=95%) (222/231) with second person marking in the plural have second person pronoun *ge* (or a variant). Very few dialects with *je* (or a variant) have second person marking in the plural (9/231). These dialects are located at the outermost edges of the area.

map 3: 9 dialects showing second person marking in the plural with pers. pron. *je*

So, it could be that this retention of *–t* in the second person plural (from Middle Dutch) is associated with the use of the pronoun *ge*. We will see a similar combination of retention of a

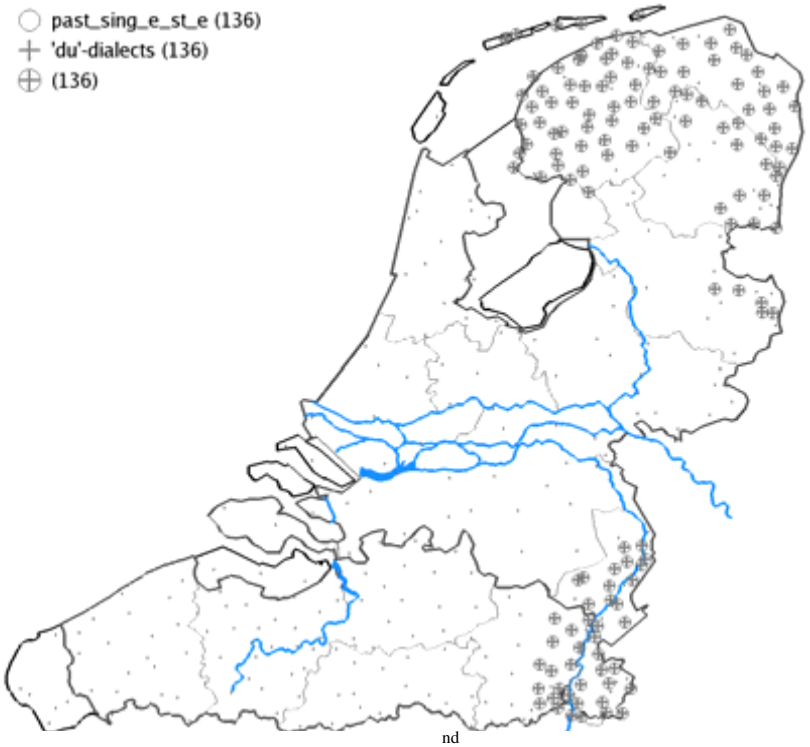
personal pronoun and inflectional ending in the following section. We conclude that, as far as the plural present tense is concerned indeed the impoverished patterns seem stable across dialects, although specific combinations of personal pronouns and inflectional endings may overrule the pattern. Let us now turn to the person markings in the past tense.

3.2. Person marking in the past tense singular

In the singular of the past tense, 451 (=74%) of the dialects show no person marking. So, the impoverishment rule in (8b) also describes a well-attested pattern. However, in the remaining 162 dialects, we do find person marking, contrary to expectation. Again, the group of dialects with person markings (going against the impoverishment pattern) falls apart in two: a. dialects with retention of 'du' as a second person pronoun (136); b. dialects with *-en* in the second person singular (26).

The first group of dialects (136 dialects) that show marking of person in the singular past tense, is characterized by the fact that they all show retention of the second person pronoun 'du'. All these dialects show the ending *–st*, or a variant thereof, in the past. This ending is also found in the present tense (2nd person) in these dialects. It turns out that this personal pronoun is very tightly connected to its original (Middle Dutch) inflectional ending. Every dialect in Dutch that still has the pronoun 'du' also has the ending *–st* in the singular present and past tense (see also Aalberse 2004). Map 4 illustrates this point.

Apparently the same factors that explain the retention of 'du' explain the retention of the *–st* suffix at the same time. We refer to Aalberse (2004, 2007) for a detailed analysis of the factors that contributed to the loss of 'du', here, it suffices to say that there is an independent explanation for the retention of person marking in the singular past tense for this large group of dialects. Interstingly, it is not coincidental that the retention of the *du* plus the ending *-st* is found at the borders of the language area. We can assume that in these areas dialect contact is not an important factor and therefore, certain patterns may be retained that are lost in areas where much more contact between speakers of different varieties is found.



map 4: 136 dialects with 'du' [2nd sing.] and *-st* in present and past 2nd pers. sing.

A second exception to the impoverishment rule (8b) is a group of 25 dialects that have *-en* as a second person marker in the past tense. This ending originally stems from the plural. Dialects that have lost 'du', have replaced (through a strategy of 'politeness') this pronoun

with a second person plural pronoun, and with this pronoun also its inflectional ending. In most cases this yielded *-t* in second person singular, but in the Northern dialects second person plural was already replaced by *–en* before *du* was lost, and therefore, we find *-en* in the second person singular in these dialects. In the 25 dialects, we still find this plural ending *–en* that entered the paradigm via a politeness strategy in the second person singular. Since these dialects are direct neighbours of the dialects that still have *du*, we may conclude that these dialects are the latest that have lost *du*.

map 5: 25 dialects with –en in the second person past tense.

So, it seems that the impoverishment pattern is indeed very stable across dialectal variation. We should be careful however in some cases, since the same 'impoverished' pattern may result from a different origin. There is a small group of 10 dialects that have three different person markings in the present tense singular (e.g. the dialect of Volendam has *klop-kloppe-klopt*), having a uniform *–e* (schwa) in the past tense. Without further assumptions, we may expect that also in the past tense a separate second person marking is present. So, *prima facie* the fact that we find the impoverished pattern in these dialects, seems convincing evidence for the impoverishment rule. However, there is a different interpretation of the impoverished pattern possible. It may also be the case that underlyingly the past tense forms are *-te -ten -te*, respectively, and that there is a deletion of final *–n* which is widely attested in many Dutch dialects. In order to test whether we have exaggerated the impoverished patterns by including these dialects, we have looked at the inflectional patterns in irregular verbs in these dialects. Interestingly, indeed some of these dialects still have second person markings in the past tense (6/10) (*ik was* 'I was'; *jij waarre* 'you were', *hij was* 'he was'), but others have lost the distinction (4/10). In some dialects (e.g. Koedijk) we see both patterns in competition (*jij was // jij waarre*). In this case it cannot be a consequence of a phonological rule. Our conclusion is that we see a transition here towards the impoverished pattern.

3.3. Person marking in the past tense plural

Finally, we turn to the plural marking in the past tense; here we find almost full adherence to the impoversihment rule (8b). Only 43 (=7%) of the dialects show person marking in the past tense plural. These dialects form a proper subset of the dialects that we encountered before, displaying person marking in the present tense plural. A typical paradigm of this sort is found in the dialect of Waregem:

(10) klopten
klopte
klopten

Map 6 gives an idea of the location of these dialects. We note that again these dialects seem to be located at the border of the language area.

43 dialects with past plur. *-ten -te –ten*

4. Conclusion

The Impoverishment rules in (8) reflect a situation of deflection, which is the result of a process of loss of inflectional distinctions (see e.g. Buitenrust Hettema 1891, Van den Berg 1949, Van Haeringen 1956, Weerman 2006). It seems that this deflection process spreads from the economic center of the Netherlands (Zuid-Holland, Utrecht) outwards and that at the borders of the language area different non-deflected patterns are still found. This would be fully in line with what is predicted by models of change as a result of dialect contact (Trudgill 1986). By taking the impoverishment rules in (8) as the core of the deflection process in Dutch verbal inflection, a picture of spreading of deflection emerges which can easily be understood in terms of dialect contact. Therefore, we believe that this view on the variation in the different dialects is essentially on the right track.

Quantitatively, in general, the patterns expressed by the impoverishment rules in (8) are reflected in the dialects. More dialects have person markings in the singular than in the plural; more dialects have person markings in the present than in the past; more dialects have person markings in the present plural and / or the singular past, than in the past plural.

Qualitatively, in areas where for some reason the impoverishment pattern is not fully established, it seems that the diachronic development is towards the impoverishment pattern. This can also be seen from the geographic distribution of the impoverished patterns (away from the center more retention of person markings are found).

The exceptions to the impoverished pattern are linked to the retention of a personal pronoun (*ge* and *du*) which are linked to an inflectional ending that, apparently, with the help of the personal pronoun, seem strong enough to resist the deflection patterns. At the borders of the language area, we find that in areas where *du* is still used as a second person pronoun,

and its inflectional ending is retained throughout the paradigm. At the border of the *du*-area, we find partial retention of second person marking. Furthermore, we find second person marking in the plural in the South, which is almost gone in the past tense, but still survives in the present tense.

Bibliography

Aalberse, S.P. (2004) "Waer bestu bleven? De verdwijning van het pronomen 'du' in een taalvergelijkend perspectief". *Nederlandse Taalkunde* 9:231-252.

Aalberse, S.P. (2007) 'The typology of syncretisms and the status of feature structure. Verbal paradigms across 355 Dutch dialects', *Morphology* 17, p. 109-149.

Baermann, M. D. Brown and G. Corbett (2005). The Syntax-Morphology Interface: A Study *of Syncretism* (Cambridge Studies in Linguistics 109). Cambridge: Cambridge University Press.

Barbiers, S. et al (2006). *Dynamische Syntactische Atlas van de Nederlandse Dialecten* (DynaSAND). Amsterdam, Meertens Instituut. URL: http://www.meertens.knaw.nl/sand/**.**

Bennis, Hans and Alies Maclean (2006) 'Variation in verbal inflection in Dutch dialects', *Morphology* 16: 291-312.

Berg, B. van den (1949). De conjugatie van het praesens in de Noordnederlandse dialecten. *Taal & Tongval* 1, 6-13.

Berg, Boudewijn L. van den (2003) Phonology & Morphology of Dutch & Frisian *Dialects in 1.1 million transcriptions. Goeman-Taeldeman-Van Reenen project 1980-1995.* Meertens Instituut Electronic Publications in Linguistics 3, Meertens Instituut: Amsterdam. (CD-ROM)

Bobaljik, Jonathan (2003)'Syncretism without paradigms: Remarks on Williams 1981, 1994", in: Booij, Geert and Jaap van Marle (eds.) *Yearbook of Morphology* 2002, p. 53-85. Kluwer: Dordrecht.

Bonet, Eulàlia (1991) *Morphology after Syntax: Pronominal Clitics in Romance*, MITWPL: Cambridge MA.

Buitenrust Hettema, F. (1891) 'De aangesproken persoon' *Taal en Letteren*:148-153

Goeman, A. (1999) *T-deletie in Nederlandse dialecten. Kwanitatieve analyse van structurele, ruimtelijke en temporele variatie*. LOT: Utrecht.

Haeringen, C.B. van (1956) *Nederlands tussen Duits en Engels*, Den Haag: Servire.

Halle, Morris (1997) 'Distributed Morphology: Impoverishment and Fission', *MIT Working papers in Linguistics* 30, p. 425-449.

Pinker, Steven (1984) *Language learnability and Language Development*, Cambridge, Mass.: Harvard Univ. Press.

Trudgill (1986) *Dialects in Contact*, Oxford: Blackwell.

Vogelaer, G. de (2005) *Persoonsmarkering in de dialecten van het Nederlands en het Fries*. http://users.ugent.be/~gdvogela/proefschrift

Weerman, Fred (2006) 'It's the economy, stupid! Een vergelijkende blik op *men* en *man*.' In: M. Hüning (red.) *Nederlands tussen Duits en Engels*, Leiden: SNL

Williams, Edwin (1981) On the notions 'lexically related' and 'head of a word', *Linguistic Inquiry* 12, p. 245-274.

Williams, Edwin (1994) 'Remarks on Lexical Knowledge', *Lingua* 92, p.7-34.

Contact:
Suzanne Aalberse / Jan Don
lsg Nederlandse Taalkunde
Spuistraat 134
1012 VB Amsterdam
j.don@uva.nl
s.p.aalberse@uva.nl

Phonologically Conditioned Allomorphy in the Morphology of Surmiran (Rumantsch)[1]

Stephen R. Anderson
Yale University

About twenty years ago, Andrew Carstairs (1986, 1988) discussed some examples that seemed to fall inconveniently between the stools of phonology and morphology: cases where some alternation whose *form* is not plausibly attributed to the operation of phonological rules nonetheless seems to be conditioned by factors that are purely phonological. These examples then lay more or less fallow for more than a decade, apart from occasional attempts to deny their existence. With the rise of Optimality Theory (Prince& Smolensky 2004, originally circulated in 1993) however, tools to handle such phenomena appeared to be at hand, and examples like those cited by Carstairs were revived to some extent as instances in which the actual input form, and not just its output correspondent, could be subject to selection by the constraint system. The principal reference here is a paper by Rene Kager (2007; actually written a number of years earlier), with additional contributions from other scholars in recent years. A current collection of papers on allomorphy (Tranel 2007) contains several papers dealing with these issues.

Most examples that have been treated in the literature occupy quite limited space in the sound pattern of the language concerned, consisting of a small set of affixes or a limited, closed class of stem alternations. This could give the impression that such phonologically conditioned allomorphy really is to be closeted in the morphology, somewhere near suppletion. In this paper, I wish to discuss an example where this kind of alternation has taken over nearly all of the phonologically conditioned variation in the language, and is clearly not marginal at all. The language to be described is Surmiran, a form of Swiss Rumantsch, though it should be noted that essentially the same situation obtains in other forms of Rumantsch, and apparently in some nearby languages of Italy as well. The Rumantsch languages of Switzerland are often seen as part of a wider group of "Rhaeto-Romance" languages, including several Ladin dialects spoken in the Dolomites of Italy and Friulian as well. A review of the issue by Haiman & Beninca (1992a), however, suggests that there is no serious evidence for such a grouping as a genetic unit. The languages in (1) are simply Romance languages with a certain amount of structural similarity spoken in close geographical proximity, but no apparent common developments that would establish a more specific connection among them.

(1) Swiss Rumantsch

Engadine	Central	Western	Dolomitic Ladin	Friulian
Puter	Surmiran	Sursilvan	Gardena	Friulian
Vallader	(Bergun)		Gadera	
(Val Mustair)	(Obervaz)		Fassa	
	Sutsilvan		Livinallongo	
			Ampezzo	

In this chart, the Swiss Rumantsch languages with distinct codified orthographies are unparenthesized, though the actual degree of variation (including the parenthesized forms, among many others) is much greater than this. Unless otherwise identified, the language described here is the standard form of Surmiran that is taught in local schools in the area of

[1] This work was supported in part by NSF award #BCS-0418410 to Yale University, and by awards from the Social Sciences Research Fund at Yale. I first began working on Surmiran while pursuing research on clitic systems, work supported by NSF award #BCS 98–76456 to Yale. The data here are drawn from dictionaries (Sonder & Grisch 1970, Signorell 1999, including the electronic edition of this work, version 2.0), from the grammar of Signorell, Wuethrich-Grisch & Simeon 1987, and in part from my own field work in Salouf and Savognin during the summers of 2002-2007. Comments from the audience at the Sixth Mediterranean Morphology Meeting in Ithaca (Greece), especially Geert Booij, have been useful in preparing this paper

Savognin, used in the newspaper *La Pagina da Surmeir*, and presented in descriptions such as Signorell et al. 1987 and Capeder 2006.

Surmiran, like other forms of Rumantsch (and the Romance languages more generally) displays a great deal of stem-shape variation in its morphology. Pedagogically oriented descriptions highlight its role in verbal alternation; in part this is simply a matter of the "irregular verbs" whose memorization is the bane of the language learner. There is also a good deal of variation which is quite systematic, however. Although controlled by phonological factors, the variation in shape displayed by verbal stems in Surmiran is nonetheless lexically determined and not simply the product of the operation of phonological rules. Once the patterns of alternation are identified, they turn out to be more pervasive, affecting derivational and other morphology as well and constituting a characteristic feature of the language's structure.

1. Regular and Irregular Verbs

Verbs in Surmiran belong in general to one of six overall classes, as distinguished by their infinitive forms and by the vowels that appear in the suffixes of certain inflected forms. The largest and most productive of these is the class of verbs in *-ar*; Figure 1 (derived from Signorell et al. 1987, p. 67) summarizes the main points of the system. Note that the difference between phonetically open ([E, O]) and closed ([e, o]) mid vowels is not reflected in Surmiran orthography, nor is the difference between unstressed [ə] (*a, e*) and full stressed vowels ([a, e]).

Inf.	Example	1pl. Pres.	1sg Imprf.	1sg Fut.	1sg Cond.	PPpl.
-ar [-ar]	*cantar* 'sing'	-agn	-ava	-aro	-ess	-o/ada
-er [-er]	*lascher* 'leave'	-agn	-eva	-aro	-ess	-ea/eda
-ier [-iər]	*spitgier* 'expect'	-agn	-iva	-aro	-ess	-ia/eida
-eir [-Ejr]	*tameir* 'fear'	-agn	-eva	-aro	-ess	-ia/eida
-er [-ər]	*tanscher* 'reach'	-agn	-eva	-aro	-ess	-ia/eida
-eir [-ejr]	*parteir* 'depart'	-ign	-iva	-iro	-iss	-ia/eida

Figure 1: *Surmiran Conjugations*

Finite (inflected) forms of the verb in Surmiran include the Present, Imperfect and Future Indicative, the Present and Imperfect Subjunctive, the Conditional and the Imperative; non-finite forms include the Infinitive, the Gerund (or Present Participle) and the (Past or Perfect) Participle. Most of our attention here will be focused on the Present Indicative, for which a representative paradigm is given in (2).

(2) *cantar* 'sing' (Pres. Indic.):
1sg (ia) cant [kant]
2sg (te) cantas ["kantəs]
3sg (el) canta ["kantə]
1pl (nous) cantagn [kən"tan]
2pl (vous) cantez [kən"tEŃ]
3pl (els) cantan ["kantən]

As can be seen here, the Present Indicative of a regular verb such as *cantar* is formed from the stem of the verb stem plus a set of suffixes: -0, -əs, -ə, -an (-in with verbs in [-ejr]), -EŃ (-iŃ with verbs in [-ejr]), and -ən. Stress falls on the root, except in the first and second person plural forms, where it falls on the suffix.

In addition to "regular verbs" like *cantar*, Surmiran has a number of irregular verbs whose paradigms are not as straightforward as that in (2). Consider the paradigms of some of these, given in Figure 2.

	eir 'go'	*neir* 'come'	(*vu*)*leir* 'want'	*deir* 'say'	*star* 'stay, live'	*saveir* 'know'
1sg	vign	vign	vi	dei	stung	sa
2sg	vast	vignst	vot	deist	stast	sast
3sg	vo	vign	vot	dei	stat	so
1pl	giagn	nign	lagn	schagn	stagn	savagn

2pl gez	niz	lez	schez	stez	savez
3pl von	vignan	vottan	deian	stattan	son

Figure 2: *Some Irregular Verbs in Surmiran*

As in other Romance languages, these (and other) irregular paradigms are not completely unrelated to those of regular verbs, or to one another, but nonetheless display considerable idiosyncrasy. As a first approximation, we might assume that while a completely regular verb like *cantar* need list nothing in the lexicon besides its stem (/kant-/), with the inflected forms of (2) being produced by rule from this. The only variation in stem shape is between stressed) [a] and (unstressed) [ə], not reflected in the orthography. It is tempting to regard this as merely a matter of low level phonetics, but it is worth noting that [a] and [ə] must be lexically distinguished, in order for instance to distinguish the infinitival endings of the first and second conjugations in Figure 1 from the fifth.

In contrast, irregular verbs like those in Figure 2 must list their individual forms, each specified for person and number. The form *stung*, for example, is listed as [StuN] with the morphosyntactic features [+ME, −PL, +PRES +INDIC] within the lexical entry for *star* 'stay, live'. By well-known principles of disjunctive application (or 'blocking': see Anderson 1992 among many other references), this specifically characterized form will take precedence over one generated by rule.

We could say that regular and irregular verbs differ structurally in this way, then. The former have only a single stem listed in their lexical entry, a base which is not specified for morphosyntactic properties, while the latter have stem forms characterized for specific combinations of features (perhaps in addition to one or more 'default' stems). As we will see in the following section, however, this distinction does not exhaust the range of possibilities in the Surmiran lexicon.

2. "Alternating" Verbs

Compare the paradigms in (3) with that of a regular verb like *cantar* shown in (2) or those of irregular verbs in Figure 2.

(3)	*ludar* 'praise'	*durmeir* 'sleep'	*lavar* 'get up'	*fittar* 'finish'
1sg	lod	dorm	lev	fet
2sg	lodas	dormas	levas	fettas
3sg	loda	dorma	leva	fetta
1pl	ludagn	durmign	lavagn	fittagn
2pl	ludez	durmiz	lavez	fittez
3pl	lodan	dorman	levan	fettan

Each of the verbs in (3) displays two distinct stems: [lod]/[lUd], [dorm]/[dUrm], [lev]/[ləv], and [fet]/[fɪt]. The first variant occurs with all three persons of the singular, and in the third plural, while the second variant occurs in the first and second person plural. It is the relation between these two stem variants that constitutes the primary focus of this paper.

2.1. The Conditions for the Alternation

We might be tempted to treat these alternating verbs together with irregular verbs like those in Figure 2. On this approach, we could list the first stem variant in each case with no morphosyntactic properties (thus making it the default), and the second with the properties [{+ME, +YOU}, +PL]. This approach clearly fails, however. We can notice immediately that it is the second of the two variants that appears in the infinitives (*ludar, durmeir, lavar, fittar*), although the infinitive does not bear the features of the first or second plural. Furthermore, this same stem variant also appears throughout several other verbal categories, as illustrated in (4).

(4)

infinitive:	ludar	durmeir	lavar	fittar
1pl Pres.	ludagn	durmign	lavagn	fittagn
1sg Imperf.	ludeva	durmiva	laveva	fitteva

1sg Fut.	ludaro	durmiro	lavaro	fittaro
1sg Condit.	ludess	durmiss	lavess	fittess
2pl Imper.	lude	durmi	lave	fitte
Pres. Ppl.	ludond	durmond	lavond	fittond

The alternative of treating the second of the two stemforms as the default, and the first as specified, does not improve matters. First of all, it is hard to construct a feature specification that includes all of the singular and (only) the third person plural. But in any event, this stem variant also appears as the basis of some other verbal categories, as shown in (5).

(5)

infinitive:	ludar	durmeir	lavar	fittar
1sg Pres. Indic	lod	dorm	lev	fet
1pl. Pres. Subj.	lodan	dorman	levan	fettan
2sg Imper.	loda	dorma	leva	fetta

Furthermore, the same category may be based on either of the two stems. Infinitives of verbs that form their infinitive with [-ər] are based on the first stem (that found in the singular of the Present Indicative). Thus, *discorrer* [dɪS"korər] 'speak' uses the same stem in the infinitive as in 1sg Present *discor*, and not that of the 1pl Present form *discurrign*. Verbs in the other conjugations of Figure 1 use the second stem (the one appearing in the first and second plural of the present indicative) as the base of the infinitive.

It does not appear, therefore, that a coherent definition of the conditions of use of each of the two shapes in which such "alternating" stems occur is possible in terms of morphosyntactic features. Signorell et al. 1987, p. 68 (and the related dictionary of Signorell 1999) suggest that the forms in (4) are actually derived from the first person plural present indicative, while those in (5) are derived from the third singular present indicative, but no such appeal to "parasitic derivations" is required. In fact, following the earlier description in Sonder & Grisch's (1970) dictionary, there is a simple regularity that governs the distribution of the two stem variants: the one appearing in the 1pl present and the other categories in (4) is used precisely when the main stress of the form is on the ending, while the stem shape associated with categories such as those of (5) is used precisely when the ending is unstressed, and main stress falls on the root. The infinitive forms illustrate this nicely: the ending [-ər] is unstressed, and so infinitives in this class have stress on the root, while all other variants of the infinitive ending in Figure 1 themselves take stress. The associated difference in stem alternants on which the infinitive is based is simply another instance of this same principle.

2.2. Stressed and Unstressed Syllables in Surmiran

Since stress is evidently the conditioning factor for the stem alternations, it is necessary to say a few words about how it is assigned before proceeding further. The influx of borrowings, especially from German, has somewhat obscured the basic principles, since many such words have exceptional stress. Within the native (and nativized) vocabulary, however, there is a relatively simple regularity. Main stress falls on the penult if the rhyme of the final syllable consists of [ə], possibly followed by [r], [l], [n] or [s]. If the final syllable contains a full (non-ə) vowel, or if the final consonant is other than one of those just listed (e.g., final syllables in [əm]), main stress is on the last syllable. Assuming that syllable rhymes of the form [ə] plus [r, l, n, s] are light, and others are heavy, this can be described by the rule in (6).

(6) Build a quantity-sensitive trochee at the right edge of the word.

Parts of compounds are stressed separately with main stress on the stress center of the final element.

In addition to the main stress, secondary stress falls on initial syllables separated by at least one syllable from the main stress. Other secondary stresses appear to be the result of cyclic word formation, although the principles at work have not yet been fully worked out.

Stressed syllables can contain a variety of vowels and diphthongs, with long vowels and diphthongs restricted to primary stressed syllables. Unstressed syllables, however, normally contain only short [ə] (written *a* or *e*), [i] or [u], although short mid vowels also occur in some unstressed syllables as a result of borrowing.

When we look again at a "regular" verb like *cantar* in (2), we see that in fact (as already noted above) the stem shows two alternants, distributed in the same way as in the "alternating" verbs of (3): [kənt] before endings that bear stress, and [kant] when main stress falls on the root. The difference is obscured here by the orthography, which represents (stressed) [a] and (unstressed) [ə] in the same way, as orthographic *a*, but there is a difference nonetheless.

The nature of the stem variation in *cantar* suggests an account of the alternating verbs. The difference between the two stems of verbs like those in (3) is a matter of vowel quality, with the "unstressed" stem always displaying an appropriate vowel from the reduced set in its final syllable, while the "stressed" stem may have essentially any vowel. Whether the main stress falls on the root or on the ending, furthermore, is reflected in whether that stress falls on the vowel that differentiates the two stem shapes or not. Perhaps the difference between the two stems is simply a consequence of a phonological rule relating the vowel systems of stressed and unstressed syllables, as seems plausible for the alternation in ["kant] 'I sing' *vs.* [kən"tan] 'we sing'. We might assign these verbs a single underlying stem shape and derive the other by a phonological rule of vowel reduction in unstressed syllables. We will see, however, that matters are not as straightforward as that.

2.3. Vowel Alternations in Verbs

In order to reduce the stem alternation to a phonological rule of vowel reduction in unstressed syllables, we must be able to find a unitary underlying form for each verb from which both of its stems can be derived by rule. It is fairly obvious that the unstressed alternant cannot serve this purpose. As we have just observed, only three vowels normally appear in unstressed syllables: [ə] (orthographic *a* or *e*), [I] (orthographic *i*) and [U] (orthographic *u*). When [ə] appears in the unstressed alternant, the corresponding stressed alternant may contain any of at least eight distinct vowels and diphthongs, as illustrated in (7).

(7)

Stressed V	Infinitive	3sg Pres. Indic.	gloss
[a]	*l[ə]var*	*lava*	'wash'
[ai]	*[ə]ntrar*	*aintra*	'enter'
[E]	*t[ə]dlar*	*tedla*	'listen'
[e]	*l[ə]var*	*leva*	'get up'
[Ei]	*p[ə]sar*	*peisa*	'weigh'
[ei]	*antsch[ə]dar*	*antscheida*	'start yeast'
[i]	*surv[ə]gneir*	*survign*	'receive'
[o]	*cl[ə]mar*	*cloma*	'call'

Minimal pairs such as *lavar/lava* 'wash' vs. *lavar/lev* 'get up' make it clear that no other properties of the environment are likely to be available to disambiguate these relations. Similarly, unstressed [I] can correspond to any of at least eight different vowels and diphthongs in the stressed alternant, as illustrated in (8).

(8)

Stressed V	Infinitive	3sg Pres. Indic	gloss
[a]	(*sa*) *tgil[I]ttar*	*tgilatta*	'sit down (scornfully, as of a cat)'
[ai]	*spisg[I]ntar*	*spisgiainta*	'feed'
[E]	*p[I]glier*	*peglia*	'take'
[e]	*f[I]mar*	*fema*	'smoke'
[ei]	*anv[I]dar*	*anveida*	'invite'
[i]	*tg[I]rar*	*tgira*	'guard'
[ie]	*s[I]var*	*sieva*	'sweat'
[o]	*dum[I]gnar*	*dumogna*	'dominate'

Finally, unstressed [U] can correspond to at least seven distinct stressed vowels and diphthongs, as illustrated in (9).

(9)

Stressed V	Infinitive	3sg Pres. Indic.	gloss
[a]	*v[Ú]rdar*	*varda*	'watch'
[O]	*d[Ú]rmeir*	*dorma*	'sleep'
[o]	*cr[Ú]dar*	*croda*	'fall'

[o:]	*p[Ú]ssar*	*pʈssa*	‘rest’
[oi]	*l[Ú]ier*	*loia*	‘arrange’
[ou]	*ram[Ú]rar*	*ramoura*	‘roll, surge’
[u]	*p[Ú]gnier*	*pugna*	‘fight, box’

An examination of the data in (7–9) should make it clear that the choice of the stressed alternant will not succeed either, since the same stressed vowel can correspond to more than one (and in some cases, such as [a] and [o], to all three) unstressed vowels. Indeed, there is no stressed vowel whose unstressed correspondent is unique. The variation between the two alternative stems for a given verb, then, cannot be derived from the phonology alone.

This is not to deny that historically, the source of these alternations was phonological. Roughly, the system that is found in stressed syllables, consisting of seven vowels, each long or short, and a number of diphthongs, was collapsed to three short reduced vowels in unstressed position. Low vowels (including [a, E, and O] reduced to [] in unstressed syllables, while non-low front vowels ([i, e]) reduced to [ɪ] and non-low back vowels to [U]. For the details of these developments, see Lutta 1923, pp. 120–136, Grisch 1939, pp. 76–94 and Eichenhofer 1989.

As emphasized by Haiman & Beninca (1992b, pp. 56–63), however, the history of individual words is nowhere near so simple in Surmiran or any of the other forms of Swiss Rumantsch as a mere rule of vowel reduction in unstressed syllables. Intermediate historical stages in the development of the unstressed vowels are preserved in certain forms, and many have been restructured so that the alternations found today are by no means always the etymologically expected ones. Furthermore, borrowed words from German and other languages have introduced unstressed vowels other than [ə, ɪ, U], contributing to the increasing opacity of the vowel reduction rule as a purely phonological generalization. It seems clear that the pattern of stem alternation was interpreted as a morphological property of individual verbs at an early point, and thereby became separated logically from the operation of the phonology so as to give rise to the system of phonologically conditioned stem allomorphy which we can observe in the modern language(s).

2.4. Consonant Alternations

The vowel changes we have seen thus far are by no means the only ways in which the two stress-conditioned alternant stems of verbs may differ. In a number of verbs, *gn* ([n]) or *ng* ([N]) following the stressed vowel of the stressed alternant corresponds to *n* ([n]) in the unstressed alternant, as in the verbs in (10).

(10)

Infinitive	3sg Pres. Indic.	gloss
manar	*magna*	‘lead’
cuschinar	*cuschigna*	‘cook’
splanar	*splanga*	‘plane’
amplunar	*amplunga*	‘pile up’

Sometimes this is accompanied by vowel change as well, as in (11).

(11)

Infinitive	3sg Pres. Indic.	gloss
(*sa*) *sdanar*	*sdegna*	‘shrink from doing s.t.’
(*s’*)*anclinar*	*anclegna*	‘bend’
smarschanar	*smarschunga*	‘loaf ’

The nasal alternations, like those affecting vowel quality, have historical roots in phonologically governed patterns. At some point in the development of Surmiran, intervocalic /n/ following main stress became [n] after front vowels and [N] after back vowels. The dialect distribution of this change was rather complex, however, and exceptions soon developed, causing the regularity to become opaque and in part lexicalized (see Grisch 1939, pp. 74f.).

In modern Surmiran, the [n]/[n] alternation is apparently predictable in many cases. Let us consider first the most productive class of verbs, those whose infinitives end in *-ar*. Of the approximately 90 verbs with infinitives ending in *-inar* that are listed in Signorell 1999,

every one for which evidence is available[2] has a stressed stem in *-egn* or *-ign*. The palatalization of /n/ after stressed [i] is thus automatic in this case. Three verbs end in the sequence *-egnar*: *cregnar* 'soak', *impegnar* 'distrain (seize to obtain payment of money owed)', and *impregnar* 'impregnate, marinate'. The final [n] in these is invariant, and the vowel quality suggests an exceptional secondary stress in the 'unstressed' stem form.

On the other hand, the [n]/[n] alternation also appears in around a dozen *-ar* verbs with other vowels (e.g. *smanar*/*smagna* 'swing'), and here it cannot be an automatic concomitant of stress shift, because comparable verbs with non-alternating *n* also exist, such as those in (12).

(12)

Infinitive	3sg Pres. Indic.	gloss
anganar	*angiona*	'swindle'
scanar	*stgona*/*scana*	'stab'

The alternation must thus be a lexical characteristic of *-ar* verbs whose unstressed stem vowel is other than /i/. Next consider *-ier* verbs. Among these, there are none that end in /-Vn/, but a number ending in /-Vn/. The same is true of verbs ending in stressed *-er*. These facts suggest that post-vocalic /n/ is obligatorily palatalized in these classes. Of the verbs ending in unstressed *-er* ([-ər]), there are a number whose stems end in non-alternating /n/. All of these appear to be prefixed forms based on a single root /pon/: *cumponer* 'compose', *deponer* 'put down', *opponer* 'oppose', etc. The suggestion, nonetheless, is that only /n/ and not /n/ appears in either stem of verbs of this class.

Among the *-eir* ([-Eir] or [-eir]) verbs, there appears to be only one that is relevant: (*s'*)*accumplaneir*/*accumplagna* 'come true'. There is no way to tell from this single case whether alternation should be considered optional or obligatory for such verbs. To summarize: verbs belonging to classes whose infinitive ending is [-iər] or [-er] display only non-alternating /n/ and not /n/ in both their stressed and unstressed stems. Verbs whose infinitive ending is [-ər] (unstressed *-er*) display only /n/ and not /n/ in both. There is very little evidence about verbs in the two *-eir* classes, although one verb does display an alternation (/n/ in the stressed stem, /n/ in the unstressed).

Verbs in the highly productive *-ar* class have final /n/ in their unstressed stems, with the exception of three verbs that have invariant /n/ after the vowel /e/. The stressed stems of these have [n] after [i] or [e], and may have either [n] or [n] after [a]. There are clearly generalizations to be found here about the distribution of [n] and [n] at the ends of verb stems. These are, however, generalizations about the lexical structure of verbal stems. In the language as a whole, [n] and [n] appear in contrast in a variety of environments, and it is not possible to state a purely phonological regularity (i.e., one that refers only to the phonological environment of the segments concerned) from which the generalizations true of verb stems would follow. The status of [n]/[N] alternations is somewhat different. This alternation only occurs in verbs of the *-ar* class, and generally only in stressed stems that have the vowel [u]. Every verb in Signorell 1999 that ends in *unar* in the infinitive and that has a stressed stem with stress on the [u] changes the /n/ to [N] in this case, with one exception: *cugliunar*/*cugliuna* 'cheat'. Similarly, every verb in *anar* that changes the stem /a/ to [u] also changes /n/ to [N]. There appears to be a phonological regularity to the effect that stressed *un* at the end of a verb stem becomes [-uN].

This is consistent with other peculiarities of [N] in Surmiran. This sound generally appears only (a) as a result of assimilation of [n] to a following velar, or (b) in final position, after a stressed vowel (almost always [+Back], apart from words in *-ing* borrowed from English like *surfing*), a position from which [n] is largely excluded. The segment [n] is likewise excluded from appearance in final position after stressed [u] apart from the word *pugn* 'fist' and its derivatives. A great many nouns in the language end in *-iun*, which is consistently pronounced [i"uN]. The adjective *bun*/*buna* 'good' has become *bung* in the masculine except in fixed expressions (e.g., *bun de* 'good day!', *da bun humor* 'of a good

[2] A number of these verbs form their stressed stem by suffixing *-esch* as will be discussed below, and so do not have a form in which the relevant vowel takes the main stress.

disposition'), and this final [N] has been extended to the feminine *bunga*, again except for fixed expressions (e.g. *buna seira* 'good afternoon/evening!', *da buna fθ* 'in good faith') where stem-final [n] is preserved.

These facts suggest a regularity of the Stem-level phonology[3] by which a nasal following stressed [u] in a final syllable is obligatorily [N]._at would account for all of the [n]/[N] alternations in verbs with one exception, *splanar/splanga* '(to) plane'. This raises some interesting issues concerning the architecture of the grammar, however. What we want to say is that the stressed and unstressed stems of a verb like *sclavunar/sclavunga* 'clean the bottomof an oven' have the same segmental form(/skləvun/), with the [N] of the stressed stem following from the stress on final syllable [u]. Note what that implies, however. The stress on this stem must be present at a point where it is still possible to determine that the /un/ sequence is stem-final, because [n] is perfectly possible after a stressed [u] in the penultimate syllable of a word (cf. *striuna* 'sorceress'). This means either that (contrary to views of morphology such as that presented in Anderson 1992) information about morphological structure must be preserved after the addition of unstressed endings (e.g. [-ə] '3sg Pres. Indic.'), or else that the regularity affecting stem-final nasals after stressed /u/ must be enforced at a point prior to the selection of this stem for use in a specific word form. This latter possibility is of course perfectly consistent with the notion that the two stems of a verb are formed (including the assignment of metrical structure) at the Stem level, and then one or the other is chosen at the Word level in association with inflectional material added at that point. It is more difficult to reconcile with monolithic (i.e., non-level-ordered) versions of OT, however, or with a variety of rule-based models.

2.5. More Complex Alternation Patterns

Thus far, our consideration of the differences between the stressed and unstressed stems of alternating verbs has been limited to cases in which it was the last vowel of the stem (and/or the character of a nasal consonant following that vowel) that differs from one stem to the other. These are the cases that most directly have their explanation in the history of vowel reduction in Surmiran, but they are by no means the only types that are found.

A number of verbs show alternation in the penultimate vowel of the stem as well as (or in a few cases, instead of) the last vowel. Some of these are illustrated in (13).

(13)

Alternation	Infinitive	3sg Pres. Indic.	gloss
a–ə~o–e	*flammager*	*flommegia*	'blaze'
e–ə~ə–e	*declarar*	*daclera*	'declare'
i–i~ə–e	*angivinar*	*angiavegna*	'solve'
i–i~ə–ei	*misirar*	*maseira*	'measure'
i–i~ə–i	*ghisignier*	*gasigna*	'taunt'
u–ə~ə–o	*murmagner*	*marmogna*	'murmur'
u–ə~ə–oi	*suarar*	*savoira*	'smell'
u–ə~ə–u	*ruschanar*	*raschunga*	'speak'
u–0~ə–ou	*luvrar*	*lavoura*	'work'

These patterns are less common than those involving only the last vowel of the stem, but they exist nonetheless. Like the simpler alternations, they have their source in the complex history of reduced vowels in Surmiran, as detailed in the handbooks. For example, the verb *luvrar/lavoura* derives from the stem of LABORARE. In the stressed stem, the stress falls on the second vowel (I) which develops straightforwardly to [ou], with the initial A becoming [ə]. In the unstressed stem, however, the first two syllables formed a single trochaic foot at some point, subordinated to the primary stress borne by the ending. Within such a foot, if the second vowel was [+Round] and the first was not, rounding was transferred to the first vowel.

[3] I assume here an architecture along the lines of Stratal Optimality _eory (cf. Kiparsky 2000, 2003 and elsewhere), with potentially distinct Root, Stem,Word and Post-Lexical phonologies each of which is represented by a system of constraints.

This change is reflected[4] in a number of other words, such as the alternation in words like *ruscha-nar*/*rashunga* and *Suagnin*, a spoken form of the place name *Savognin* (whose official form has been remodeled from dialects in which this change did not take place). In the unstressed stem of 'to work' the second, reduced vowel of the initial foot was syncopated, yielding *luvrar* as the infinitive. This and a number of similar changes are no longer part of the synchronic phonology of the language, but have left their traces in the stem alternations of (13).

A few verbs display an alternation that looks at first glance like a variety of metathesis. *Bargeir* 'cry', 3sg. Pres. Indic *bragia*, *patartger*/*patratga* 'think', *sgartar*/*sgratta* 'scratch' and one or two others seem to display an *ar*/*ra* alternation between their unstressed and stressed stems. When we consider the verbs *crescher* ["krešər] 'be brought up', 1pl. Pres. Indic. *carschagn*, and *sgarmar* 'remove the cream from milk', 3sg. Pres. Indic. *sgroma*, however, this appears in a somewhat different light. Rather than metathesis, what we have in these verbs is an alternation between a vowel that appears in the stressed stem and 0, with an epenthetic schwa inserted in the (otherwise vowel-less) unstressed stem: /zgrat/, /kreš/, /zgrom/ as stressed stems corresponding to unstressed /zgrt/, /krš/, /zgrm/ with schwas inserted before /r/ to give [zgərt], [krəš] and [zgərm] as the surface forms of the stem. An alternation between [ə] and 0 before liquids occurs in other forms in the language; whether a unitary principle of epenthesis (or deletion) governs all of these cases remains to be established, but the claim of secondary nature for the [ə] (orthographic *a*) in the relevant verbs does not seem problematic.

The alternation between a full vowel in the stressed stem and 0 in the unstressed stem (with subsequent epenthesis of [ə] before /r/) of the verbs just discussed has precedent in another verb, *glisnarger* 'simulate', 3sg. Pres. Indic. *glisnaregia*, and possibly also in *cloccar*/*clocchegia* 'gurgle' and *tuslar*/*tuslegia* 'spank, clobber' (with simplification of the clusters /kj/ and /slj/ in the unstressed stems, rather than epenthesis)[5].

Finally, a ə/0 alternation is also found, but in the opposite direction, in the verbs *sbusarar*/*sbusra* 'go wrong' and *sdaranar*/*sderna* 'lay down'. Once again, a [ə] is found before /r/ in the unstressed stem, which may suggest a connection with verbs like *sgarmar*/*sgroma*. Neither alternation, however, seems purely amatter of the synchronic phonology of the language.

2.6. The Stem Suffix *-esch*

Thus far, we have established that a great many verbs have two distinct stems, distributed in accord with a rather straightforward principle: one stem appears in word forms involving an affix which itself bears the main word stress (e.g., the first and second person plural present indicative suffixes) while the other appears with affixes that do not themselves bear stress. The stem allomorphy here is clearly phonologically conditioned, rather than being governed by themorphological categories of the words in which one stem or the other appears. The question that naturally arises is whether the alternations we have seen are entirely matters of the phonology of Surmiran.

It is uncontroversial that these alternations are *historically* phonological in their origin: the principles governing the reduction of vowels in unstressed syllables have clearly played a central role, for one thing. But the survey of alternations that we have made in the preceding sections makes itmuch less likely that these historical processes could bemotivated for the synchronic grammar of the language. We are thus led to the conclusion that the stem variation found in individual verbs is largely (if not entirely) a matter of lexical specification, with the lexical entry for each verb specifying not one single underlying stem form but two, one that appears in words containing a stressed suffix and one that appears when the stress

[4] See Grisch 1939, pp. 93f. for discussion and further examples.

[5] These may all be reflexes of verbs in -IDIARE; other verbs of that class (cf. Grisch 1939, p. 163) have vowel/ə alternations in the corresponding position instead of vowel/;, as in *manager*/*manegia* 'mean' (stem from German *meinen*), (*sa*) *turpager*/*turpegia* 'be ashamed'.

falls on the stem itself. This situation is of course familiar at least in part from many other Romance languages, but seems to be particularly robust in Surmiran (and Rumantsch more generally).

The conclusion that the stem variation we have seen cannot have its origin in the operation of independently motivated phonology is confirmed by another large (and productive) class of verbs. For these verbs, the stressed and unstressed stems differ not in terms of vowel alternations or the like, but rather in the presence (*vs.* absence) of a suffix *-esch* ([-Eš]).

Consider the present indicative paradigmof the verb *luschardar* ([lužər"dar]) 'strut' in (14) for example.

(13)

1sg	luschardesch
2sg	luschardeschas
3sg	luschardescha
1pl	luschardagn
2pl	luschardez
3pl	luschardeschan

Here the unstressed stem /lužərd/, which appears with the affixes *-agn* and *-ez* (as well as with all other stressed affixes in other tenses), is extended by the suffix *–esch* whenever stress would be expected fall on the stem. The basis of the variation between the presence *vs.* absence of *-esch* is exactly the same as in the paradigm of ordinary alternating verbs like (3), but here the possibility of a phonological regularity that inserts (or deletes) the sequence /Eš/ does not exist.

A search of the Surmiran lexicon turns up a great many verbs whose paradigm is like (14). Newly borrowed verbs from other languages often tend to be assigned this pattern. A number of verbs that are listed in Sonder & Grisch 1970 with one or another of the patterns of alternation we have surveyed above are identified as *–esch* verbs in the later dictionary of Signorell (1999). When speakers cannot recall the correct alternation pattern for a given verb, they sometimes produce an *-esch* form instead.

The motivation for this productivity is clear. When stress is assigned to a stem ending in *-esch*, it is *-esch* itself, rather than the lexical part of the stem on which it falls. In a paradigm like (14), the result is that this lexical portion of the stem is always unstressed, and thus displays no stressed/unstressed alternation. The use of this pattern, then, has the advantage that the speaker does not need to retrieve any information about the specific alternation pattern of the stem in order to produce all of the correct forms. Otherwise, it would be necessary to choose (for the 3sg Pres. Indic. form of *luschardar*, for example) among a variety of possibilities such as **luscharda*, **luscheirda*, **luschorda*, **laschurda*, **laschorda*, etc. Each of these patterns is more or less secure with reference to at least some verbs in the Surmiran lexicon, but the availability of the paradigm in (14) makes it possible to avoid the choice when positive evidence is not readily available.

The correctness of this account is indirectly supported by the description for native speakers in the normative grammar of Signorell et al. 1987, pp. 73f.. After noting that "a number of verbs, almost all with the infinitive ending *-ar* or *-eir*" show the paradigm in (14), and explicitly relating this to the phenomenon of alternating verbs, the authors go on to characterize its scope in negative terms. Essentially, the *–esch* form is avoided when the correct alternation pattern is known. Otherwise, however, a potentially alternating verb can be treated in this way.

The full scope of usage of the verbal paradigm based on a stressed stem consisting of the unstressed stem extended by *-esch* is not clear, though it is apparent that it is quite robust. Since the principles governing the distribution of the two stems remain the same as those governing (other) alternating verbs, and the presence *vs.* absence of *-esch* itself cannot be due to the operation of a purely phonological rule, this supports the claim that the stem alternation pattern in general is a matter of phonologically conditioned suppletion in the sense of Carstairs (1988), not simply phonology.

2.7 Alternations in Other Word Classes

Up to this point, we have described the stress-based alternation patterns in Surmiran as a fact about the verbal system. Closer inspection, however, reveals the fact that these patterns are not at all limited in that way. Apart from their role in verbal inflection, similar stem alternations appear in derivationally related forms as illustrated in (15).

(15)

ei~ə	neiv	'snow'	navada	'much snow'
ei~i	stgeir	'dark (adj.)'	stgirantar	'get dark'
ou~u	pour	'farmer'	puraglia	'peasantry'
o~u	fora	'opening'	furela	'entrance'
e~i	fem	'smoke'	fimera	'dense smoke'

Typically, when a verb has "stressed" and "unstressed" stems, derivationally related forms will be built on one or the other, depending on where stress falls in the derived form as illustrated in (16).

(16)

ludar/loda	'to praise':		
(igl) lod	'praise (n.)'	ludevel	'praiseworthy'
clamar/cloma	'to call':		
(igl) clom	'call (n.)'	(la) clamada	'calling (n.)'
gartager/gartegia	'to succeed':		
(igl) gartetg	'success'	malgartagea	'ill brought up'
stimar/stema	'attend to, value':		
(la) stema	'worth'	(la) stimadeira	'valuation'

But in a significant number of forms, illustrated in (17), the "stressed" stem appears in a form where it does not take the stress.

(17)

"sfend[ər]/sfandagn	'(to) split':		
sfandia 'cracked (adj)'		sfendibel	'splittable'
durmeir/dorma	'(to) sleep':		
durmigliun 'late riser'		dormulent	'sleepy'
satger/setga	'(to) dry [intr.]':		
setg	'dry (adj.)'	setgantar	'(to) dry [trans.]'
acccumadar/accumoda	'adjust':		
accumodabel	'adjustable'	accumodamaint	'adjustment'
accumpagner/accumpogna	'accompany':		
accumpagneder	'accompanist'	accumpognamaint	'accompaniment'

These may result from cyclic application, with stemchoice taking place on one cycle and further morphology (and alteration of stress pattern) taking place on a later cycle (cf. Kamprath 1987 for discussion of motivations for cyclic interaction in a closely related form of Rumantsch, the language spoken in Bergun/Bravuogn). Further exploration of the Lexical Phonology of the language is necessary before this suggestion can be considered confirmed.

There is one major difference between the alternations in verbal inflection and what we find in other categories. In particular, the "stressed" stemin *-esch* never shows up outside of verbal inflection. Verbs that take *-esch* in the stem-stressed forms always use the "unstressed" stem as the base for derivation (e.g., *fixar/fixescha* 'fix, harden'; *fix* 'fast, unmovable', *fixaziun* 'fixation'). Evidently the appearance of *-esch* as a way of avoiding stem alternation is a fact about the lexical entries for verbs (or about the rules of verbal inflection) and not a fact about the more abstract stems of those verbs.

2.8 An Excursus on a Defective Paradigm

Before concluding the discussion of stem alternations, it is interesting to consider the facts concerning the verb *dueir* 'must, should'. This is a classic example of a "defective" paradigm: in the Present Indicative, only two forms exist: 1pl. *duagn*, and 2pl. *duez*. All of the singular forms and the third person plural are replaced by forms of a different verb, *stueir*. This verb is itself suppletive: *ia stç* 'I must', *te stost* 'you (sg.) must', *el stç* 'he must'; *els ston* 'they must'; Subjunctive *ia stoptga*, etc. In addition to these missing forms of the Present Indicative, the Subjunctive of *dueir* is also missing, and replaced by forms of *stueir*. Other verbal forms are

built normally, however: Imperfect (*ia dueva*, etc.), Conditional (*ia duess*, etc.), Future (*ia duaro*, etc.); Gerund (*duond*), Past Participle (*duia, dueida*).enumerate

The pattern of defectiveness displayed by *dueir* is not at all arbitrary or random, of course. The evident generalization is that all and only those forms built on "unstressed" stem exist, and these are constructed in completely regular fashion. Where the "stressed" stem would be called for, the form in question is substituted by one taken from the paradigm of *stueir*. The history that has led to this state of affairs is fairly complex, involving not only phonological developments but also semantic competition between the reflexes of DEBERE and STOPERE (cf. Decurtins 1958, pp. 155ff. for discussion), but the synchronic situation is quite stable. It would seem that this defective paradigm could easily be repaired: even if no inherited "stressed" stem exists, surely there are other, phonologically similar verbs that could serve as the model for the creation of one. But in fact, all other verbs of the shape *C0ueir* are either substantially irregular (like *stueir* 'must, should') or use the stem extension *-esch* in the stem-stressed forms (e.g. *cueir* 'allow'; *flueir* 'flow'; *prueir* 'sprout', *scueir* 'begrudge', (*sa*) *snueir* 'shudder'). As a result, there is no availablemodel to provide a stressed stem.

The alternative, of course, would to use the stem extension *-esch* for the forms in question, giving e.g. **ia duesch* 'I must'. This is quite ungrammatical, however. One possible explanation is that *dueir* is unlikemost verbs in being amodal auxiliary, rather than a 'normal' lexical verb, and no other modal or other auxiliary verbs use *-esch* in their conjugation. Since the inflexion of auxiliary verbs are otherwise identical with those of lexical verbs (*modulo* individual irregularities of formation and suppletion), it is not clear how this proposal is to be implemented. The essential observation is clear, however: *dueir* is defective in having no "stressed" stem, and no valid model on which one can be constructed.

2.9 Conclusions: the Descriptive Properties of Alternating Stems

We conclude, therefore, that Surmiran shows a rich system of phonologically conditioned allomorphy, where the conditions governing the use of one allomorph or the other (but not the actual difference in shape) are a matter of the phonological environment in which the stem appears: specifically, the pattern of stress distribution associated with its affixes.

For the vast majority of bases in the language (perhaps all), at least two distinct stems must be given in the lexicon, with the choice based on the location of main stress in a given inflected form. Of course, this does not at all entail that the relation between the stems of a given base is entirely unstructured. In Surmiran as in other languages, the lexicon is not simply a list of forms, but also includes a set of regularities (cf. Jackendoff 1975, Anderson 1992, Anderson & Lightfoot 2002) that identify some lexical patterns as 'regular' and well integrated into the language and others as arbitrary and isolated. Nonetheless, from the point of view of the morphology and phonology of the language, Surmiran bases show two distinct stems whose distribution must be accounted for. It is to this that we turn next.

3. The Analysis of Stem Alternation

Although the stem alternation in Surmiran (and other Rumantsch languages) have their origin in strictly phonological processes, those have become opaque, and are now lost as phonological rules. The residual allomorphy, however, is governed by a strictly phonological condition: one stem or the other is chosen depending on the location of main stress in the output form.

The resulting systemdisplays a number of unusual properties when compared with other examples of phonologically conditioned allomorphy that have figured in the literature. For one thing, it is difficult to consider this to be a marginal pattern within the language: unlike the most widely cited instances of phonologically conditioned allomorphy, this pattern affects most content words in the language, and not just a small set such as a few affixes, or the 'mobile diphthongs' of Italian (van der Veer & Booij 2007).

The correct analysis of these alternations is not self-evident, and poses a problem for some approaches that have been adopted in other cases. Since it is stems, not affixes that

alternate in Surmiran, a sub-categorization solution of the sort advocated by Paster (2007) and Bye (2007a, 2007b) among others does not seem appropriate.

On the other hand, an approach that provides two phonological representations for each stem, and treats the choice between them as purely a matter of optimization based on phonological conditions (as in Kager 2007, Rubach & Booij 2001) could have trouble with the fact that the choice of the 'wrong' stem would apparently result in a perfectly well-formed word in many cases. Consider the verb *vurdar* 'watch', 3sg *varda* for example. If we simply listed the two stems as /vurd/ and /vard/, a constraint system would not appear to have any basis for preferring /vard/ to /vurd/ when stress falls on the stem, or for the opposite preference when stress falls on the desinence. Even supplementing the phonological constraints with a stipulated ranking of alternants (as in Bonet, Lloret &Mascaro 2007, Wolf 2007), the same issue will pose problems. Finally, conventionalOT solutions will have to deal with the fact that the defectiveness of *dueir* appears to consist in its having only one stem (the unstressed one). The fact that when stress falls on this stem the resulting form is excluded is just the kind of phenomenon that ranked, violable constraints are not equipped to express.

3.1 Alternating Verbs

Apparently, then, the analysis of stem alternations in Surmiran will require some additional assumptions. The overall OT framework of work such as Kager 2007 and Rubach & Booij 2001 seems to point in the right direction, but what is required is some way to characterize individual stem shapes in such a way that they will correctly be preferred in stressed *vs.* unstressed environments.

One way to achieve this would be to represent the stressed stems as already bearing metrical structure, in the form of a monosyllabic (trochaic) foot constructed over the final syllable of the stem. The unstressed stem, in contrast, would carry no metrical structure in its lexical representation. This marking would be supplemented by a stipulated preference for using the stressed stem if possible, and high ranking Faithfulness constraints requiring that input metrical structure be preserved. As a consequence, when an ending that itself attracts stress (such as the 1pl or 2pl Present Indicative) is combined with a base, the preferred stressed stem is tried first. The lexical stress in this form could be considered to violate a constraint against a clash with the stress of the ending; and rather than fail to realize the lexical stress (violating Faithfulness), the form would then be constructed on the basis of the unstressed stem instead.

Although this solution has only been sketched, it seems clear that it is not ideal. The notion of "Stress clash" that must be appealed to is not self-evident, since a desinential stress may not be strictly adjacent to the stem (in e.g. *vurd-a"ro*, **vard-a"ro* 'I will watch'), and some additional mechanism would be required to prevent the simple reduction of the lexical stress on the stressed stem /"vard/ from primary to secondary. In addition, the presence of lexical stress on these stems would obscure the distinction between genuinely unpredictable stress (in loan words, primarily) and stress which surfaces exactly when it conforms to the general principles of metrical structure in the language.

An alternative is suggested by the observation above in section 1 that in at least some instances, the vowels of unstressed syllablesmust be represented differently from those of stressed syllables. We have seen that the infinitive ending *-er* of the fifth class of verbs in Figure 1 must be represented as /ə/ as opposed to the /a/ of the first class or the /e/ of the second. Let us assume that this difference can also appear in stem forms.

Now suppose that for a verb like *l[ə]var* 'get up', 3sg *leva*, we represent the unstressed stem as /ləv/ and the stressed stem as /lev/. We could now say that constraints militate against the appearance of the reduced vowel [ə] in stressed syllables, and conversely, the appearance of a "full" vowel like [e] in unstressed syllables.

Suppose now we construct a verbal form from this base, where two stems are potentially available. If the form involves an ending such as that of the 1pl form *l[ə]vagn*, stress falls on the final syllable and the stem vowel is unstressed. As a result, the constraint system will prefer the stem /ləv/, whose vowel is acceptable in an unstressed syllable, to the

stem form /lev/ which would have a full vowel in unstressed position. With an ending such as the [-ə] of the 3sg form *leva*, in contrast, the stress falls on the stem vowel, and so a stem with a full vowel in this position will be preferred to one with the reduced vowel [ə].

The same solution is readily extended to the other verbs of the "alternating" type. We can note that in every one of these, the last vowel in the unstressed stem is either *a* ([ə]), *i* or *u*. Just as unstressed *a* (sometimes also written *e*) is phonetically distinct in quality from stressed *a*, the unstressed forms of *i* and *u* are laxer and shorter than the stressed forms. Suppose, then, that we allow these lax vowels /ɪ/ and /U/ to appear in the lexical representations of stems, and extend the constraints referred to above so that not only [ə] but also [ɪ] and [U] are preferred in unstressed syllables and dispreferred in stressed ones. We can now extend the treatment of *lavar/leva* to *durmeir* 'sleep', 3sg *dorma* and *fittar* 'finish', 3sg *fetta* by representing the stems of the former as {/dUrm/, /dorm/} and of the latter as {/fɪt/, /fet/}. The choice of the correct stem will then be made on the basis of a preference for stress to occur on an appropriate (full) vowel, and not on an inappropriate one ([ə, ɪ, U]).

In fact, all of the verbs of the 'Alternating' class can be accommodated by this analysis. Each will be assigned a lexical representation with two stems, differing as required in vocalism and in some cases (as discussed in earlier sections) in other ways as well. The crucial difference, however, lies in the character of the last (possibly the only) vowel of the stem: if this is one of [ə, ɪ, U], the corresponding form will be preferred where stress does not fall on it (and dispreferred if stressed), while if it is a full vowel the preference will work in the opposite direction.

3.2 'Regular' and 'Irregular' Verbs

What, then, of the "regular" verbs that appear to show no alternation? These are verbs such as *cantar* 'sing', 3sg *canta*; *chintar* 'calculate', 3sg *chinta*; *cuntschier* 'tinker', 3sg. *cuntscha*. All such verbs have a stem whose last vowel is *a, i* or *u*, and it is straightforward to assimilate them to the (somewhat larger) "alternating" class. These verbs also have two stem forms, but the difference between them is concealed by the orthography: on the present analysis, the bases of 'sing', 'calculate' and 'tinker' each have a pair of shapes: {/kənt/, /kant/}; {/kɪnt/, /kint/}; and {/kUnft/, /kuntf/}, respectively. Again, the correct choice is made by the constraints that associate full vowels with stressed syllables and reduced vowels with unstressed ones. It may seem counter-intuitive to assign two lexical stems to these verbs, where the only difference is one that seems to follow from phonetic vowel reduction. Recall, however, that in the course of history the relation between reduced and unreduced vowels has become increasingly opaque, and no longer the province of a productive phonological rule. The patterns [a]/[ə], [i]/[ɪ] and [u]/[U] are just very common instances of a relation between stressed and unstressed vowels that is much more general than this basic set.

The pattern of two lexical forms for verbal stems, then, is quite pervasive in Surmiran. It can even be discerned under the complexity of many of the 'irregular' verbs of the language, which take idiosyncratic shapes for some cells of their paradigm.

Consider the verb *pudeir* 'can, be able to', whose Present Indicative is given in (18).

(18)

1sg	ia poss
2sg	te post
3sg	el po
1pl	nous pudagn
2pl	vous pudez
3pl	els pon

When we look at other forms of this verb, we find that the Present Subjunctive, normally built on the stressed stem, has a regular paradigm (*ia poss, te possas, el poss; nous possan, vous possas, els possan* if we take the stemto be /pOs/. Forms like the Imperfect (*ia pudeva*, etc.), Conditional (*ia pudess*, etc.), Future (*ia pudaro*, etc.) and others that are normally built on the unstressed stem are similarly regular on the assumption that this is /pod/. For this verb, then, we need only list the 2sg, 3sg and 3pl Present

Indicative forms (/post/, /po/, /pon/) in addition to the stem pair {/pos/, /pod/}[6]. A number of verbs have idiosyncratic forms in the Present Indicative singular and 3pl; some have idiosyncratic stems in the Subjunctive (e.g., *saveir* 'know', Subjunctive *ia saptga*, with the same stem appearing in the Imperative) or some other category, but the basic outline of the two-stem system can be found in all except the most irregular of verbs (*esser* 'be' and *aveir* 'have').

3.3 Verbs in *-esch*

Does this mean that *every* verb in the Surmiran lexicon has two associated phonological shapes for its stem (together, perhaps, with some additional idiosyncratic forms)? If only a single stem were present, that would imply that the verb stem did not alternate at all - not even between [a] and [ə], [i] and [ɪ], or [u] and [ʊ], which could only happen if there were no difference between stem-stressed and unstressed forms. That that is in fact just what we saw for a large class of verbs, those with the stem extension *-esch* exactly in the forms that would otherwise have stem stress. These verbs only have unstressed forms (ignoring the stress that falls on *-esch* itself), and so require only a single stem.

As noted above, the extension *-esch* only appears in verbal inflection. If these verbs were listed with two stems (with and without *-esch*), that would suggest that any related de-verbal forms with stem stress might also show the extension, which does not happen. We need to say, then, that this is specific to verbal inflection, introduced by a rule such as (19).

(19) /X/ –> /XɛS/ [+VERB]

The place of this rule within the system of constraints requires some explanation. Since it expresses no property (Morphosyntactic or otherwise) of the verb, its application is not required by any constraint of the **Max** family (requiring that input material have some correspondent in the output). And since the phonological material it introduces is not present in the input, constraints of the **Dep** family will be violated when it applies. We might expect, then, this rule could never actually affect the output, unless this eliminated violations of some other constraint(s). And in fact, it will have that effect, exactly when a verb appears in a form with stem stress, but only has an unstressed stem. In exactly that case, the constraint against stress falling on one of the vowels [ə, ɪ, ʊ] would be violated, something that is prevented by the introduction of *-esch*. A verb that has a stressed as well as an unstressed stem has an alternative way of avoiding this violation, so *-esch* will never show up in the paradigm of such a verb. Stating a specific constraint ranking that would have this effect would require more precision with respect to the relevant constraints than is presently available, but the outline of the analysis proposed should be clear.

Verbs for which no inherited alternation is motivated, or where the actual alternation is forgotten or insecure, can thus be represented with only a single (unstressed) stem, and the result will be that these (and only these) will show the extension *–esch* exactly in their stem-stressed forms.

3.4 *Dueir*

We are still left with the problem of the defective verb *dueir* 'should, must'. This verb, it will be recalled, lacks precisely those forms where stem stress would be required, and replaces these with forms of another virtually synonymous verb, *stueir*. All forms built on the unstressed stem, however, are essentially regular. This suggests that the verb has only a single stem, the unstressed one (/dʊ/); but from what was said in the preceding section, that would appear to entail that the stem-stressed forms, rather than being lacking, would be built with an extended stem */dʊɛS/. Since that does not in fact happen, some other account is necessary. Note that it will not suffice to say simply that *dueir* 'borrows' a stressed stem from *stueir*, since it is the full range of irregular forms of this (*ia stç, te stost, el stç, els ston*; Subjunctive *ia stoptga*, etc.) that replace those of *dueir* where stress would fall on the stem.

[6] Note that the 3sg Present Indicative *poss* is completely regular on this assumption.

We might attempt to appeal to the status of *dueir* as a modal auxiliary. On that basis, we could perhaps modify rule (19) above to include a feature [−MODAL] in its structural description, to block its application to such verbs. There are two problems with this, however. First, there is no evidence for a feature like [±MODAL] in the morphology of Surmiran; modal auxiliaries in this language (as opposed, say, to English) inflect exactly like any other verbs, modulo individual lexical idiosyncrasies. And secondly, even if we blocked the application of (19) to the stem /dU/ on this (or some other) basis, what we would expect would be that the unstressed stem, as the only one available, would show up anomalously as the basis of the relevant stem-stressed forms. Optimality, that is, is well equipped to allow for the emergence of unexpected forms,but not at all set up to describe cases in which constraint violation leads to complete ungrammaticality.

Of course, the result is not in fact complete ungrammaticality, but rather the substitution of forms from the paradigm of another (essentially synonymous) verb for the problematic ones. This implies that some constraint(s) relevant to faithful lexicalization must in fact be integrated with the phonology, such that a phonological violation (in this case, having stress fall on a stem vowel that can only be unstressed) is repaired by choosing a different lexical item altogether. This suggestion has interesting implications for the overall architecture of grammar, and for the theory of defective paradigms in particular, but it is impossible to develop it further here.

4. Conclusions

Phonologists typically assume that variation in the shape of individual lexical items is either governed by purely phonological rules or essentially arbitrary, and correlated with morphological categories. Surmiran, in contrast, makes extensive use of variation that cannot be reduced to regularities of sound structure alone, but which is still correlated with phonological rather than morphological factors. If the analysis offered in this paper is on the right track, Surmiran (and the other Rumantsch languages for which much the same account could be offered) underwent major restructuring in the evolution from earlier Romance. The changes in question were not dramatic remodelings of the surface forms of words, but a much subtler change in the organization and status of inter-word relations. An original productive rule reducing vowels in unstressed syllables, something found in most languages with dynamic stress, became increasingly opaque through a combination of developments in individual words, other phonological processes, and borrowings from other languages (and from other forms of Rumantsch). Alternations between the shapes of stems that were originally predictable from the location of stress were largely preserved, but came to be predictable only by taking the identities of individual words into account — that is to say, lexicalized. The result is a system in which phonologically conditioned but lexically specified allomorphy, widely considered a rather marginal phenomenon in the languages of the world, has come to dominate most of the phonological variation in the language. As such, Surmiran serves as an interesting example for the phonologist or morphologist of how readily and completely a language can abandon the sort of simple and coherent internal organization (based on unitary lexical items and purely phonological variation) that we often assume as the ideal of linguistic structure.

The original phonological processes that gave rise to this situation (especially vowel reduction) still have a role to play in the description of Surmiran, but it is no longer that of phonological rules of the usual sort. Instead, these now have the status of regularities governing the internal structure of the lexicon. The two stems associated with a verb are not, in general, related arbitrarily in form, but instead fall into one or another of a limited set of patterns described by analogs of the original phonological rules. The lexicon is not, as often portrayed, simply an inert list of isolated forms, but instead a system of knowledge that forms part of a speaker's overall competence in the language (cf. Anderson & Lightfoot 2002, chap. 7). In order to capture the regularities that undeniably characterize the relations between stem shapes in Surmiran, linguists need to take more seriously the kind of observations made many years ago by Jackendoff (1975), and develop a more articulated theory of that knowledge than

exists today.

References

Anderson, Stephen R. 1992. *A–Morphous Morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. & Lightfoot, David W. 2002. The *Language Organ: Linguistics as Cognitive Physiology*. Cambridge: Cambridge University press.

Bonet, Eulalia, Lloret, Maria-Rosa & Mascaro, Joan 2007. Allomorph Selection and Lexical Preferences: Two Case Studies. *Lingua* **117,** 903-927

Bye, Patrick 2007a. Allomorphy — Selection, not Optimization, *in* M. Kramer, S. Blaho & P. Bye (eds), *Freedom of Analysis?* Berlin: Mouton de Gruyter.

Bye, Patrik 2002b. The Nature of Allomorphy: the Burushaski Evidence, *in* Tranel (2007).

Capeder, Reto 2006. *An lengia directa — Grammatica surmiran*. Coira: Leia Rumantscha. [Based on Vallader version by Jachen C. Arquint, and ultimately on an Italian original by Corrado Conforti and Linda Cusimano].

Carstairs, Andrew 1986. *Allomorphy in Inflexion*. London: Croom Helm.

Carstairs, Andrew 1988. Some Implications of Phonologically Conditioned Suppletion. *Yearbook of Morphology* **1**, 67-94.

Decurtins, Alexi 1958. *Zur Morphologie der unregelmδíigen Verben im Buóndnerromanischen*. Vol. X.I. of *Romanica Helvetica*, Bern: A. Francke Verlag.

Decurtins, Caspar 1982. *Rδtoromanische Chrestomathie*. Vol. X.I.: Sursettisch, Sutsettisch, Chur: Octopus Verlag. Reprint of 1914 edition.

Eichenhofer, Wolfgang 1989. *Diachronie des betonten Vokalismus im Bóndnerromanischen seit dem Vulgarlatein*. Vol. of *Romanica Rζtica*, Zurich: Juris Druck +Verlag AG.

Grisch, Mena 1939. *Die Mundart von Surmeir*. Vol. 12 of *Romanica Helvetica*, Paris: E. Droz.

Haiman, John & Beninca, Paola 1992a. The *Rhaeto-Romance Languages*. London: Routledge.

Haiman, John & Beninca, Paola 1992b. The *Rhaeto-Romance Languages*. London: Routledge.

Jackendoff, Ray S. 1975. Morphological and Semantic Regularities in the Lexicon. *Language* **51**, 639-671.

Kager, Rene 2007. Lexical Irregularity and the Typology of Contrast, *in* K. Hanson & S. Inkelas (eds), *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge MA: MIT Press.

Kamprath, Christine 1987. *Suprasegmental Structures in a Raeto-Romansh Dialect: A Case Study in Metrical and Lexical Phonology*. Doctoral Dissertation, University of Texas at Austin.

Kiparsky, Paul 2000. Opacity and Cyclicity. *The Linguistic Review* **17**, 351–367.

Kiparsky, Paul 2003. Finnish Noun Inflection, *in* D. Nelson & S. Manninen (eds), *Generative Approaches to Finnic and Saami Linguistics*. Stanford, CA: CSLI Publications, 109–161.

Lutta, C. Martin 1923. *Der Dialekt von Bergón*. Vol. 71 of *Beihefte zur Zeitschrift fór romanische Philologie*, Halle (Saale): Max Niemeyer.

Paster, Mary 2007. Phonologically Conditioned Suppletive Allomorphy: Cross-Linguistic Results and_eoretical Consequences, *in* Tranel (2007).

Prince, Alan & Smolensky, Paul 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

Rubach, Jerzy & Booij, Geert 2001. Allomorphy in Optimality Theory: Polish iotation. *Language* **77**, 26–60.

Signorell, F. (ed.) 1999. *Vocabulari surmiran-tudestg / Wφrterbuch Deutsch-Surmiran*. Coira: Departamaint d'educaziun digl Grischun.

Signorell, Faust,Wuethrich-Grisch, Mena & Simeon, Gion Pol 1987. *Normas Surmiranas*. Coira: Tgesa editoura cantunala per stampats e meds d'instrucziun.

Sonder, Ambros & Grisch, Mena 1970. *Vocabulari da Surmeir*. Coira: Leia Rumantscha.

Tranel, B. (ed.) 2007. *Understanding Allomorphy: Perspectives from Optimality Theory*. London: Equinox Press.

van der Veer, Bart & Booij, Geert 2007. Allomorphy in OT: The Italian Mobile Diphthongs, *in* Tranel (2007).

Wolf, Matthew 2007. Lexical Insertion Occurs in the Phonological Component, *in* Tranel (2007).

Aspects of Dialectal Variation in the Nominal Inflection of Greek: A Feature-Based Approach[1]

Vassilios Spyropoulos
University of the Aegean
spiropoulos@rhodes.aegean.gr

Konstantinos Kakarikos
University of Athens
k_kakarikos@hotmail.com

1. Introduction

This paper examines some aspects of dialectal variation in the nominal inflection of Greek. More specifically, we investigate the status of the [±human] feature as well as of definiteness in determining certain aspects of the nominal inflection, such as case syncretism, the feature inventory of formatives and the determination of inflectional classes. The aim of our investigation is to reveal the role of the relevant features in the dialectal subsystems and to propose a formal description of the facts in terms of a feature-based approach to nominal inflection.

Our analysis is couched within a feature-based approach to nominal inflection and it is based on the analysis of Greek nominal inflection proposed by Ralli (2000, 2005). This approach has the benefit of providing the basis for capturing in formal ways the systematic dialectal variation as well as the systematic relations holding among the systems of the various dialects and Standard Greek. Its advantage is that it allows for systematic dialectal variation to be explained as the parametric effect of the differences in the morphological function of certain features in the relevant subsystems. These parameters can be formulated as follows:

(1) a. A feature may be inflectionally active or not
b. A feature may acquire different specifications
c. A feature may belong to the feature inventory of a formative or not
d. A feature may trigger certain morphological operations or not
e. A feature may determine inflectional patterns or not

In this paper, we show that both definiteness and the [±human] feature are inflectionally active and control certain aspects of nominal inflection in a number of Asia Minor Greek dialects such as Cappadocian and Pontic Greek. In Standard Greek, these two features have no effect on the nominal inflection. However, their existence in the morphosyntactic structure of Greek is revealed by their effect on certain agreement rules. Thus, we propose that these two features are part of the feature inventory of the Greek nominal morphosyntax. In Standard Greek they are inflectionally inactive, in the sense that they do not affect the distribution of the inflectional endings and they do not condition any rules of inflection. As a result of a different parametric setting, these two features are inflectionally active in the nominal system of the aforementioned Asia Minor Greek varieties. Definiteness triggers *Differential Object Marking* and thus determines the distribution of the nominative and accusative formatives in the singular. The [± human] feature has a more pervasive effect, since it may determine the

[1] We would like to thank Anthi Revithiadou for useful discussions on the data and the analysis. We also thank the audience of the 6th Mediterranean Morphology Meeting and especially Angeliki Ralli for their comments. All errors are our own responsibility. Contact Information: Department of the Mediterranean Studies, University of the Aegean, 1 Demokratias Av., Rhodes 85100, Greece.

inflectional pattern of nouns and, thus, define inflectional classes, it controls the distribution of certain endings within an inflectional class and it can trigger case syncretism.

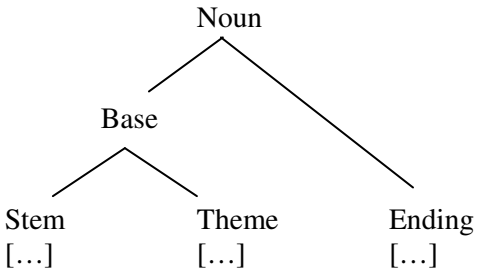
2. Background: standard Greek nominal inflection

Ralli (2000, 2003, 2005) has proposed a feature-based analysis of Standard Greek inflection, the basic premises of which we follow for the purposes of this paper. According to this analysis, Greek nominal inflection involves eight inflectional classes (IC), which are defined in terms of the allomorphic variation of the bases and the system of endings attached to them. Each ending is in fact an inflectional formative defined as a bundle of features with the relevant specification. These features belong to an independently defined set of features that determine the nominal morphology. In Standard Greek this set includes the features presented in (2) together with their potential values:

(2) Nominal features of Greek:
- Inflection Class (IC): 1-8
- Number: singular, plural
- Case: nominative, genitive, accusative
- Gender: masculine, feminine, neuter

In addition, we adopt Halle and Vaux's (1998) description of the morphological structure of nouns by means of a tripartite scheme which consists of the base (further divided into the stem and the thematic vowel or theme) and the ending:

(3)



Gender and Inflectional Class are assumed to be properties of the base, because of their inherent nature. On the other hand, number and case are considered to be properties of the ending. Thus, the formatives that manifest this terminal node are specified for these two features. The feature inventory of the formatives also includes a specification for Inflectional Class, which is checked against the relevant specification of the base. In this way the compatibility of a particular base with a particular set of formatives is formally accounted for (see also Alexiadou and Müller 2008). The feature inventory of formatives does not include the feature of gender, as it has been convincingly argued for by Rally (2003, 2005), because gender does not control the distribution of these formatives. In Greek, nouns of different gender may follow the same inflectional pattern, i.e. they may share the same set of formatives. This is illustrated in (4):

(4) *a΄nθropos* 'man' (MSC) - *ðia΄lektos* 'dialect' (FEM)

	SINGULAR	
NOM	a΄nθrop-os	ðia΄lekt-os
GEN	anθróp-u	ðialékt-u
ACC	a΄nθrop-o	ðia΄lekt-o
	PLURAL	
NOM	a΄nθrop-i	ðia΄lekt-i
GEN	anθróp-on	ðialékt-on
ACC	anθróp-us	ðialékt-us

In addition to these features there is evidence that the nominal morphosyntax of Standard Greek involves two other features, namely definiteness and [±human]. Definiteness is a phrasal property of the noun phrase which is marked syntactically mainly by the choice of the relevant article:

(5) a. MASCULINE
o a´nθropos énas a´nθropos
the man a man
b. FEMININE
i γinéka mja´ γinéka
the woman a woman
c. NEUTER
to peði éna peði
the child a child

The [±human] feature is a class feature related to animacy and it constitutes a lexical (semantic) property of the stem. The formal postulation of this feature relies on its role in determining the resolution of gender conflict in conjunction, as this is revealed by the participation of the conjoined structure in agreement constructions (Holton *et al.* 1997, Chila-Markopoulou 2003, Spyropoulos 2005):

(6) When the conjunct DPs are [+human] → Adj[masculine]
a. o pe´tros ke i mari´a i´ne fi´li
the Petros-MSC.NOM and the Maria-FEM.NOM are friends- MSC.PL.NOM
'Petros and Maria are friends'
b. i γine´kes ke ta peðja´ na i´ne e´timi
the women-FEM.NOM and the children-NT.PL SUBJ are ready-MSC.PL.NOM
'The women and the children should be ready'

(7) When the conjunct DPs are [–human] → Adj[neuter]
a. o komunismo´s ke i ele´fθeri aγora´
the communism-MSC.NOM and the free market-FEM.NOM
i´ne asimvi´vasta
are incompatible-NT.PL.NOM
'Communism and free market are incompatible'
b. i ðro´mi ke i plati´es i´tan γema´ta ko´smo
the streets-MSC.NOM and the squares-FEM.NOM were full-NT.PL.NOM people
'The streets and the squares were full of people'

In situations where two [+human] DPs of different gender are conjoined the resulting phrase acquires the masculine value for the gender feature. This is evident in examples such as those in (6): the adjectival predicate which obligatorily agrees with the conjoined phrase is specified as masculine. On the contrary, when two [–human] DPs of different gender are conjoined, the resulting phrase is specified as neuter, as evident by the neuter specification of the agreeing adjectival predicate (7).

Despite their existence in the nominal morphosyntax, [±human] and definiteness cannot be considered inflectional features in Standard Greek, because they do not determine any aspect of the nominal inflection. Nevertheless, these two features seem to affect the inflection of nouns in various ways in various dialectal varieties. In what follows, we present evidence from the nominal inflection of two dialect groups of Asia Minor Greek, namely Cappadocian and Pontic Greek,[2] which shows that the [±human] feature can (a) control the choice of certain plural formatives, (b) determine the direction of case syncretism in plural and (c)

[2] Cappadocian Greek data are drawn from the invaluable grammatical description by Dawkins (1916). Pontic Greek data are drawn from Oikonomides (1958) and Papadopoulos (1955).

define inflectional classes. Furthermore, we propose an analysis of *Differential Object Marking* in Cappadocian Greek which illustrates how definiteness determines the realization of case of DP-objects in these varieties.

3. Formatives specified as [±human]: Cappadocian Greek

In Cappadocian Greek the [±human] feature has undertaken the role of gender, which has been convincingly argued to be absent from the nominal system of the dialect (Dawkins 1916, Janse 2004, Spyropoulos and Kakarikos 2007).[3] Significantly, the [±human] feature is able to determine the distribution of certain plural formatives. In the variety of Delmeso, nouns of Greek origin ending in *-as*, such as *papa´s* 'priest' and *kerata´s* 'snail' take different formatives in the plural:

(8)

	SINGULAR	
NOM	papa´-s	kerata´-s
GEN	papað-ju´	keretað-ju´
ACC	papa´-∅	kerata´-∅
	PLURAL	
NOM	papa´ð-es	kerata´ð-ja
GEN	-	-
ACC	papa´ð-es	kerata´ð-ja

These nouns belong to the same inflectional class, because they exhibit the same pattern of base allomorphy and they take the same set of singular formatives. Plural formatives are grammatically conditioned and their choice depends on the [±human] property of the stem. Thus, the *-ja* formative is a special plural formative selected by non-human nouns. On the contrary, the plural formative *-es* has a wider distribution in the variety and it is used for both human and non-human nouns in other inflectional classes:

(9) *ne´ka* 'woman', *tzi´na* 'sparrow', *ni´fi* 'bride', *psi´i* 'crub'

	SINGULAR			
NOM	ne´ka-∅	tzi´na-∅	ni´fi-∅	psi´i-∅
GEN	ne´ka-s	tzi´na-s	ni´fi-s	-
ACC	ne´ka-∅	tzi´na-∅	ni´fi-∅	psi´i-∅
	PLURAL			
NOM	ne´k-es	tzi´n-es	ni´f-es	psi´-es
GEN	-	-	-	-
ACC	ne´k-es	tzi´n-es	ni´f-es	psi´-es

We therefore conclude that *-ja* is specified as [–human], whereas *-es* is a default plural formative. Interestingly, the same phenomenon appears in other Asia Minor Greek dialects, such as the one of Pharasa (Dawkins 1916, Andriotis 1948):

(10) *papa´*s 'priest' vs. *prakana´*s 'beetle'

	SINGULAR	
NOM	papa´-s	prakana´-s
GEN	papa´-∅	prakana´-∅
ACC	papa´-∅	prakana´-∅
	PLURAL	
NOM	papa´ð-es	prakana´ð-ja
GEN	-	-
ACC	papa´ð-es	prakana´ð-ja

[3] In the grammatical descriptions of the references cited in the text the [±human] feature is referred to as animacy.

4. The [±human] feature and case syncretism in plural: Cappadocian and Pontic Greek

Evidence that the [±human] feature is inflectionally active comes from the fact that it is able to affect the inflection of certain forms by triggering case syncretism. In the following subsections we examine the effects of a case syncretism rule that applies in both Cappadocian and Pontic Greek. The rule is formulated as follows:

(11) [nom] → [acc] / [___, -animate, plural]

4.1. Cappadocian Greek

In the Cappadocian Greek varieties of Delmeso, Potamia and Malakopi, case syncretism between nominative and accusative takes place in the plural of inflectional patterns which employ distinct formatives for each of the two cases. More specifically, nouns of Greek origin ending in *-os*, such as *a΄θropos* 'human', *tzoba΄nos* 'shepherd', take the *-i* formative for the nominative plural and the *-us* formative for the accusative plural:

(12)

	SINGULAR	
NOM	tzoba΄nos	a΄θropos
GEN	tzoba΄nu	aθro΄pu
ACC	tzoba΄no	a΄θropo
	PLURAL	
NOM	tzoba΄ni	a΄θropi
GEN	-	-
ACC	tzoba΄nus	aθro΄pus

However, when such a noun is non-human, nominative becomes syncretic to accusative and it is expressed by the formative *-us*:

(13) *to΄pos* 'place', *mi΄los* 'mill', *γa΄mos /qa΄mos* 'wedding'

PLURAL	*Delmeso*		*Potamia*	*Malakopi*
NOM	to΄pus	mi΄lus	γa΄mus	qa΄mus
ACC	to΄pus	mi΄lus	γa΄mus	qa΄mus

4.2. Pontic Greek

In Pontic Greek [–human] nouns are subject to the same case syncretism in plural. Significantly, this syncretism applies to all [–human] masculine and feminine nouns,[4] irrespective of the inflectional class they belong to. Thus, non-human masculine nouns ending in *-os* take the accusative formative *-us* for both the nominative and the accusative plural:

(14) *stavro΄s* 'cross' vs. *a΄nθropos* 'man'

	PLURAL	
NOM	stavr-u΄s	anθro΄p-i
ACC	stavr-u΄s	anθro΄p-us

Non-human masculine nouns ending in *-as* (15) as well as non-human feminine nouns ending in *-a* (16a) and *-i* (16b) take the accusative formative *-as* for both the nominative and the

[4] Unlike Cappadocian Greek, gender exists as a feature in the nominal morphosyntax of Pontic Greek with a tripartite specification (masculine, feminine, neuter) and it controls the choice of the definite article in a similar way as in Standard Greek. An interesting phenomenon attested in Pontic Greek is that [–human] nouns neutralize their gender specification in the plural and behave as neuters; this is evident by the choice of the neuter form of the definite article in these cases.

accusative plural. It should be noticed that whereas the corresponding human masculine nouns in the plural take the distinct *-es* and *-as* formatives for nominative and accusative respectively, the human feminine nouns take the syncretic nominative plural formative *-es* for both the nominative and the accusative:

(15) masculines in *-as*: *mi´nas* 'month' vs. *a´ndras* 'man'

	PLURAL	
NOM	mi´n-as	a´ndr-es
ACC	mi´n-as	a´ndr-as

(16) a. feminines in *-a*: *θa´lasa* 'sea' vs. *θeγate´ra* 'daughter'

	PLURAL	
NOM	θa´las-as	θeγate´r-es
ACC	θa´las-as	θeγate´r-es

b. feminines in *-i*: *ni´ki* 'victory' vs. *batzi´* 'virgin'

	PLURAL	
NOM	ni´k-as	batzi´ð-es
ACC	ni´k-as	batzi´ð-es

5. The [+human] feature defines the inflection pattern: Cappadocian Greek

The Cappadocian Greek variety of Ulaghatsh is characterized by a radical reanalysis of the nominal system under the pressure of the agglutinative typological pattern of Turkish. As a consequence, most of the inflectional patterns became agglutinative. Interestingly, nouns of Greek origin ending in *-os* split into two inflectional patterns depending on whether they denote human or non-human entities.[5] Thus, human nouns, such as *xeri´fos* 'man', follow a synthetic inflectional pattern, whereas non-human nouns, such as *li´kos* 'wolf', follow the general agglutinative inflectional pattern:

(17)

	SINGULAR	
NOM	xeri´fo-s	li´kos-∅
GEN	xerif-ju´	li´kos-ju (> li´kozju)
ACC	xeri´fo-s	li´kos-∅
	PLURAL	
NOM	xeri´f-ja	li´kos-ja (>li´kozja)
GEN	-	-
ACC	xeri´f-ja	li´kos-ja (>li´kozja)

We can safely conclude that human nouns ending in *-os* form their own inflectional class with a distinct inflectional pattern. In this way, the [+human] feature defines a certain inflectional class. The defining characteristic of this inflectional pattern is the syncretic *-s* formative for the nominative and accusative singular. These facts led to the reanalysis of this *-s* singular formative as an [+human] marker, as revealed by the spread of its use in examples like *pana´jas* (< *pana´ja*) 'Holy Mother', *sofja´s* (< *sofja´*) 'Sophia' (Kesisoglou 1951: 31).

6. Interim summary: the [±animate] feature

[5] A similar situation is also attested in the Cappadocian Greek variety of Axos (see Dawkins 1916 and Mavroxalyvidis and Kesisoglou 1961).

The evidence presented above show that the feature [±human] is active in the nominal inflection of Cappadocian and Pontic Greek. This [±human] feature is a class feature related to animacy. It conditions rules that affect the feature constitution of the ending terminal node and it may determine the inflectional pattern of an inflectional classes. What is more, certain formatives may be specified for this [±human] feature in a similar way as for the inflectional class feature. This specification determines the distribution of these formatives in accordance with the [±human] specification of the stem.

7. Case and definiteness: Differential Object Marking in Cappadocian Greek

Another characteristic of some Cappadocian Greek varieties is that they exhibit *Differential Object Marking* (DOM) with respect to definiteness. When a noun that follows an inflectional pattern with distinct nominative and accusative case formatives in the singular is the object of the clause, it appears in the nominative form instead of the accusative when it is interpreted as indefinite (Dawkins 1916, Janse 2004, Spyropoulos and Tiliopoulou 2006). In the following examples from the Delmeso and Potamia varieties the generic and indefinite objects appear in the nominative case form with the characteristic *-s* formative, instead of the expected accusative forms *γa⇔mo* 'marriage-SG.ACC' and *laγo΄* 'hare-SG.ACC':

(18) *Potamia (Dawkins 1916: Potamia 1, p.456: 1)*
i⇔stera pi⇔kan γa⇔mos
afterwards made-3PL marriage-SG.NOM
'after that they got married'

(19) *Delmeso (Dawkins 1916: 94)*
Δe⇔ke e⇔na laγo⇔s
hit-3SG a hare-SG.NOM
'He struck a hare'

Janse (2004) follows the traditional description by Dawkins (1916) and argues that the *-s* formative has been reanalyzed as an indefiniteness marker. According to his analysis, forms with *-s* are cited as indefinite accusative singular in the relevant nominal paradigms. The main objection against such an analysis is that the formative *-s* seems to predominately mark the nominative case of subject function irrespective of definiteness/indefiniteness. This is evident in example (20) where the noun *e⇔ljos* 'sun' which is the subject of the clause, takes the characteristic formative *-s* although it has a definite reading:

(20) *Delmeso (Dawkins 1916: Delmeso 1, p.312: 14-15)*
e⇔ljos ma⇔vrosen to pro⇔sopo
sun-NOM.SG blackened-3SG the-ACC.SG face-ACC.SG
'the sun blackened my face'

On the basis of this evidence, Spyropoulos and Tiliopoulou (2006) argue that such constructions reveal an instance of direct structural interference from Turkish. According to their analysis Cappadocian Greek DOM involves a situation where the nominative form substitutes for the accusative one resulting in the indefinite reading. Cappadocian Greek DOM is therefore formulated in terms of feature substitution as follows:

(21) Cappadocian Greek DOM
[acc] → [nom] / [___, -definite]

Nevertheless, such a formulation goes against the mechanism of DOM itself. In DOM situations accusative is not substituted for by nominative but rather by an absolutive form which coincides with the nominative (Aissen 2003). This absolutive form is marked with the

null formative -∅, which is the default morpheme. However, in Cappadocian Greek the situation is reversed, since the null formative -∅, which is the default morpheme (Spyropoulos and Kakarikos 2007), marks the accusative case. Thus Cappadocian Greek DOM involves a situation in which the [–definite] specification of the DP neutralizes the distinction between accusative and nominative and allows for the insertion of the more specified nominative formative *-s*.

Spyropoulos and Kakarikos (2007) propose an analysis which is based on the feature decomposition of case (Halle and Vaux 1998, McFadden 2004) combined with a division of labour between narrow syntax and morphology as far as case assignment is concerned (Español-Echevarría and Ralli 2000, Spyropoulos in progress). They assume that morphological case (m-case) is set apart from syntactic (abstract) case (s-case) (Marantz 1992, Schütze 1997, McFadden 2004, Bobaljik 2006) in the sense that syntax does not determine m-case, i.e. the surface case specification, but rather those aspects of case that are related to the licensing of DPs (Español-Echevarría and Ralli 2000, Spyropoulos in progress). The full case feature specification is determined in morphology in terms of case domains and hierarchies (McFadden 2004, Bobaljik 2006). Case is therefore decomposed in a number of case features which represent its morphological and syntactic aspects:

(22) Case features (Halle and Vaux 1998, McFadden 2004)
- [±structural]: [+structural] is assigned to DPs on the basis of their position in the syntactic structure (Halle and Vaux 1998).
- [±oblique]: [+oblique] is assigned to DPs by certain functional heads, such as *vAPPL*, P etc. (McFadden 2004).
- [±genitive]: [+genitive] is a lexically specified case feature assigned by certain functional or lexical heads (McFadden 2004, Español-Echevarría and Ralli 2000).
- [±inferior]: [+inferior] is assigned to DPs due to the presence of a higher argument within the case-domain (McFadden 2004, Bobaljik 2006).

Each case is therefore the result of the appropriate specification of the relevant case features. Thus, nominative and accusative are defined as follows:

(23) Feature decomposition of nominative and accusative
- a. nom = [+case, –oblique, –inferior]
- b. acc = [+case, –oblique, +inferior]

Nominative and accusative are syntactically identical, being the *par excellence* structural cases. This means that the [inferior] feature which differentiates them is specified in the syntax-morphology interface as a result of the position of the relevant DP in its case domain. Thus, syntax provides the same terminal node for nominative and accusative, which is given in (24):

(24) [+case, +structural, –oblique, –genitive, α inferior]

In Cappadocian Greek, DOM is the effect of a rule which negatively specifies the [inferior] feature in a [–definite] environment:

(25) [α inferior] → [–inferior] / [___, –definite]

After the application of this rule the terminal nodes for nominative and accusative are defined as follows:

(26)
- a. nominative
  [+case, –oblique, –inferior]
- b. accusative definite
  [+case, –oblique, +inferior | +definite]
- c. accusative indefinite
  [+case, –oblique, –inferior | –definite]

The *-s* formative, being the characteristic formative for the nominative, is specified as [–inferior]. Thus, Cappadocian Greek DOM is analyzed here as an operation on case features, according to which the [–definite] specification is able to determine the case feature specification of the terminal node of the ending. By doing so, it creates the appropriate environment for the insertion of the *-s* formative in situations where this is not expected to occur. Thus, indefiniteness controls the distribution of the *-s* formative, as implied by Janse's and Dawkins' descriptions, but it does so in an indirect way and not totally. The formative *-s* is not specified as [–definite]; this specification is carried only by the terminal node that hosts this formative, which gives the impression that *-s* is an indefiniteness marker. In fact, the [–definite] feature only forces the negative specification of the [inferior] feature in the terminal node, which in turn results in the insertion of the *-s* formative which carries the [–inferior] specification.

What is important for our discussion is that (in)definiteness has the ability to trigger a specific rule that affects the inflectional manifestation of case. In this sense it can be argued that (in)definiteness is inflectionally active in these dialectal varieties.

8. Summary: Conclusions and extensions

The examination of the data from Cappadocian and Pontic Greek has revealed that definiteness and the [±human] feature determine certain aspects of their nominal inflection, a fact that differentiates these dialect groups from Standard Greek. More specifically, the two features were shown to determine the realization of case by triggering rules that affect the feature content of the terminal node of the ending. The [±human] feature was additionally shown to behave as a class feature and to be able to determine the inflectional pattern of certain nouns and to control the distribution of certain plural endings. The complexity of the dialectal facts was accounted for by means of (a) morphological operations which affect the feature content of the terminal nodes in the morphological structure and (b) an inventory of formatives with the relevant feature specification (see also Spyropoulos and Kakarikos 2007). Such a feature-based analysis allows us to explain some very important aspects of dialectal variation in the nominal system of Greek as the result of micro-parametric variation with respect to the status and function of certain features.

References

Aissen, Judith. 2003. Differential object marking: iconicity vs. economy. *Natural Language and Linguistic Theory* 21: 435-483.

Alexiadou, Artemis and Gereon Müller. 2008. Class features as probes. Ms. To appear in Asaf Bachrach and Andrew Nevins (eds.), *Inflectional Identity*. Oxford: Oxford University Press, 101-155.

Andriotis, Nikolaos P. 1948. *Το Γλωσσικό Ιδίωμα των Φαράσων [The dialect of Pharasa]*. Αθήνα: Ίκαρος.

Bobaljik, Jonathan David. 2006. Where's Phi? Agreement as a postsyntactic operation. Ms. University of Connecticut.

Chila-Markopoulou, Despina. 2003. Γένος και συμφωνία στη Νέα Ελληνική [Gender and agreement in Modern Greek]. In Anna Anastasiadi-Simeonidi, Angeliki Ralli and Despina Chila-Markopoulou (eds.), *Το Γένος [Gender]*. Αθήνα: Πατάκης, 100-132.

Dawkins, Richard McGillivray. 1916. *Modern Greek in Asia Minor: A Study of the Dialects of Silly, Cappadocia and Pharasa with Grammar, Texts, Translations and Glossary*. Cambridge: CUP.

Español-Echevarría, Manuel and Angeliki Ralli. 2000. Case mismatches in Greek: Evidence for the autonomy of morphology. *Acta Linguistica Hungarica* 47: 179-203.

Halle, Morris and Bert Vaux. 1998. Theoretical aspects of Indo-European nominal morphology: The nominal declensions of Latin and Armenian. In J. Jasanoff *et al.* (eds.), *Mír Curad*: *Studies in Honor of Calvert Watckins.* Innsbruck: Innsbrucker Beiträge zur Sprachenwissenschaft, 223-240.

Holton, David, Peter Mackridge and Irene Philippaki-Warburton. 1997. *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.

Janse, Mark. 2004. Animacy, definiteness and case in Cappadocian and other Asia Minor Greek dialects. *Journal of Greek Linguistics* 5: 3-26.

Kesisoglou, Iordanis I. 1951. *Το Γλωσσικό Ιδίωμα του Ουλαγάτς [The dialect of Ulaghatsh]*. Athens: French Institute of Athens.

Marantz, Alec. 1992. Case and licensing. *ESCOL '91*: 234-253.

McFadden, Thomas. 2004. *The Position of Morphological Case in the Derivation: A Study on the Syntax-Morphology Interface*. Ph.D. Thesis, University of Pennsylvania.

Mavroxalyvidis, Georgios and Iordanis I. Kesisoglou. 1960. *Το γλωσσικό ιδίωμα της Αξού [The dialect of Axos]*. Athens.

Oikonomidis. D. H. 1958. *Γραμματική της Ελληνικής Διαλεκτου του Πόντου [Grammar of Pontic Greek]*. Εν Αθήναις.

Papadopoulos, Anthimos. 1955. *Ιστορική Γραμματική της Ποντικής Διαλέκτου [Historical Grammar of Pontic Greek]*. Αθήναι.

Ralli, Angeliki. 2000. A feature-based analysis of Greek nominal inflection. *Glossologia* 11-12: 201-228.

Ralli, Angeliki. 2003. Ο καθορισμός του γραμματικού γένους στα ουσιαστικά της νέας ελληνικής [The determination of grammatical gender of nouns in Modern Greek]. In Anna Anastasiadi-Simeonidi, Angeliki Ralli and Despina Chila-Markopoulou (eds.), *Το Γένος [Gender]*. Αθήνα: Πατάκης, 57-99.

Ralli, Angeliki. 2005. *Μορφολογία [Morphology]*. Αθήνα, Πατάκης.

Schütze, Carson. 1997. *Infl in Child and Adult Language: Agreement, Case and Licensing*. Ph.D. Thesis, MIT.

Spyropoulos, Vassilios. 2005. Some remarks on conjunction and agreement in Greek: Implications for the theory of agreement. Paper presented at the *7th International Conference on Greek Linguistics*, University of York, 8-10 September 2005. (To appear in proceedings)

Spyropoulos, Vassilios. In progress. Case conflict in Greek free relatives: Case in syntax and morphology. Ms. University of the Aegean.

Spyropoulos, Vassilios and Konstantinos Kakarikos. 2007. A feature-based analysis of Cappadocian Greek nominal inflection. Paper presented at the *3rd International Conference on Modern Greek Dialects and Linguistic Theory*, University of Cyprus, Nicosia, 14-16 June 2007. (To appear in proceedings)

Spyropoulos, Vassilios and Marianna Tilipoulou. 2006. Definiteness and case in Cappadocian Greek. In Mark Janse, Brian Joseph and Angela Ralli (eds.), *Proceedings of the 2nd International Conference on Modern Greek Dialects and Linguistic Theory*. Patra: University of Patras, 366-378.

On Deriving Polarity Effects

Antje Lahne[1]

December 7, 2007

In this paper, I argue for a new, theory-neutral approach to polarity effects resting upon the idea that the choice of marker for a given feature specification is determined by the choice of marker for a minimally different specification. In paradigms instantiating polarity effects, the matching of morpho-syntactic and phonological features proceeds by natural class based-rules, but is partly overriden by two principles, Discreteness of Environment and Minimality, the latter of which is an independently motivated assumption underlying syntactic derivations, and arguably a basic property of language in general.

1 Background

The concept of natural classes is one of the most basic concepts of linguistic description. However, there are data that seem to escape analyses making use of natural classes. The arguably most striking example are polarity effects in inflectional paradigms, that is, "complementary distributions of inflectional markers in such a way that syncretism constitutes itself in mirror-image identity of non-contiguous paradigmatic cells" (Baerman et al. 2005:104). These effects can be found in Old French masculine ostems (Rheinfelder 1976), where the syncretism pattern that occurs is that of a chessboard: nominative singular syncretises with the accusative plural, and the nominative plural syncretises with the accusative singular (table 1). Likewise, in the Somali definite article (Saeed 1999:112), the masculine singular and the feminine plural are syncretic, and the masculine plural and the feminine singular are syncretic (table 2).

These "chessboard distributions" do not seem to be analysable by having recourse to natural classes (cf. Baerman et al. 2005), as the two morphosyntactic feature specifications associated with one phonological form do not have a common value for any given feature (or do not have a common distinctive structure for any given dimension), no matter which particular feature representation is chosen, and therefore cannot be referred to by means of standard natural class-based rules.

Table 1

		SG	PL
Old French	NOM	-s	-ø
Masculine o-stems	OBJ	-ø	-s

Table 2

		SG	PL
Somali	FEM	-ta	-ka
Definite article	MASC	-ka	-ta

2 On the Systematicity of Polarity Effects

Chess board distributions can be treated by morphological theories in two different ways:

(I) Natural classes are taken to be the sole underlying concept. Consequently, polarity effects come as a completely accidental pattern (i.e., they are ignored by the morphological theory).

(II) The systematicity of chess board distributions is integrated into the morphological theory. There are basically two ways of accomplishing this:

[1] For comments and discussion, I would like to thank Gereon M¨uller, Andrew Nevins, Catherine Taylor, Jochen Trommer, and audiences at WOTM 3 (Leipzig, June 2007) and MMM 6 (Vathy/ Ithaka, September 2007). This research was supported by the 2007/08 Research Grant for the State of Saxony.

a. The theory defines new natural classes such that seemingly ununifiable specifications come out as forming a class (e.g. exactly because of their property of being un-unifiable);
b. The matching of morpho-syntactic and phonological features is modeled in such a way that it proceeds according to natural class based-rules, but can be overriden by other morphological principles or processes.
Section 2.1 discusses theories of type (I). Two powerful techniques of type (IIa) are discussed in section 2.2. To my knowledge, there is no type (IIb) theory of polarity effects up to now (though Weißer (2007), an analysis of Lshaped syncretisms, can perhaps be extended to account for polarity effects). In this paper I would like to propose a new analysis of the data which makes use of strategy (IIb).

2.1 Polarity Effects as an Accidental Pattern

It seems at first sight that polarity effects are not a systematic, but an accidental pattern. The seeming syncretisms in the paradigms in tables 1 and 2 could therefore be seen as a case of marker homonymy, as shown in (1) for the Old French example[2].

(1) Vocabulary items for Old French noun inflection:
/-s/ ↔ [+obj +pl +m +x[3]]
/-s/ ↔ [–obj –pl +m +x]
/-ø/ ↔ [ ]

The alternative – if polarity effects are understood as an accidental pattern – is to assume that impoverishment rules, or rules of referral, respectively (Halle and Marantz 1993; Noyer 1997; Bobaljik 2002; Stump 1993, 2001) are at work in these data. The basic idea of this device is that a marker which is expected to appear in the context of a certain morphosyntactic feature specification is replaced by a different marker due to a feature modification operation which is prioritised in the sense that it applies before matching with phonological features takes place. As a result, when the phonological features are inserted, the set containing the modified feature(s) is matched with a previously unexpected, potentially underspecified marker. (2) shows the prioritised rule for the Somali definite article in two notational variants: (2a) is a possible impoverishment rule from the Distributed Morphology framework (multiple non-standard feature-changing impoverishment as proposed in Noyer 1998), and (2b) is a rule of referral from the framework of Word-and-Paradigm Morphology.

(2) a. Impoverishment rule for the Somali definite article:
[+f +pl] > [–f –pl] / [+art +def]
b. Rule of referral for the Somali definite article[4]:
I{[+f +pl]} → I{[–f –pl]} / [+art +def] .

The effect of this prioritised rule is that whenever the system encounters the morphosyntactic context [+f +pl], this context is changed to [–f –pl]. The vocabulary items for the Somali definite article are given in (3).

(3) Vocabulary items for the Somali definite article:
/-ka/ ↔ [–f –pl]
/-ta/ ↔ [ ]

The a priori expected marker for the context [+f +pl] is /-ta/. However, as the features [+f +pl] of the head (or cell) have been overwritten by [–f –pl], the morphosyntactic context now matches the specification for /-ka/, thus /-ka/ is inserted in the context [+f +pl].
A difference between impoverishment rules and rules of referral is that impoverishment rules are conceived as being more restrictive inasfar as they are either deletions of features or changes of values from [+x]→[–x], or [ ]x]→[x] (i.e., a retreat to the general case causing

[2] Throughout the paper, the association of markers and forms is given in the notation used in Distributed Morphology. The analysis is however theory-neutral; it can be implemented in any mophological theory.
[3] x= class feature defining o-stems.
[4] I{+x} represents the inflection marker for context {+x}, and ! represents "is replaced by".

the insertion of a less specific marker; Halle and Marantz 1993, 1994; Noyer 1998), whereas rules of referral by definition modify features without restrictions.

2.2 Polarity Effects as a Systematic Pattern

As has been shown by Baerman (2007), polarity effects are far from being a rare phenomenon in the world's languages; they occur e.g. in Hebrew gender marking, Old Church Slavonic neuter noun and adjective suffixes, voicing reversal in Luo, Estonian partitive endings, Nehan definite articles, Tubatulabal aspect marking, and tense marking in Trique. It therefore seems to be reasonable to model morphological theory in such a way that polarity effects follow as a systematic pattern. Two solutions have been previously suggested.

One possible way of capturing polarity effects as a systematic pattern is to establish natural classes by abstracting over feature values (α-notation, Chomsky and Halle 1968). (4) shows the vocabulary items for the Somali definite article using α-notation.

(4) Vocabulary items for Somali definite article:

/-ta/ ↔ [_f –_pl]

/-ka/ ↔ [ ]

This solution has been modelled as a powerful device of forming natural classes. However, as is shown in (5) for the Somali definite article, if the variable in the insertion rule is resolved, then α-notation turns out to be a mere notational variant for homonymous markers.

(5) /-ta/ ↔ {[+f –pl], [–f +pl]}

A second possible solution, proposed by Bejar and Hall (1999), is to assume a new form of underspecification by which seemingly un-unifiable specifications come out as forming a natural class. The basic idea of this geometrybased approach is that natural classes are defined by the degree of featural markedness (where markedness is defined in terms of presence or absence of structure in a feature-geometrical representation). Let me sketch this approach for the Old French data. The analysis is based on the following case and number decomposition[5]:

(6) Old French: feature geometry

Singular:	Plural:	Nominative:	Objective:
ind	ind	case	case
|	|	|	|
min	group	subj	obj

The combination of case and number features yields the featural representations of the four contexts:

(7) Old French: case-number combinations

NOM SG:	NOM PL:	OBJ SG:	OBJ PL:
ind case	ind case	ind case	ind case
	|	|	| |
	min	obl	min obl

The specifications that form a natural class in this approach are obj sg and nom pl in that both are specified for the dimension node on one branch ([ind] in the case of obj sg, and [case] in nom pl), and the dimension node plus a further node on the other branch ([case|obj] in obj sg, and [ind|group] in nom pl). The vocabulary items for Old French are given in (8).

(8) Vocabulary items for Old French:

/-ø/ dim1 dim2 | /-s/ ↔ [ ]

[5] Abbreviations in this paper: ind=individuation; min=minimal; part=participant; spk=speaker; addr=addressee; subj=subject; obj=object; obl=oblique.

X

In this case, the zero marker is the most specific vocabulary item. It is inserted in nom pl and obj sg contexts. /-s/ is the elsewhere marker and thus matches all other contexts (nom sg, obj pl). The zero marker cannot be inserted in the context nom sg: though on one branch this context is specified for a dimension node only ([ind] or [case], respectively), it is not specified for a dimension node plus a further node on the other branch. Note that the system crucially relies on the assumption that underspecification is not at work in vocabulary insertion operating on degrees of structural markedness; thus, the zero marker cannot be inserted in obj pl contexts. This assumption is potentially incompatible with the need for full underspecification (i.e., the specification of /-s/ as the elsewhere marker). This dilemma can possibly be solved by assuming that the grammar allows for a mixed system of both feature-based and markedness-based insertion rules, where the former are subject to underspecification, and the latter must be fully specified for the context where they can be inserted.

2.3 A Motivation for Polarity Effects

Up to now, a single argument has been given in favour of a systematic analysis of polarity effects: that it is not unknown among the world's 6000 languages. The desideratum however is to recover a deeper motivation for why this seemingly random pattern is actually systematic. The starting point for such a motivation is the empirical observation that the matching of phonological forms with morphological or syntactic specifications can be accomplished by embarking on one of three basic strategies:

(A) Targeting minimal ambiguity with maximal formal inventory (i.e., yielding no syncretisms at all);

(B) Making use of syncretisms in natural classes;

(C) Targeting minimal ambiguity with minimal formal inventory (i.e. yielding evenly distributed syncretisms).

Strategy (A) occurs e.g. in Greek aorist or Russian present indicative, while strategy (B) seems to be favoured in the world's languages. The typical chessboard distributions of morphological polarity are instantiations of strategy (C). The implication of this typology of matching strategies is that polar distribution of inflectional markers is in no way unexpected, but the most efficient way of referring to feature specifications minimally ambiguously with a minimal formal inventory (i.e., minimal formal inventory, but at the same time minimal ambiguity).

3. Claim

The goal of this paper is to propose a new analysis of polarity effects as a systematic pattern resting upon the idea that the choice of marker for a given feature specification is determined by the choice of marker for a minimally different specification. In the new approach, polarity effects are a result of the interaction of natural class-based rules and two economy principles, Discreteness of Environment and Minimality.

(9) Discreteness of Environment (general version)
Adjacent cells must be discretely marked.

(10) Minimality
If the association of a marker M1 with a matching morpho-syntactic environment [α] violates a principle P, then insert a marker M2 such that M2 meets P and the features of M2 are minimally distinct from the features of M1.

The effect of (9) and (10) is that the choice of marker for a given feature specification is determined by the choice of marker for a minimally different specification: whenever a marker matches a feature specification but is prohibited by Discreteness, a marker with a minimally different specification is chosen to fill the given cell.

Minimality is one of the basic properties of language (alongside e.g. compositionality, double articulation, and cyclicity). Syntactic examples of minimality effects are manifold, e.g. wh-movement in English:

(11) a. (I wonder) who$_1$ t$_1$ invented what$_2$
b. * (I wonder) what$_2$ who$_1$ invented t$_2$
c. *What$_2$ did who$_1$ invent t$_2$?

Data like (11) show that in English multiple wh-questions, extraction of the lower wh-item is accepted by native speakers to a much lesser degree than extraction of the higher wh-item. The same principle is at work in extraction from complex NPs:

(12) a. They heard [DP$_1$ a rumour that [DP$_2$a linguist] dined and dashed at Cafe Kowalski]
b. [DP$_1$What] did they hear t$_1$?
c. * [DP$_2$Who] did they hear [DP$_1$ a rumour that t$_2$ dined and dashed at Cafe Kowalski?]

A third example is object shift in Icelandic, where the higher, but not the lower object can be moved above the negation (Collins and Thrainsson 1996):

(13) a. ´Eg l´ana Mar´ıu$_1$ ekki t$_1$ bækurnar$_2$
I lend Maria.dat not books.acc
b. *´Eg l´ana bækurnar$_2$ ekki Mar´ıu$_1$ t$_2$
I lend books.acc not Maria.dat
'I do not lend the books to Maria'

The underlying principle has been formulated first in Chomsky (1964) as A-over-A Principle, of which a generalised version (F-over-F) is given in (14a). It was reformulated as Superiority Condition (14b)[6].

(14) a. F-over-F Principle
*In a structure $a_{[*F*]}$ ... [$\beta_{[F]}$... [γ_F ...] ...] ..., movement to [*F*] can only affect the category bearing the [F] feature that is closer to [*F*].*
b. Superiority Condition (Chomsky 1973):
In a structure $\alpha_{[*F*]}$ ... [... $\beta_{[F]}$ ... [... γ_F ...] ...] ..., movement to [*F*] can only affect the category bearing the [F] feature that is closer to [*F*].

The most recent formulation of the minimality principle is the combination of these two constraints, known as the Minimal Link Condition:

(15) Minimal Link Condition (Chomsky 2000, 2001):
If β and γ both match a probe α and β asymmetrically c-commands, a syntactic operation γ cannot involve α and γ.

Minimality is also an underlying principle in language processing (Late Closure, Minimal Chain Principle, Minimal Attachment; see e.g. Frazier and Fodor 1978; Bornkessel and Schlesewsky 2006) and phonology (e.g. association lines in autosegmental phonology; Goldsmith 1976). It therefore seems reasonable to assume that the minimality property of language is at work in morphology, too (M¨uller 2007).

4. Formal Implementation

The new analysis works independently of a particular feature representational system. Section 4.1 demonstrates the implementation in an approach resting upon feature geometry; in section 4.2 the analysis is implemented for a binary feature system.

4.1 Implementation for Feature Geometry Representations

The basic assumption of approaches working with feature geometries is that grammatical categories are represented as decomposed into geometrically organized privative features (Harley and Ritter 2002, among many others). The feature geometry for Old French masculine o-stems is the one given in (6). The basic idea of this analysis is that paradigms are

[6] The difference between (14a) and (14b) is that _ dominates in the F-over-F Principle, whereas in the Superiority Condition, _ c-commands .

generated in such a way that the system detects the most specific vocabulary item and the morphosyntactic specification associated with it, and then detects a contiguous specification and the marker associated with it. The system thus proceeds until all cells have been filled. The transition from cell to cell (or specification to specification) is accomplished by detecting that the featural specification of the "new" cell can be reached starting from the feature specification of the "old" cell by a transition from a node β in the feature geometry to an adjacent node β. In paradigms showing polar distributions, this transition from node to node is subject to the Discreteness Principle, which is given in (16) in its feature geometry version.

(16) Discreteness of Environment (feature geometry version)
Adjacent nodes in the geometry must be discretely marked.

This constraint has the effect that whenever a marker M*β* is detected by a transition from a node *α* to a node *β*, then the marker associated withM*β* has to be phonologically distinct from M*α*. If M*α* and M*β* are associated with the same phonological features, then the choice of marker is determined by the Minimality Principle (10). The vocabulary items for Old French masculine o-stems are given in (17).

(17) Vocabulary items for Old French (masc. o-stems)
/-s/ ↔ [case|obj ind|group]
/-ø/ ↔ [ ]

The paradigm develops in such a way that at first the most specific vocabulary item (/-s/ ↔ [case|obj ind|group]) is inserted in the matching context (Specificity Principle).

(18)	MORPHOSYNTACTIC REPRESENTATION	PHON. REPRES.
	ind|group case|obj	-s

Now the most proximate specification is detected (either [ind|group case|subj] or [ind|min cas|obj]; it is of no import which way the system takes in unfolding the paradigm). The transition to either of these specifications is achieved by a transition to an adjacent node in the geometry ([group→ind→min], or [obj→case→subj]). The matching marker for these nodes is /-ø/ (underspecification; cf. Identity Default Rule, Stump 2001). The insertion of the zero marker conforms to the Discreteness Principle: adjacent nodes in the geometry are marked differently.

(19)	MORPHOSYNTACTIC REPRESENTATION	PHON. REPRES.	
	ind|group case|**subj**	-ø	
	|		
	ind|*min* case|obj <– ind|*group* case| **obj**	-ø	-s

The system now detects the remaining specification [ind|min case|subj], which can only be matched with the zero marker. However, the insertion of this marker is prevented by the Discreteness Principle: [ind|min case|subj] and [ind|min case|obj] are adjacent, and [ind|min case|subj] and [ind|group case|subj] are adjacent, but in both cases the adjacent specifications are both associated with the zero marker, thus they are not discretely marked:

(20)	MORPHOSYNTACTIC REPRESENTATION	PHON. REPRES.	
	ind|*min* case|**subj** <– ind|*group* case|subj	*ø	-ø
	|		
	ind|min case|**obj** ind|group case|obj	-ø	-s

In this case the choice of marker is determined by a principle of vocabulary insertion, Minimality:

(21) (=10) Minimality
If the association of a marker M1 with a matching morphological environment [α] violates a principle P, then insert a marker M2 that meets P iff. the feature specification of M2 is minimally distinct from that of M1.

The effect of this principle is that a marker with a minimally different specification is chosen to fill the given cell. The avaliable marker for [ind|min case|subj] with minimally different

specification in Old French is the only alternative marker available in the Old French case system: /-s/. Thus, /-s/ is inserted in the context [ind|min case|subj].

(22)

MORPHOSYNTACTIC REPRESENTATION		PHON. REPRES.	
ind|*min* case|**subj** <–	ind|group case|subj	-s	-ø
ind|min case|**obj**	ind|group case|obj	-ø	-s

4.2. Implementation for Binary Feature Representations

In a system based upon a binary feature representation, the principles of Discreteness and Minimality can be formulated in the form of the following algorithm[7]:

(23) Discreteness of Environment/Minimality (binary feature version):

a. Let x,y = associations of morpho-syntactic and phonological features ("cells").

Let F = morpho-syntactic feature ∈ x,y.

Let P= set of phonological features ∈ x,y.

b. f(sim(x,y)) = n =
{
n=0;
for all F
if val(x)(Fi)= val(y)(Fi)
then n=n+1:
}

c. {
if n=1, then Px 6= Py.
if n6=1, then Px = Py
}

(23b) is a function over the morphosyntactic similarity of a "cell" (i.e., associations of morpho-syntactic and phonological features) x and a "cell" y. In 4-cell paradigms, two primitive features are minimally needed (and thus optimal) to uniquely characterise each of the cells. These features can be cross-classified, as shown in (24):

(24) Feature distribution in 4-cell paradigms:

	–F1	+F1
–F2	–F1 –F2	+F1 –F2
+F2	–F1 +F2	+F1 +F2

The function f(sim(x,y)) operates with all possible pairs of (x,y), that is, all possible pairs of "cells". In a 4-cell paradigm, there are 6 possible combinations:

(25) {–F1 –F2, –F1 +F2};
{–F1 –F2, +F1 –F2};
{–F1 –F2, +F1 +F2};
{+F1 +F2, –F1 +F2};
{+F1 +F2, +F1 –F2};
{+F1 –F2, –F1 +F2}.

For any given pair (x,y), f compares the values of a given feature Fi. If the value Fi of x equals the value Fi of y, then the counter n is set to 1. If the value Fi of x is not equal to the value Fi of y, then n is not raised. Now (23c), the algorithm controlling the phonological realisation, comes into effect: if f returns a number unequalling 1 for a given pair of "cells", then the set of phonological features of x and y must be identical; if f returns 1 for a given pair of "cells", then the phonological features of x and y must not be identical.

Let me exemplify the analysis by means of the Old French masculine o-stem paradigm. The case and number features are decomposed as given in table 3.

[7] Andrew Nevins, p.c.

Table 3

		SG	PL
	NOM	–pl –obj	+pl –obj
Old French		/-s/	/-ø/
Masculine o-stems	OBJ	–pl +obj	+pl +obj
		/-ø/	/-s/

Let x equal [–pl –obl], and y=[+pl –obl]. F1 is thus [pl], and F2 = [obl]. The algorithm starts with n=0. The value of (x)([pl]) equals '–', and val(y)([pl])='+', thus val(x)(F1) 6= val(y)(F1); n is not raised. Now x and y are compared with regard to F2. The value of (x)([obl]) equals '–', and val(y)(obl])='–', thus val(x)(F2) = val(y)(F2). The equation f(sim([+pl – obl],[+pl –obl])) thus returns 0+1 = 1. As a result, (23c) requires Px to unequal Py. If x=[–pl +obl] and y=[+pl –obl], then val(x)(pl) 6= val(y)(pl), and val(x)(obl) 6= val(y)(obl). In this case, f returns 1+1=2. (23c) thus requires Px to equal Py.

5. Partial Polarity Effects

Partial polarity effects are complementary distributions of markers embedded in larger paradigms (Baerman et al. 2005). Examples are Romanian i-stem verbs with -esc/est t-infix (Popovici 2003) and Old Irish masculine ostems (Luhr 2004):

Table 4

		SING	PLURAL
Romanian	1	-esc	-im
I-stem verbs with	2	-est	-iti
-esc/est -infix	3	-este	-esc

Table 5

		SING	PLURAL	DUAL
	NOM	fer	fir	(d´a) ˙fer
Old Irish	ACC	fer	firu	(d´a) ˙fer
Masculine o-stems	GEN	fir	fer	(d´a) ˙fer
	DAT	fiur	feraib	(dib) feraib

The Romanian data can still be described in terms of an epiphenomenon by natural class-based rules alone. The marker /-esc/ then must be the elsewhere marker:

(26) Vocabulary items for Romanian i-stem verbs with -esc/est-infix:
/-este/ ↔ [–1 –2 –pl]
/-iti/ ↔ [+2 +pl]
/-im/ ↔ [+1 +pl]
/-esti/ ↔ [+2 –pl]
/-esc/ ↔ [ ]

However, such an analysis is impossible in the case of Old Irish unless a special device (prioritised operation, _-notation) or a homonymous form /fir/ (or /fer/) is assumed, as the morpho-syntactic environments associated with /fir/ (and, likewise, /fer/) do not have a common value for any given feature.

I would like to put forth a uniform solution for full and partial polarity. So far, however, the new analysis can only satisfactorily account for polarity effects that constitute themselves in strict chessboard distributions of inflectional markers. The solution is a refined notion of Discreteness: recall from section 4.1 that in the new theory, paradigms "unfold" by first inserting the most specific vocabulary item into its morphosyntactic context, and then moving on to a contiguous cell, inserting the matching marker there, and so on; the transition from cell to cell is accomplished by a transition from node to node in the feature geometry. These node transitions are subject to the Discreteness Principle, which requires adjacent nodes in the geometry to be marked discretely. Partial polarity effects can be integrated into the theory by assuming that the Discreteness Principle can be active only on certain node transitions.

Let me illustrate this by means of the Romanian and Old Irish data. In Romanian verb inflection, there are two basic patterns: pattern A shows number syncretism in the 3rd person;

pattern B is constituted by a syncretism of the 1st singular and 3rd plural forms[8]. This is illustrated in tables 6 and 7:

Table 6

Romanian C-final A-stems (PATTERN I)

	SING	PLURAL
1	-ø	-ăm
2	-i	-aţi
3	**-ă**	**-ă**

	SING	PLURAL
1	A	D
2	B	E
3	**C**	**C**

Table 7

Romanian infixal I-stems (PATTERN II)

	SING	PLURAL
1	**-esc**	-im
2	-eşti	-iţi
3	-eşte	**-esc**

	SING	PLURAL
1	**A**	D
2	B	E
3	C	**A**

The feature geometry relevant to these data is given in (27):

(27) Romanian: feature geometry

SING:	PLURAL:	1ST:	2ND:	3RD:
ind	ind	part	part	part
|	|	|	|	
min	group	spk	addr	

The new assumption made is that for verb classes following pattern II there is a Discreteness constraint Dr on the transition sg → pl (i.e., [min → ind → group]):

(28) Discreteness constraint for Romanian (Dr):
[min → ind → group] must be discretely marked / +V +y[9]

The paradigm generation proceeds analogously to Old French (cf. section 4.1), using the vocabulary items given in (29).

(29) Vocabulary items for Romanian I-stems with -esc/est-infix:

/-esc/ ↔ [ind|min part|spk]	(1sg)
/-esti/ ↔ [ind|min part|addr]	(2sg)
/-im/ ↔ [ind|group part|spk]	(1pl)
/-iti/ ↔ [ind|group part|addr]	(2pl)
/-este/ ↔ [ ]	(3)

According to the vocabulary specifications, the marker for 3pl is the elsewhere marker /-este/. However, the insertion of /-este/ is prevented by D_r. Now the Minimality Principle comes into effect. The candidates for filling the cell in question and their feature specifications are shown in table 8 (/-este/, which is not a possible candidate, is marked grey).

Table 8

Romanian Candidates for minimal discreteness

	SG	PL
1	/-esc/ ind part | | min spk	/-im/ ind part | | group spk
2	/-eşti/ ind part | | min addr	/-iţi/ ind part | | group addr
3	/-eşte/ ind part | min	? ind part | group

[8] Pattern A occurs in a-stems, V-final i-stems, ˆı-stems without infix, and some "irregular" verbs in the present indicative; Pattern B occurs in ea-stems, e-stems, C-final i-stems, istems with -esc/est-infix, ˆı-stems with ˇasc/ˇast-infix, and some "irregular" verbs in the present indicative.

[9] y=class feature(s) defining verbs following pattern 2.

The detection of a minimally distinct marker is accomplished by comparing the two nodes A=3 sg and B=3 pl between which Discreteness is violated. The constraint by which the minimally distinct marker /-c/ is detected is given in (30):

(30) Minimal Distinctness

A marker /c/ associated with a node C in the feature geometry is minimally distinct from two nodes A and B in the geometry iff.

a. C is reached from A by a node transition in a single dimension.

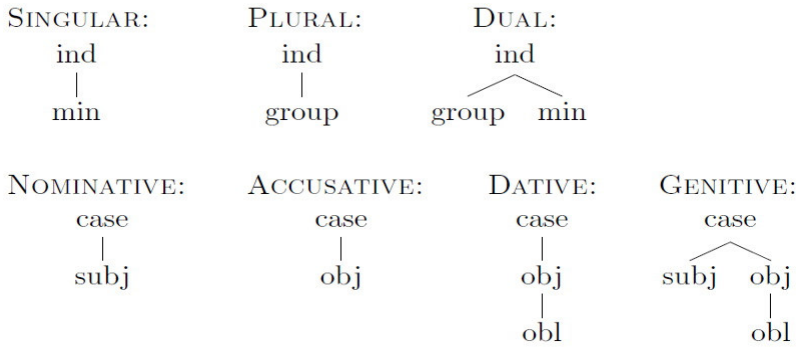
b. C retains at least 1 feature of B.

In the case of Romanian verb inflection, C has to be identical with A=3sg in one dimension (30a). There are two specifications meeting this constraint, 1sg and 2sg. (30b), on the other hand, has no effect here: with B=3pl being underspecified for [part], all nodes that have a [part] dimension meet (30b). (30) thus does not help to decide between the candidates 1sg and 2sg. However, [part|spk] (1sg) is preferred over [part|addr] (2sg) because 'speaker' is the default interpretation for the organising node 'participant' (Harley and Ritter 1999).

Consequently, the marker associated with 1sg is inserted into the 3pl context. A question that comes up at this point is, why is it the nom pl marker that is adjusted, and not the nom sg marker? The answer is that the Discreteness constraint is defined only for the transition from singular to plural ([min→ind→group]), not for [group→ind→min]. Thus, the transition from 3pl to 3sg does not lead to a violation of Dr. The constraint will however not fail to apply, as each single possible node transition is used in the paradigm generation.

The polarity effect in the Old Irish nominal inflection constitutes itself in the i/e umlaut[10]. The analysis is based on the following feature geometry:

(31) Old Irish: feature geometry



For Old Irish masculine o-stems, too, there is a Discreteness constraint on the transition sg → pl ([min → ind → group]). As shown in (33), /fir/ is the marker for genitive singular, and /fer/ is the elsewhere marker.

(32) Discreteness constraint for Old Irish (Di):

[min → ind → group] must be discretely marked / +N +z[11] +m

(33) Vocabulary items for Old Irish fir 'man'[12]:

/fir/ ↔ [case|subj|obj|obl ind|min] (gen sg)
/fiur/ ↔ [case|obj|obl ind|min] (dat sg)
/feraib/ ↔ [case|obj|obl ind] (dat pl, dual)
/firu/ ↔ [case|obj ind|group] (acc pl)
/fer/ ↔ [case ind] (elsewhere)

[10] The umlaut is due to *i > *e / [_a/o in IE.

[11] z= class feature(s) defining o-stems.

[12] Though Old Irish masculine o-stems are arguably best analysed by means of subanalysis (/fir-/, /fer-/, /-aib/, /-u/), I will stick to the pattern instantiated by the whole word forms for the purpose of demonstration.

The only matching marker for both nom sg and nom pl is the elsewhere marker /fer/. The insertion of /fer/ in the context nom pl however is banned by D_i, so that a different marker has to be chosen. As can be seen from table 9, the possible candidates are /firu/ (acc pl), /fir/ (gen sg), /fiur/ (dat sg), and /feraib/ (dat pl+dual). The other feature specifications fail to be legitimate candidates (and are thus marked grey), as they are associated with the marker /fer/, which is disallowed for insertion into the nom pl context.

Table 9 *Old Irish*
Candidates for minimal discreteness

	SG	PL	DUAL
NOM	/fer/ ind case min subj	? ind case group subj	/fer/ ind case min group subj
ACC	/fer/ ind case min obj	/firu/ ind case group obj	/fer/ ind case min group obj
GEN	/fir/ ind case min subj obj obl	/fer/ ind case group subj obj obl	/fer/ ind case min group subj obj obl
DAT	/fiur/ ind case min obj obl	/feraib/ ind case group obj obl	/feraib/ ind case min group obj obl

The marker for the nom pl context is determined by the definition of Minimal Distinctness: The nodes that can be reached by A=nom sg by node transitions in one dimension only are gen sg and dat sg (condition (30a); both are identical with A in their [ind|min] node, and there is no legitimate candidate which is identical with A in its [case|subj] node). Of those two, gen sg wins, as it also has the features [case|subj] in common with nom pl, whereas acc sg has no feature in common with nom pl (condition (30b)).

6 Consequences

The new analysis has a number of advantages. Firstly, only two insertion rules are needed to model full polarity effects, while the morphology is now making use of two principles, the latter of which is an independently well motivated assumption underlying syntactic derivations. Secondly, the choice of marker is predictable (in contrast to impoverishment rules or rules of referral, where the choice is to a high degree arbitrary). Thirdly, the analysis can be implemented in any morphological theory; it is fully compatible with lexical-incremental approaches (Lieber 1992; Wunderlich 1996), lexical-realisational approaches such as Distributed Morphology (Halle and Marantz 1993, 1994), and inferential-realisational approaches such as the Word-and-Paradigm model (Spencer 2001; Stump 2001). Fourthly, in this analysis, polar distributions of inflectional markers come for free as a systematic pattern.

References

Baerman, Matthew (2007): 'Morphological Reversals', Journal of Linguistics 43, 33–61.

Baerman, Matthew, Dunstan Brown and Greville G. Corbett (2005): The Syntax-Morphology Interface: A Study of Syncretism. Cambridge University Press, Cambridge.

B´ejar, Susana and Daniel Currie Hall (1999): Marking Markedness: The Underlying Order of Diagonal Syncretisms. In: R. Daly and A. Riehl, eds, Proceedings of the 1999 Eastern States Conference on Linguistics. Cornell Linguistic Circle Publications, Ithaca, N.Y.

Bobaljik, Jonathan (2002): Syncretism without Paradigms: Remarks on Williams 1981, 1994. In: G. Booij and J. van Marle, eds, Yearbook of Morphology 2001. Kluwer, Dordrecht, pp. 53–85.

Bornkessel, Ina and Matthias Schlesewsky (2006): 'The Extended Argument Dependency Model: A Neurocognitive Approach to Sentence Comprehension Across Languages', Psychological Review 113(4), 787–821.

Chomsky, Noam (1964): Current Issues in Linguistic Theory. Mouton, The Hague.

Chomsky, Noam (1973): Conditions on transformations. In: S. R. Anderson and P. Kiparsky, eds, A Festschrift for Morris Halle. Holt, New York, pp. 232–286.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. In: R. Martin, D. Michaels and J. Uriagereka, eds, Step by Step: Essays 17 on minimalist syntax in honor of Howard Lasnik. MIT Press, Cambridge, Mass., pp. 89–155.

Chomsky, Noam (2001): Derivation by Phase. In: M. Kenstowicz, ed., Ken Hale: A Life in Language. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam and Morris Halle (1968): The Sound Pattern of English. Harper Row, New York.

Collins, Chris and H¨oskuldur Thr´ainsson (1996): 'VP-Internal Structure and Object Shift in Icelandic', Linguistic Inquiry .

Frazier, Lynn and Jane Fodor (1978): 'The Sausage Machine: A new twostage parsing model', Cognition 6, 1–34.

Goldsmith, John (1976): Autosegmental Phonology. PhD thesis, MIT. Halle, Morris and Alec Marantz (1993): Distributed Morphology and the Pieces of Inflection. In: K. Hale and S. J. Keyser, eds, The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger. MIT Press, Cambridge, Mass., pp. 111–176.

Halle, Morris and Alec Marantz (1994): Some Key Features of Distributed Morphology. In: H. H. Andrew Carnie and T. Bures, eds, MITWPL: Papers on Phonology and Morphology. Vol. 21, MITWPL, Cambridge, pp. 275–288.

Harley, Heidi and Elizabeth Ritter (1999): 'Meaning in morphology: Motivating a feature-geometric analysis of person and number'. Manuscript, University of Arizona, University of Calgary.

Harley, Heidi and Elizabeth Ritter (2002): 'Person and Number in Pronouns: A Feature-Geometric Analysis', Language 78(3), 482–526.

Lieber, Rochell (1992): Deconstructing Morphology. University of Chicago Press, Chicago.

Luhr, Rosemarie (2004): Lectures on Old Irish, 2002-2004, Universit¨at Jena.

Muller, Gereon (2007): 'Linguistische Mythen'. Talk delivered at "Studium Universale – Kosmos Sprache", Leipzig, 17 October 2007.

Noyer, Rolf (1997): Features, Positions and Affixes in Autonomous Morphological Structure. Garland, New York.

Noyer, Rolf (1998): Impoverishment theory and morphosyntactic markedness. In: S. G. Lapointe, D. K. Brentari and P.M. Farrell, eds, Morphology and its Relation to Morphology and Syntax. CSLI, Stanford, pp. 264–286. 18

Popovici, Victoria (2003): Lectures and Courses in Romanian, 2000-2003, Universitat Jena.

Rheinfelder, Hans (1976): Altfranz¨osische Grammatik. Formenlehre. Vol. 2, Hueber, Munchen.

Saeed, John (1999): Somali. John Benjamins, Amsterdam.

Spencer, Andrew (2001): 'The Word-and-Paradigm Approach to Morphosyntax', Transactions of the Philological Society 99, 279–313.

Stump, Gregory T. (1993): 'On Rules of Referral', Language 69(3), 449–479.

Stump, Gregory T. (2001): Inflectional Morphology. Cambridge University Press, Cambridge.

Weißer, Philip (2007): 'Gap Morphology and the Theory of Case Borrowing', Linguistische Arbeitsberichte 85.

Wunderlich, Dieter (1996): Minimalist Morphology: The Role of Paradigms. In: G. Booij and J. van Marle, eds, Yearbook of Morphology. Kluwer, Dordrecht, pp. 93–114.

Classification of Chinese compounds

Antonella Ceccagno and Bianca Basciano

(University of Bologna) (University of Verona)[1]

1. Introduction

This article, developed within the Morbo/comp project on compounds[2], aims at contributing to a widening of the debate on the classification of Chinese compounds and their internal structure initiated by Ceccagno and Scalise (2006). To this end, we will briefly introduce the existing classifications put forth by different scholars for compounds and in particular we will focus on the approaches proposed for Chinese compounding.

We will then discuss how to apply to Chinese the classification for compounds argued for by Bisetto and Scalise (2005). After highlighting some shortcomings of this classification with respect to Chinese compounds, we will propose a refinement of it through a new definition of the three macro-types proposed by the two authors.

We will present a new table of Chinese compounds based on an analysis of a corpus of neologisms[3].

2. Classification Schemes for Compounds

Compounds have received much attention in recent years; several proposals for classification have been put forth in the current literature. One of the most prominent problems in compound taxonomy has been the heterogeneousness of the adopted criteria. Most of the classification schemes (Spencer, 1991, Bauer, 2001, Haspelmath, 2002, Booij, 2004 among many others) exhibit a flat structure, since different criteria have been put on the same level: for example, grammatical relations between the constituents are at the same classification level with presence or absence of lexical head (i.e. coordination and subordination are put on the same level with endocentricity and exocentricity). This has been discussed in detail by Bisetto and Scalise (2005; forthcoming).

In order to avoid the described shortcomings, Bisetto and Scalise (2005) propose a new classificatory scheme based on hierarchical arrangement of homogeneous criteria, following the

---

[1] The general outline of the work was discussed jointly by the authors, however Bianca Basciano is responsible for sections 1, 2, 3, 4.1.1, 4.1.2 and Antonella Ceccagno for sections 4.1.2, 4.2, 4.3, 5.

[2] The Morbo/Comp project based in Bologna is led by Sergio Scalise. It concerns the analysis of compounds in some thirty languages, distributed in such a way as to cover the six macro-areas that have been identified by recent typology (cf. Dryer 1992). To present, fifteen languages have been studied, for a total in excess of 70.000 compounds. Cf. http://morbocomp.sslmit.unibo.it/index.php?section=home

[3] The database consists of 1077 neologisms, presented as a separate section in *The Contemporary Chinese Dictionary* (2002). In our analysis for reasons of simplicity we have chosen to focus on dysillabic words, excluding words with more than two syllabs (368). The analysis of the compounds is presented in Ceccagno and Basciano (forthcoming a).

path of earlier classifications introduced by Bloomfield (1933), Bally (1965) and Marchand (1969). Bisetto and Scalise (2005) identify three macro-types in compounding, i.e. subordinate, attributive and coordinate. Each type may be endocentric (with a lexical head) or exocentric (without a lexical head). Besides, Scalise and Guevara (2006) distinguish between categorial and semantic head. The categorial head is "the constituent which shares with – and percolates to – the whole compound all of its formal features: lexical category and subcategorization frame". The whole compound, thus, is expected to have the same distributional properties of its categorial head. The semantic head "is the constituent which shares with – and percolates to – the whole compound all of its lexical/ conceptual information. The whole compound, thus, is expected to be a hyponym of its semantic head". According to the authors, an endocentric compound has at least one categorical head and at least one semantic head (the two must coincide).

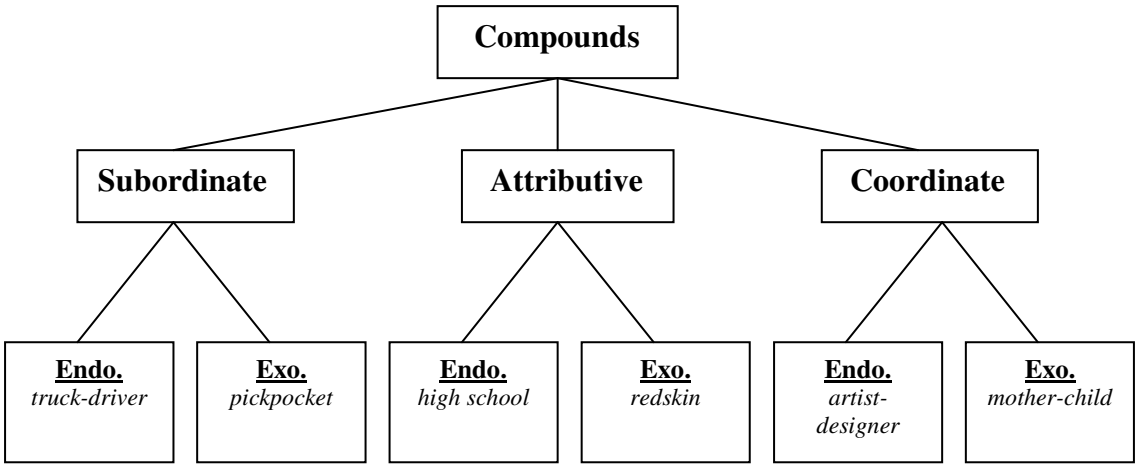
According to Bisetto and Scalise (2005), each macro-type is characterized by a different grammatical relation between the constituents of the compounds. Subordinate compounds (SUB) entail a relation of complementation between the head and the non-head. This is clear in compounds with a deverbal head constituent, such as *truck-driver.* A similar relation can be found in compounds that do not have a deverbal head, for example [N+N] compounds where the constituents are typically linked by what they call an 'of-relation', as in *doorknob* ('knob *of* a door'), or by other clearly subordinating relations, as in *catfood* ('food *for* cats').

Attributive compounds (ATT) are those in which the constituents are linked by a relation of attribution. The prototypical case involves compounds in which the first constituent is an adjective, as in *high school.* Other structural types are found as well, for instance [N+N] attributive compounds, in which the non-head is used as a metaphoric attribute of the head, as in *swordfish* ('fish with a *sword-like* snout') and not as a mere complement of it (*fish *of* a sword, fish *for* a sword).

In coordinate compounds (CRD) the constituents are linked by a coordinating relation. This relation can be a conjunctive natural coordination, as in *artist-designer.* However, in other languages we can find other types of coordination as well, as we will see for Chinese coordinate compounds.

Figure 1 summarizes the classification scheme presented in Scalise and Guevara (2006):

Figure 1 – Classification of Compounds



3. Theoretical Approaches in Analysing Chinese Compounds

Different approaches have been adopted also in the analysis of Chinese compounds (for a detailed description of the approaches and a criticism of them, see Packard, 2000; Ceccagno and Scalise, 2006). One approach describes compounds in terms of the relationship running between the two constituents of a compound. Xia (1946, in Pan, Yip and Han, 1993) identifies four kinds of relationship: 'meaning limiting', 'oppositional', 'modificational' and 'cause-effect'.

Another approach is represented by the modification structure description: compounds can be analysed on the basis of the modification relationship between the constituents. The modification structure can take a juxtapositional form, where neither constituent has a modifying or subordinate relation to the other, or a hierarchical form, with one constituent modified by and therefore structurally dominating the other (this approach has been discussed in Packard, 2000. For a detailed discussion of the many types of semantic relations between the modifying and modified constituent of nominal compound words, see Li and Thompson, 1981).

Compounds can also be analysed through a semantic approach, i.e. by describing how the meaning of the whole word is built up from the meanings of its parts.

One widespread approach is the syntactic description: compounds constituents are regarded as having the same identities as the syntactic constituents that make up a sentence (e.g. subject and object). The first who fully developed syntactic description in analysing Chinese compounds are Chao (1948) and Lu (1964), but over time this approach has expanded and has been adopted by others with slight modifications (Chao, 1968; Tang, 1989, among others)[4]. Following Tang classification (1989, cit. in Yu, 2003), compounds are divided into: subject-predicate; modifier-head; verb-object; verb-complement; coordinate construction (where, it should be noted, the highlighted relation between the components is semantic, not syntactic).

[4] The types of compound are those described by Tang (1989) cited by Yu (2003).

Packard (2000) adopts a form class description. His approach is essentially based on the priority of the lexical category of the compound word with respect to the categories of the constituents. Ceccagno and Scalise (2007) point out that this approach is hard to integrate in a theory of morphology because according to the lexicalist theory it is the head that confers the category to the whole of the compound and not viceversa. Besides, in actual practice, when analysing compounds Packard very often cannot avoid referring to other types of information.

Ceccagno and Scalise (2007) propose that the whole set of category, functional and semantic levels has to be taken into account for an exhaustive analysis of compounds: "an analysis capable of identifying: the lexical category of the constituents, their grammatical relationship and therefore the classification of the whole compound, the semantics of the constituents, the semantics of the compound and the position of the head (if any)". They posit that in the absence of one of these aspects the analysis is incomplete if not misleading.

4. A New Classification of Chinese Compounds

In analysing Chinese compounds we first adopted the classification scheme argued for by Bisetto and Scalise (2005). However, in our view, the proposed distinction between subordinate and attributive compounds does not seem to be clear enough. In fact, following this classification scheme a number of Chinese compounds do not easily fit in one or the other macro-type. For example, [V+N] compounds, as 賣場 *màichǎng* 'sell + large place where people gather for a specific purpose = big marketplace *for* selling commodities', and [N+V] compounds, as 函售 *hánshòu* 'letter + sell = order by mail', show a relation between the constituents that is hardly attributable to one or the other macro-type if we follow the macro-types' description argued for by Bisetto and Scalise (2005).

Therefore a more detailed description of the macro-types' classification is needed. After discussing some problematic examples in our database we shall propose a stricter definition of the macro-types within the proposed classification scheme.

4.1 Classification Issues

4.1.1 $[V+V]_V$ Compounds Across Macro-types

Compounds with the same structure can show different relations between the constituents. This is evident, for instance, in $[V+V]_V$ compounds which can be classified in all three macro-types on the basis of the relation between the two constituent verbs. However, sometimes it is not simple to establish the macro-types in which to include some of these compounds.

The prediction of the syntactic and semantic relation between the two constituents of [V+V] compounds in Chinese represents a very challenging topic of research, also because compound

verbs lack morphological markings (cf. Chan, Chen and Huang, 2000; Chang and Chen, 1999, among others)[5]. For this reason, we have to focus on the interpretation of the whole compound. For example, the compound 競買 *jìngmǎi* 'compete + buy' could be classified as coordinate, 'compete and buy', but the meaning of the compound, 'compete to buy', makes it clear that the relation between constituents is of the subordinate type ('serial verbs'), in which case the constituent on the left acts as the head of the compound. The relation between the two constituents is of the verb-complement subordinate type.

The same is true for 拒載 *jùzài* 'refuse + carry = (of a tax driver) refuse to take a passenger', where the meaning of the whole compound makes clear that the event express by the verb on the right depends on that expressed by the verb on the left, so that the compound is a subordinate verb of the serial verbs type.

An attributive interpretation is possible for $[V+V]_V$ compounds as well. For compounds such as 躍升 *yuèshēng* ('leap + rise') native speakers provide the interpretation of 躍進式升長 *yuè jìn shì shēng zhǎng* 'rise in form of leaps'. This leads us to consider this type of compound as right-headed attributive compounds in which the left constituent acts as a modifier.

In conclusion, on the basis of the corpus of neologisms analysed in Ceccagno and Basciano (forthcoming a), it can be stated that Chinese $[V+V]_V$ compounds can cover the entire spectrum of macro-types: they can be coordinate compounds, in which the two constituents are linked by a coordinating relation, be it of conjunction, as in 推展 *tuīzhǎn* 'recommend + exhibit = recommend and exhibit for sale', or synonymy, as in 操控 *cāokòng* 'control + control = operate and control'; furthermore, Chinese exhibit right-headed attributive compounds, in which the non-head verb constituent acts as a modifier of the head, as in 躍增 *yuèzēng* 'leap+ increase/add = grow by leaps'; finally Chinese shows left-headed subordinate compounds, both of the resultative type[6], as in 入住 *rùzhù* 'come into/enter + live/stop = move into', and of the 'serial verbs' type[7], as in 競賣 *jìngmài* 'compete + sell = compete to sell'.

[5] The same difficult of interpretation is found in syntax too, where a sequence of two verbs without any overt subordinative or coordinative markers is often ambiguous and can generate different interpretations (Li & Thompson, 1981, Paul, 2006 among others).

[6] Resultative constructions have been widely studied in literature. Lin (1990), Li (1990), Cheng (1997), Packard (2000), Li (2005) and Sun (2006) all seem to assume that these constructions exhibit a verb-complement relation, where the resultative verb (on the right) acts as a complement of the verb on the left. As highlighted by Li (2005), Levin and Rappaport (1995) provide evidence that English resultatives consist of a verb taking a complement XP: in a resultative construction the result phrase is in the complement position. This supports the analysis of resultative compounds as left-headed. A different analysis of Chinese resultative constructions has been put forth by Starosta et alii (1998), who consider resultative verbs as derivate words headed by the resultative suffix on the right.

[7] Serial verbs compounds are modelled on the syntactic structure of serial verbs constructions in Chinese, where the event expressed by the verb on the right depends on that expressed by the verb on the left, so that they are considered left-headed compounds . As far as we know, this kind of compound has never been discussed in literature, which has focused instead on resultative compounds (often referred to as resultative serial verbs constructions). What we call serial verb compounds show the same relation

4.1.2 Subordinate and Attributive $[N+V]_V$ Compounds

Chinese $[N+V]_V$ compounds do not easily fit in the macro-types' description provided by Bisetto and Scalise (2005). For this reason, in their analysis of Chinese compounds, Ceccagno and Scalise (2006), sometimes classify these structures as subordinates, e.g. 筆伐 *bǐfá* 'writing brush + attack = condemn or denounce in writing', sometimes as attributives, e.g. 口算 *kǒusuàn* 'mouth + calculate = do a sum orally'. Actually both compounds, apart from sharing the same structure, also share the same relation between constituents - the noun is an adjunct of the verb. Therefore they need one and the same classification.

While the attributive interpretation is unequivocal for compounds with a $[N+V]_V$ structure, such as 升 *biāoshēng* 'whirlwind + rise = (of price, quantities) soar', for other types of compounds, 函售 *hánshòu* 'letter + sell = order by mail', the classification appears to be more problematic. The noun constituent acts as an adjunct[8] which modifies the verb.

We therefore propose to classify $[N+V]_V$ compounds as attributives where the noun constituent acts as a modifier of the head. According to Ceccagno and Basciano (forthcoming a), subordinate verbs are all left-headed (or exocentric), whereas right-headed verbs can only be found in attributive compounds. This could be a further -structural- justification in favour of classifying these compounds as attributives[9].

4.1.3 Subordinate and Attributive $[V+N]_N$ Compounds

In Chinese, $[V+N]_N$ compounds, despite having the same structure and the same output category, can entail different relations between the constituents. For instance, 文胸 *wénxiōng* 'cover up + breast = bra' is an exocentric compound in which the constituents have a verb-argument relation, while 飄塵 *piāochén* 'float + dust = floating dust' is a right-headed compound where the verb constituent modifies the head. We therefore classify the first type as subordinate (exocentric) compounds, while list the second as attributive (endocentric) compounds. In fact, unlike subordinate compounds with $[V+N]_N$ structure, in the case of attributive nouns with $[V+N]_N$ structure the interpretation of the compounds coincides with that of the noun constituent (that acts as the head); in this case the non-head constituent (i.e. the verb) specifies a feature of the head. For instance, 賣場 *màichǎng* is a 'square' where people

---

between the constituents as that shown by syntactic serial verb purpose clause and complement clause constructions (Li & Thompson, 1981).

[8] As for verbal phrases, Chinese exhibits a (left-headed) VO order with direct internal argument, but constructs adjuncts to the left (head on the right). Therefore the $[N+V]_V$ structure, with the noun being an adjunct, is perfectly in line with the syntactic order of Chinese.

[9] As discussed in Ceccagno and Basciano (forthcoming a, forthcoming b), Chinese is a language with three head position in compounding, since it consistently forms right-headed compounds (noun, verbs and adjectives), left-headed compounds (subordinate verbs) and two-headed compounds (coordinate compounds).

sell. Therefore, while in $[V+N]_N$ subordinate compounds the verb constituent confers a thematic role and the compound receives an agentive, locative, instrumental or eventive interpretation, in $[V+N]_N$ attributive compounds the verb acts as a modifier of the head.
This type of attributive compound too, just like those with an adjective modifier, exhibit the same order of the Chinese syntax, where the nominal phrase is left-branching. In particular, in $[V+N]_N$ attributive compounds the constituents follow the syntactic order of the noun phrase where a relative clause precedes the noun head.

The case of $[V+N]_N$ compounds also highlights the importance of analysing all of the aspects of a compound, discussed in 3. When some aspects are neglected, one can reach the same wrong conclusion as Huang (1998) who concluded that: "[…] Chinese is a headless language in its morphology since neither the rightmost nor the leftmost member of a compound uniquely determines the category type of a compound"[10]. Huang reached this conclusion after analysing the output categories of a substantial corpus of Chinese disyllabic compounds. He observed that compounds with the same structures can have different output categories. For example, [V+N] compounds can be nouns as well as verbs, and sometimes adjectives. What Huang failed to consider is the relation between the constituents. In fact, the position of the head also depends on the relation running between the two constituents. As we have seen, in a compound such as 賣場 màichǎng 'to sell + place = big marketplace for selling commodities', with a $[V+N]_N$ structure, the head is the noun constituent and the verb acts as a modifier, i.e. 'the place in which one sells'. On the other hand, in a $[V+N]_V$ compound, such as 投資 tóuzī 'to put + money = to invest', the head is the left constituent, the verb, while the noun acts as its internal argument. Finally, a $[V+N]_N$ compound such as 監事 *jiānshì* 'supervise + matter/responsability = supervisor' has the same structure and the same output category as the compound in the first example, but the relation between the constituents is completely different: it is an exocentric compound where the constituents have a verb-argument relation. The examples highlight that not only input and output categories are important, but also the relation between the constituents contributes to determine the position of the head in a compound.

4.2 Classification of Chinese Compounds

The issues discussed above highlight the need for a better definition of the macro-types. While adopting the classification scheme argued for by Bisetto and Scalise (2005), we propose a new description of macro-types.
Subordinate compounds (SUB) are those in which constituents have an argument-head (or head-argument) relation. A first type shows either a verbal or deverbal head which projects an argument satisfied by the non-head constituent. Moreover, we consider subordinate compounds

[10] For a critical analysis of Huang's positions see Ceccagno and Scalise (2007).

also those compounds that show a verbal head that takes a verb as its complement. Examples are presented below:

(1) 毒販 *dúfàn* $[N+N]_N$ 'drug + vendor = drug trafficker'

(2) 禁毒 *jìndú* $[V+N]_V$ 'prohibit + poison = ban the production, sale and abuse of drugs'

(3) 攀高 *pāngāo* $[V+A]_V$ 'climb + high = climb up; rise

入住 *rùzhù* $[V+V]_V$ 'come into/enter + live/stop = move into'

(4) 拒載 *jùzài* $[V+V]_V$ 'refuse + carry = (of a tax driver) refuse to take a passenger'

毒販 *dúfàn* (1) is a compound with a deverbal head, where the leftmost constituent acts as its argument.

禁毒 *jìndú* (2) is a verbal compound of the verb-object type, where the leftmost constituent is the head of the compound and the noun acts as the internal argument of the verb.

攀高 *pāngāo* and 入住 *rùzhù* (3) are verbal compounds of the resultative type, in which the non-head constituent is in a complement relation with the head constituent, specifying the result produced by the event of the head (the left constituent).

拒載 *jùzài* (4) is a compound of the serial verbs type, in which the event expressed by the verb on the right depends on that expressed by the verb on the left.

As highlighted by Ceccagno and Basciano (forthcoming a), subordinate nominal compounds are right-headed while subordinate verbal compounds are left-headed.

A second type of subordinate compound shows a relational noun as head, where the non-head acts as a semantic argument saturating the noun head, as in:

(5) 警嫂 *jǐngsǎo* $[N+N]_N$ 'police + sister = respectful term for a policeman's wife'

(6) 價差 *jiàchā* $[N+N]_N$ 'price + difference = price difference'.

Attributive compounds (ATT) are those in which the constituents have a modifier-head[11] relation. These can be compounds where:

1) the non-head is an adjective or a noun which expresses a property of the head, as in 黑金 *hēijīn* $[A+N]_N$ 'black/illegal + money = money used for bribery and other illegal activities within official circles' or 天價 *tiānjià* $[N+N]_N$ 'sky + price = prohibitive price';

2) the non-head constituent acts as an adjunct modifying the head, as in 口算 *kǒusuàn* $[N+V]_V$ 'mouth + (to) do a sum = (to) do a sum orally';

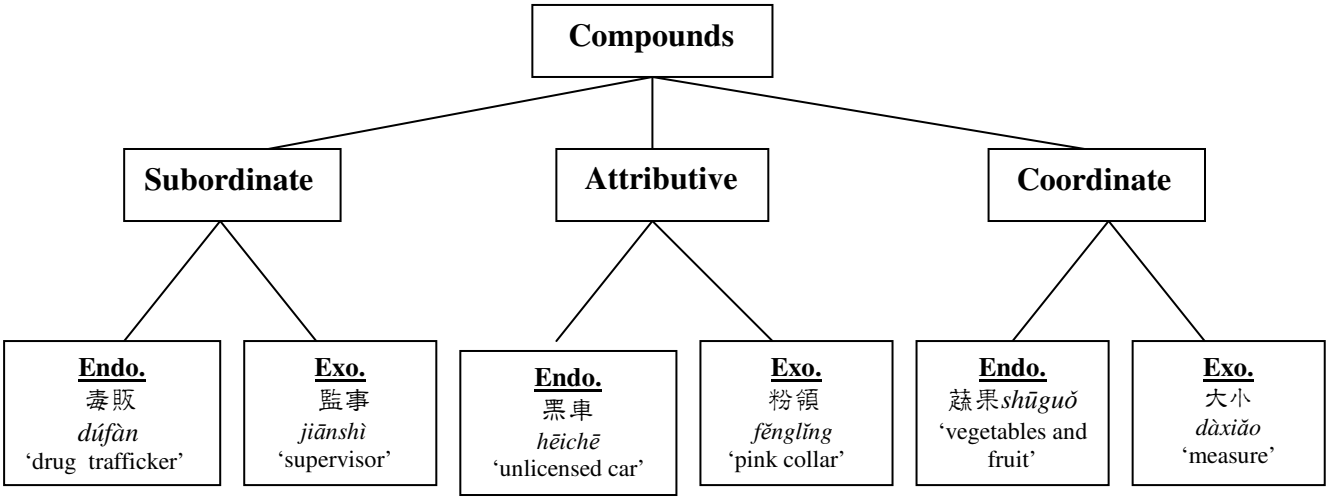
3) a verbal non-head acts as a modifier of the head, as in 賣場 *màichǎng* $[V+N]_N$ 'sell + large place where people gather for a specific purpose = big marketplace for selling commodities'.

[11] A similar distinction between 'argument-head' (or 'argument-predicate') and modifier-head has been proposed for English by Giegerich (2004) and Plag et alii. (2007) among others.

Coordinate compounds (CRD) are those which show a logical coordination between the constituents ('and'), as in 新鋭 *xīnruì* $[A+A]_A$ 'new + sharp = new and sharp', or a semantic relation of sinonymy, as in 勝績 *shèngjì* $[N+N]_N$ 'victory + achievement = win/victory'; antonymy, as in 呼吸 *hūxī* $[V+V]_V$ 'exhale + inhale = breath'; redundancy, as in 松樹 *sōngshù* $[N+N]_N$ 'pine + tree = pine tree'; reduplication, as in *tiāntiān* $[N+N]_{Adv}$ 'day + day = every day'.

Figure 2 shows the classification scheme illustrated with Chinese examples.

Figure 2- Classification of Chinese Compounds



4.3 Structures of Chinese Compounds

The new classification scheme, enables us to present a table of Chinese compounds, which includes all the different types of compound emerged from the analysis in Ceccagno and Basciano (forthcoming a). Table 1 illustrates an example for each available type of compound. The table shows the compound, its transcription in *pinyin*, output category, internal structure, classification, position of the head, glosses and translation[12].

Table 1 – Types of compound

LEGEND: A= adjective, V= verb, N= noun, Adv= adverb

COMPOUND	PINYIN	CLASS	STRUCT	CAT	HEAD	GLOSS
房型	fángxíng	SUB	[N+N]	N	right	house + model = layout of a house
市道	shìdào	SUB	[N+N]	N	exocentric	market + way = market prices

[12] While most of the compounds are the neologisms from the corpus analysed in Ceccagno and Basciano (forthcoming a), compounds in gray are those presented by Ceccagno and Scalise (2006) from a more traditional corpus, not found among the neologisms analysed by Ceccagno and Basciano.

楼花	lóuhuā	SUB	[N+V]	N	exocentric	floor + spend/use = building that is put up for sale before it is completed
監事	jiānshì	SUB	[V+N]	N	exocentric	supervise+ matter/responsibility = supervisor
待崗	dàigǎng	SUB	[V+N]	V	left	wait for + post = wait for a job
割肉	gēròu	SUB	[V+N]	V	exocentric	cut with a knife + meat = sell sth at a price lower than its original price
攀高	pāngāo	SUB	[V+A]	V	left	climb + high = climb up; rise
入住	rùzhù	SUB	[V+V]	V	left	come into/enter + live/stop = move into
拒載	jùzài	SUB	[V+V]	V	left	refuse + carry = (of a tax driver) refuse to take a passenger
膽小	dǎnxiǎo	SUB	[N+A]	A	exocentric	guts/ courage + small = coward
失范	shīfàn	SUB	[V+N]	A	exocentric	lose/deviate from the norm + model = irregular
天價	tiānjià	ATT	[N+N]	N	right	sky + price = proibitive price
色狼	sèláng	ATT	[N+N]	N	exocentric	lust + wolf = sex maniac
婚介	hūnjiè	ATT	[N+V]	N	exocentric	wedding + introduce = matchmaking
飄塵	piāochén	ATT	[V+N]	N	right	float + dust = floating dust
蹦床	bèngchuáng	ATT	[V+N]	N	exocentric	jump + bed = trampoline
速遞	sùdì	ATT	[A+V]	N	exocentric	fast + hand over/pass = express delivery
黑車	hēichē	ATT	[A+N]	N	right	black/illegal + vehicle = unlicensed car
黃毒	huángdú	ATT	[A+N]	N	exocentric	yellow/pornographic + poison = pornographic books
互動	hùdòng	ATT	[Adv+V]	N	exocentric	mutually + move = interaction
升	biāoshēng	ATT	[N+V]	V	right	whirlwind + rise = (of price, quantities) soar
品讀	pǐndú	ATT	[V+V]	V	right	decide with discrimination + read = read carefully; ponder on
航拍	hángpái	ATT	[V+V]	N	exocentric	navigate (by water or air) + take a photograph = aerial photograph

完勝	wánshèng	ATT	[A+V]	V	right	whole + win victory = win a complete victory
突審	tūshěn	ATT	[Adv+V]	V	right	unexpectedly + interrogate = interrogate sb. by surprise
利淡	lìdàn	ATT	[A+A]	N	exocentric	favourable + weak= unfavourable information for the market which may lead to a fall in prices
雪白	xuěbái	ATT	[N+A]	A	right	snow + white = white like the snow
高發	gāofā	ATT	[A+V]	A	exocentric	high/above the average + deliver = frequent
花心	huāxīn	ATT	[A+N]	A	exocentric	attractive but unreal or insincere + heart = unfaithful
統合	tǒnghé	ATT	[Adv+A]	A	right	all/totally + whole = uniform/overall
頻密	pínmì	ATT	[Adv+A]	A	exocentric	frequently + close/dense = frequent
梯次	tīcì	ATT	[N+N]	Adv	exocentric	steps + order = by echelon or by group/ in order of age, size, etc.
蔬果	shūguǒ	CRD	[N+N]	N	two-headed	vegetables + fruit = vegetables and fruits
東西	dōngxi	CRD	[N+N]	N	exocentric	east + west = thing
峯位	fēngwèi	CRD	[N+N]	N	right	peak + place = peak
警示	jǐngshì	CRD	[V+V]	N	exocentric	warn + show = warning
高矮	gāoǎi	CRD	[A+A]	N	exocentric	high + low = height
	yánfā	CRD	[V+V]	V	two-headed	research/study + develop = research and develop
疏離	shūlí	CRD	[V+V]	V	exocentric	scatter/not familiar + be away from= be estranged and keep apart
亮麗	liànglì	CRD	[A+A]	A	two-headed	bright + beautiful = brilliant/bright and beautiful

5. Conclusions

In this article we have adopted the classification scheme put forth by Bisetto and Scalise (2005) - based on hierarchical arrangement of homogeneous criteria- where three macro-types in compounding (subordinate, attributive and coordinate) are identified.

While highlighting some issues of Chinese compounds, shortcomings of the adopted classification scheme have emerged. Therefore we have proposed a new scheme with stricter definition of the macro-types.

Finally, we have presented a new table, which presents all the different types of Chinese compounds emerged from the corpus of neologisms analysed by Ceccagno and Basciano (forthcoming a); in this table Chinese compounds have been classified according to the NEW classification scheme proposed in this article.

REFERENCES

Bally, C. (1944). *Linguistique générale et linguistique française*, 2. ed. Berne: Francke.

Bauer, L. (2001). *Compounding*. in M. Haspelmath *et al*. (eds.), *Language Typology and Language Universals*. Vol. I, Berlin: Walter de Gruyter, 695-707.

Bisetto, A., and Scalise, S. (2005). Classification of Compounds. *Lingue e Linguaggio*, 2, 319-332.

--- (forthcoming). Classification of Compounds. In R. Lieber and P. Stekauer (eds.), *The Handbook of Compounding.* Oxford University Press.

Bloomfield, L. (1933). *Language*. New York, Holt.

Booij, G. (2004). 'Compounding and derivation: evidence for Construction Morphology', in W. U. Dressler *et al*. (eds.) *Morphology and Its Demarcation*. Selected papers from the 11th Morphology Meeting, Vienna, February 2004, Amsterdam/Philadelphia: John Benjamins, 109-132.

Ceccagno, A., and Basciano, B. (forthcoming a). Compound Headedness in Chinese: An analysis of Neologisms. *Morphology.*

--- (forthcoming b). Chinese compounding. In R. Lieber, P. Stekauer (eds.). *The Handbook of Compounding.* Oxford University Press.

Ceccagno, A., and Scalise, S. (2006). Classification, Structure and Headedness of Chinese Compounds*.* *Lingue e Linguaggio*, 2, 233-260.

--- (2007), 'Composti del cinese: analisi delle strutture e identificazione della testa', in A. Palermo (ed.)*, La Cina e l'altro*. Napoli: Università degli studi di Napoli 'L'orientale': 503-543.

Chang, L.-L., and Chen, K.-J. (1999). 動詞詞構與語法功能互動初探 *Dòngcí cí gòu yǔ yǔfǎ gōngnéng hùdòng chū tàn*. *Proceedings of the 12th Research on Computational Linguistics*. Tsinchu, Taiwan: 67-85.

Chang, L.-L., Chen, K.-J., and Huang, C.-R. (2000). A Lexical-Semantic Analysis of Mandarin Chinese Verbs: Representation and Methodology. *Computational Linguistics and Chinese Language Processing*. Vol. 5, No. 1, February 2000: pp. 1-18.

Chao, Y. R. (1948). *Mandarin Primer: an Intensive Course in Spoken Chinese*. Cambridge, Mass: Harvard University Press.

--- (1968). *A Grammar of Spoken Chinese.* Berkeley and Los Angeles: University of California Press.

Cheng, L. L. (1997). Resultative Compounds and Lexical Relational Structures. *Chinese Languages and Linguistics III: Morphology and Lexicon,* 167-197.

Giegerich, H. (2004). Compound or phrase? English noun-plus-noun constructions and the stress. criterion. *English Language and Linguistics* 8, 1-24.

Haspelmath, M. (2002). *Understanding Morphology*, London, Arnold.

Huang, S. (1998). Chinese as a Headless Language. In J. Packard (Ed.), *Compounding Morphology*: *New Approaches to Chinese Word Formation* (pp. 261-283), Berlin and New York: Mouton de Gruyter.

Levin, B., and Rappaport Hovav, M. (1995). Unaccusativity: At the Syntax-Lexical Semantics Interface. *Linguistic Inquiry Monograph,* Vol. 26. Cambrige, Massachussets: MIT Press.

Li, C. N., and Thompson, S. A. (1981). *A Grammar of Spoken Chinese: A Functional Reference Grammar.* Berkeley: University of California Press.

Li, Y. (1993). Structural Head and Aspectuality. *Language* 69: 480-504.

--- (2005). *A Theory of the Morphology-Syntax Interface*. Cambridge (MA): MIT Press.

Lin, F. (1990). *The Verb-Complement (V-R) Compounds in Mandarin*. M.A. Thesis, National Tsing Hua University.

Lu, Z. (1964). *汉语的构词法 Hanyu de goucifa* (Chinese Morphology). Beijing: Kexue Chubanshe

Marchand, H. (1969). *The Categories and Types of Present Day English Word Formation*.

München: Beck'sche Verlagsbuchhandlung.

Morbo/Comp, http://morbocomp.sslmit.unibo.it/index.php?section=home

Packard, J. (2000). *The Morphology of Chinese*. Cambridge: Cambridge University Press.

Pan, W., Ye. B., and Han, Y. (2004). (*Hanyu de goucifa yanjiu*- Research on Word Formation in Chinese). Shanghai: Huadong Shifan Daxue Chubanshe.

Paul, W. (2005) The "serial verb construction" in Chinese: A Gordian knot. In Oyharçabal B. and Paul W (Eds.), *Proceedings of the workshop La notion de « construction verbale en série » est-elle opératoire ?* December 9, 2004, Ehess, Paris.

Plag, I., Kunter, G., and Lappe, S. (2007). Testing Hypotheses About Compound Stress Assignment in English: A Corpus-based Investigation. *Corpus Linguistics and Linguistic Theory* 3.2, 199-233.

Scalise, S., and Guevara, E. (2006). Exocentric Compounding in a Typological Framework. *Lingue e Linguaggio*, 2, 185-206.

Spencer, A. (1991). *Morphological Theory*. Oxford: Blackwell.

Starosta, S., Kuiper, K., Ng, S., and Wu, Z. (1998). On Defining the Chinese Compound Word: Headedness in Chinese Compounding and Chinese VR Compounds. In J. Packard (Ed.), *New Approaches to Chinese Word Formation* (346-369). Berlin and New York: Mouton de Gruyter.

Sun, C. (2006), *Chinese: A Linguistic Introduction*, Cambridge University Press.

The Contemporary Chinese Dictionary [Chinese-English Edition], 現代漢語詞典[漢英双語] (2000). Beijing: Foreign Language Teaching and Research Press.

Xia, M. (1946). *Methods of composing two-character words* [Cited in Pan, Yip and Han (1993: 37), no further citation given].

Yu, M. (2003). 'Chinese Morphology: An Exploratory Study of Second Language Learners' Acquisition of Compounds, *Concentric: Studies in English Literature and Linguistics* 29.2: 1-35

Tang, T. (1989). *Studies on Chinese Morphology and Syntax: 2.* Taipei: Student Book Co.

On the Semantics of Denominal Adjectives

Bernard Fradin
Laboratoire de linguistique formelle
CNRS & University of Paris 7 Denis Diderot
bernard.fradin@linguist.jussieu.fr

1. Introduction

Denominal adjectives are complex adjectives morphologically constructed on a nominal basis, the base noun (abbreviated here as BseN). They can be formed by many different derivational processes as tableau1 attests for French.

Tableau 1. Varieties of denominal adjectives

SFX	BseN	Gloss	A	Gloss
-AIN	Afrique	'Africa'	*africain*	'African'
-E	*iode*	'iodine'	*iodé*	'iodized'
-EL	*an*	'year'	*annuel*	'annual'
-ESQUE	*Dante*	'Dante'	*dantesque*	'Dantesque'
-EUX	*lait*	'milk'	*laiteux*	'milky'
-IEN	*caméléon*	'chameleon'	*caméléonien*	'chameleonic'
-IN	*opale*	'opal'	*opalin*	'opalin'
-IQUE	*basalte*	'basalt'	*basaltique*	'basaltic'
-U	*branche*	'branch'	*branchu*	'branched'

If we classify adjectives according to the classical distributional criteria given in tableau 2, we see that denominal adjectives range in several distinct subclasses, as shown in tableau 3.

Tableau 2. Distributional criteria for A classification

Predicative position	NP COPULA A	Jane (is + became) sad
Prenominal position	A N	*big tree*
Postnominal position	N A	*stars visible*
Gradable	DEG A	*very angry / completely full*

Tableau 3. Adjectival subclasses

	(Aa)	(Ab)	(Ba)	(Bb)	(Ca)	(Cb)
1. Pred.	–	–	+	+	+	+
2. A N	–	–	–	–	–	+
3. N A	+	+	+	+	+	+
4. Grad.	–	–	–	–	+	+
5. Denom.	–	+	–	+	+	+

Examples : (Aa) *cadet* ‘cadet’, (Ab) *présidentiel* ‘presidential’, (Ba) *borgne* ‘one-eyed’, (Bb) *mensuel* ‘monthly’, (Ca) *pansu* ‘paunchy’, (Cb) *courageux* ‘courageous’, *léger* ‘light’.

In tableau 3, denominal adjectives have value ‘+’ for property 5 ‘denominal’. All denominal adjectives occur in the postnominal attributive structure. The denominal adjectives I am interested in here are true relational adjectives, namely those in (Ab), which cannot occur in a predicative position e.g. **La voiture garée là-bas est présidentielle* ‘the car parked over there is presidential’. This property sets them apart from other subgroups of denominal adjectives, inasmuch as the latter can appear after the copula (cf. (Bb), (Ca), (Cb)). Moreover, denominal adjectives of type (Cb), e.g. *osseux* ‘bony’, may shift from one group to another in function of the semantic relationship existing between the noun which heads the NP (the head noun, HdN) and the base noun. For instance, while *osseux* behaves like a relational adjective in (1), it behaves like a plain predicative A, e.g. *léger* ‘light’, in (2).

(1) a. *tuberculose osseuse* ‘bone tuberculosis’
b. **tuberculose très osseuse* ‘very bone tuberculosis’
c. *?*sa tuberculose est osseuse* ‘his tubercolosis is (a) bone tuberculosis’

(2) a. *visage osseux* ‘bony face’
b. *visage très osseux* ‘very bony face’
c. *son visage est osseux* ‘his face is bony’

The reason why I limit myself to adjectives (Ab) of tableau 3 is that their properties place them at one end of the scale on which denominal adjectives can be ranked, the other end being occupied by adjectives like *courageux*, which do not differ from ordinary descriptive adjectives. Adjectives such as *mensuel* ‘monthly’ and *pansu* ‘paunchy’, on the other hand, occupy positions in-between since they also share properties of ordinary adjectives. To that extent, relational adjectives (Ab) appear to be more typical denominal adjectives than the other ones. In my opinion, it is easier to describe typical specimens of a category than less typical ones.

The scope of this article is to account for the interpretation of denominal adjectives in a postnominal position. In section 2, after reviewing general properties of denominal adjectives, I discuss some existing analyses and examine possible mechanisms in charge of their

interpretation. Section 3 proposes such a mechanism for two standard cases, one where the denominal adjective exhibits an eventive reading and the other where we have a spatial reading. The last sections deal with more complex cases. Section 4 focuses on prefixed denominal adjectives such as *pré-électoral* 'pre-election' and section 5 extends the analysis to morphologically underived adjectives such as *intergalactique* 'intergalactic'.

2. General properties of denominal adjectives

2.1. Morphological structure

I assume that denominal adjectives have the structure examplified in figure 1, couched here in a typed-feature formalism (cf. Sag & Wasow 1999). *A-den* (denominal adjective) notes the lexical type of the adjective. Feature MORPH-ST (Morphological structure), whose value is a sign, gives the structure of the N which is the base of the adjective. Schematic though it is, the structure of fig. 1 notes two important properties of denominal adjectives. First, the fact that their phonology (PHON) is equivalent to the concatenation of the phonology of the base with a suffix (chosen among those mentioned in tableau 1). Second, the fact that the semantics of the derived adjective is the same as the semantics of its base noun. This is expressed through the shared value [2] of the feature REL (RELATION), itself a value of the feature CONT (CONTENT). Figure 2 gives the schematic representation of the denominal adjective PRÉSIDENTIEL.[1]

$$\left[\begin{array}{l} A-den \\ \text{PHON } [1] \oplus \text{sfx} \\ \text{SYN-SEM} \left[\begin{array}{l} \text{HEAD } [\text{CAT a}] \\ \text{CONT } [\text{REL } [2]] \end{array}\right] \\ \text{MORPH-ST} \left[\begin{array}{l} Noun \\ \text{PHON } [1] \\ \text{SYN-SEM} \left[\begin{array}{l} \text{HEAD } [\text{CAT n}] \\ \text{CONT } [\text{REL } [2]] \end{array}\right] \end{array}\right] \end{array}\right]$$

Figure 1. Structure of denominal adjectives

[1] On the apparition of the semi-vowel /j/, see Thornton (1999).

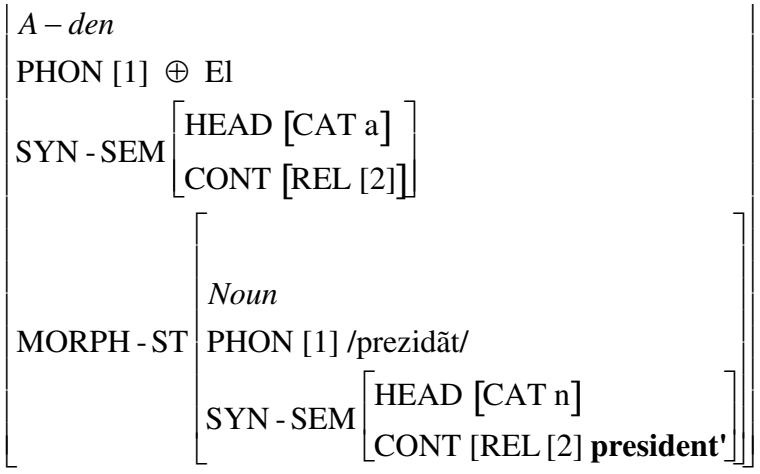


Figure 2. Structure of adjective *présidentiel*

2.2. The semantics of relational adjectives

As the above-mentioned representations state, relational adjectives are semantically transparent. I contend that their only semantic import is to indicate that their base noun is a potential argument of a semantic relation R. Two views exist as regards relation R : a purely contextual one and a mixed one.

The contextual approach claims that R's value is always supplied by the context. It has been argued for by Mezhevich (2004) and endorsed by several other researchers (Mcnally & Boleda 2004). On this view, expression (3a) receives semantic representation (3b) and the value of R is contextually determined.

(3) a. *presidential adviser*
b. (λx. **adviser'**(x) ∧ **R**(x, **president'**))

The mixed approach says that either the semantic representation of either HdN or BseN can provide us with R. Otherwise, or in addition, the relation is supplied by the context (the default option). For instance, if we assume that (4b) is the representation associated with (4a), any of the predicates mentioned in (4c) can be used to instanciate R. Predicates **live-in'** and **build'** come from the semantic representation associated to *palace* (cf. below), while **possess'** would be a default relation triggered by the animacy of the BseN's referent.

(4) a. *palais présidentiel* 'presidential palace'
b. *T*(*palais présidentiel*) = (λx. **R**(x,y) ∧ **palace'**(x) ∧ **president'**(y))
c. R = {**live-in'**, **build'**, **possess'**…}

What are the arguments supporting one approach rather the other? The next subsection discusses this issue and is devoted to the survey of Mezhevich's position.

2.3. Mezhevich's point of view

Mezhevich's argumentation is articulated in three points. On the basis of data like (5), she first argues that the values of R in constructs with a denominal adjective (DA constructs)[2] seem arbitrary and that their range is potentially unlimited.

(5) a. *moloč-n-yj* *magazin*
milk-AZR-M.SG store:M:SG

[2] Denominal adjective constructs (DAC) are incomplete NPs with a DA (NPDA) e.g. *palais présidentiel*. They are incomplete in French because they lack a determinant.

‘store that sells dairy products’

b. *moloč-n-aja* *ferma*
milk-AZR-F.SG farm:F:SG
‘farm that products dairy products’

c. *moloč-n-yj* *stakan*
milk-AZR-M.SG glass:M:SG
‘glass for milk’

d. *moloč-n-yj* *koktel*
milk-AZR-F.SG cocktail:M:SG
‘milk cocktail’

We agree with Mezhevich that the range of readings allowed by denominal adjective is very wide, but this does not mean that the value of R is arbitrary. If it were such, any of the DA construct above could show any of the attested reading, which is not true. For instance, normally, (5c) cannot mean ‘glass made of milk’, or (5b) ‘store that products dairy products’ because of the meaning attached to *stakan* and *magazin*. Secondly, she states that R’s value crucially depends on the meaning of HdN, BseN and the knowledge of the world. But she adds, third point, that the arbitrariness of the relation makes it more dependent on the contextual knowledge. Example (6) is a case in point, since it clearly exhibits a purpose relation whereas, according to Mezhevich, neither *zamok* ‘lock’ nor *skvažina* ‘hole’ expresses a relation of purpose.

(6) *zamoč-n-aja* *skvažina*
lock-AZR-F.SG hole:F:SG
‘keyhole’

Point 2 is true, but point 3 lacks compelling support. *Skvažina* seems to denote a hole dug on purpose as examples (7) show, whereas other Russian translation of *hole* e.g. *dyra* ‘hole’, *jama* ‘pit, hole’, *ščel’* ‘chink, crack, slit’, *otverstie* ‘slot, opening’ are not so overtly telic. The first three generally denote holes the existence of which results from natural causes.

(7) a. *nefta-n-aja* *skvažina*
oil-AZR-F.SG hole:F:SG
‘oil well’

b. *bur-ov-aja* *skvažina*
drill-AZR-F.SG hole:F:SG
‘drill well’

The same reasoning extends to other examples of (5), since all HdNs thereof denote a functional-artefact (store, glass, farm).

Jensen and Vikner’s experiment on genitive interpretation suggests another argument in favor of the idea that, in most cases, relation R can be recovered from the interaction of HdN and BseN’s meanings and that making use of the context to find an adequate pragmatic interpretation is a last resort strategy, not a basic strategy. In their paper (Jensen & Vikner 2002), Jensen and Vikner report an experiment they performed on the interpretation of genitive phrases such as **the N of N**, **the N’s N**, etc. In a corpus of fictional and non fictional texts in English, they found 2,333 genitive forms, of which only 9 had a pragmatic interpretation. As an example of pragmatic interpretation, they give the phrase “the deaf mute’s pen” which had to be interpreted as ‘the pen he had bought from the deaf mute’ in the book’s context. The very low rate of contextual interpretation is at odds with the view that it is the basic interpretive strategy.

As has often been stressed (Partee & Borschev 2003, 2001), the problems of interpretation raised by NPDAs are akin to those raised by genitives, which makes Jensens and Vikner's finding relevant for the present discussion. It remains clear however that a conclusive argument about the interpretation of denominal adjective hinges on checking their interpretation on corpora in French through experiments similar to the one carried out by Jensen and Vikner. I leave this task for future research.

The third argument that can be put to the fore against a purely contextual approach is tied to the contrast between (1) *visage très osseux* and (2) **tuberculose très osseuse*. The point is that this constrast in grammaticality is determined only on the basis of information supplied by the HdN and the BseN (Fradin 2007, 2008 (to appear)). This downplays the role of context in the story, since we can do without it in this latter case.

To sum up, the mixed approach, according to which relation R is established on the basis of the meaning of both the HdN and the BseN on the one hand, and the context on the other hand whenever needed, seems heuristically more promising and will be adopted here.

2.4. The semantic mechanism for DAs interpretation

Since denominal adjectives limit themselves to transmitting the semantic content of their BseN at an upper level, their interpretation takes place at the level of the NP in which they occur, or, more appropriately, at the level of the denominal adjective construct, which corresponds to the lexematic (underlined) constituents in (8) (alternative analyses concerning the realisation of *la* in *l'élection* can be proposed, but they have no bearing on the point under discussion).

(8) $[_{NP}\ [_{DET}$ la$]\ [_{NP}\ [_{N}$ <u>élection</u>$][_{A\text{-}DEN}$ <u>présidentielle</u>$]]]$

The semantic interpretation rule is triggered by the feature A-DEN (denominal adjective) and varies in function of the information associated both to the HdN and the relational A. Relation R is usually instanciated by a predicate supplied either by the HdN or the BseN, the choice between them depending on the semantic sort of these units.

The DA's semantic mechanism shows two important properties. The first one is rule sensibility. I assume that Ns can be classified in function of their semantic properties and that the semantic nature of the N is noted by a feature. The features indicate, for instance, whether the HdN denotes an event, an artefact, an individual, etc., or whether the BseN denotes an object, a place, an agent, etc. This device allows us to adjust the instruction associated with the semantic rule, which changes in function of the feature combination involved. The second property is rule peeking: the semantic rules use information located inside the HdN or the BseN's semantic representation. This idea is not new and has been put in application since Pustejovsky's qualia at least.

I will assume that the semantic information involved in semantic rules comes from two different sources: surface level (or external meaning) and lexical level (or internal meaning cf. Pustejovsky's qualia). The external meaning is the semantic translation of the syntactic units that form sentences (cf. (9)). The internal meaning corresponds to lexically encoded information and usually lacks any syntactic correlate in the sentence (cf. (10)). Note however that standard formal semantics approaches, as Dowty (1979) for instance, do not allow using internal meaning.[3] What makes such a move licit here, I would suggest, is the presence of the denominal adjective in the construction.

(9) a. T(HOUSE) = (λx. **house'**(x)) = **house'**
b. T(FATHER) = (λx. λy. **father'**(x,y)) = **father'**
c. T(ELECTION$_{EXT}$) = (λx. **election'**(x)) = **election'**

(10) a. T(ELECTION$_{INT}$) = (λy. λx. λe. **elect'**(e,y,x))
b. T(MAGAZIN$_{INT}$) = (λz. λy. λx. λe. **sell'**(e,y,x) ∋ LOC(**in'**(z),e)))

[3] Many thanks to Patrick Caudal, who reminded me of this point.

The values taken by relation *R* can be subsumed under three types: eventive, spatial, and equative. As I said before, the choice between one type or the other hinges on the semantic content associated to the HdN or BseN. Tableau 4 gives a synopsis of the main cases we can come across. Only cases in bold will be dealt with in the present study. To give just a hint about the equative type, I would say that it corresponds to readings such as 'fishing that is (an instance of) industry' or 'wrought iron that is (an instance of) art' that can be associated to *pêche industrielle* and *ferronnerie artistique* respectively.

Tableau 4. Synopsis of the semantic relations arising in NPDA

Source	Eventive	Spatial	Equative
HdN	***élection populaire*** 'popular election', *palais présidentiel* 'pre-sidential palace', ***carte routière*** 'road map'	*zone pavillonnaire* 'private housing aera' *centre commercial* 'commercial centre'	*pêche industrielle* 'industrial fishing', *ferronnerie artistique* 'craftsmanship in wrought iron'
BseN		***carte murale*** 'wall map', *élection professionnelle*, 'trade election', *pêche côtière* 'inshore fishing'	

3. Analysis of standard cases

3.1. The eventive interpretation

Let us have a look at examples (11). Assuming that the semantic relationship 'X elect Y (N)' can be recovered from the N ÉLECTION 'election', we see that (11a) and (11b) do not share the same interpretation. While (11a) involves semantic relationship (12a), (11b) involves (12b). This latter interpretation is illustrated in (13).

(11) a. *L'élection présidentielle est terminée.*
'The presidential election is over'
b. *Une élection populaire est difficile à organiser.*
'A popular election is difficult to organize'

(12) a. 'X élire Y président' 'X elect Y president'
b. '(le) peuple élire Y' 'people elect Y'

(13) ***Election populaire*** *des autorités villageoises au Tibet*. (fr.cctv.com/français /special/)
'Popular election of authorities ruling villages in Tibet'

The information associated with the deverbal N ÉLECTION is given under (14) and the translation of PRESIDENTIEL and POPULAIRE is figured in (15). The sorts used here (cf. (14a)) are semantic types needed to account for phenomena such a coercion (Godard & Jayez 1993). They are organised as a hierarchy allowing multiple inheritances. For the purpose of this article, I will adopt Godard and Jayez's three top level objects event (ev), object (o) and properties (p), to which I add the two modes extensive vs. non extensive, where *extensive* means 'having a spatio-temporal dimension' (Flaux & Van De Velde 2000). T_{EXT}(N)—the external meaning—corresponds to the "argument structure" of (Jensen & Vikner 2002).

(14) a. SORT(ELECTION) = ev $\wedge$ extensive
b. T_{EXT}(ELECTION) = (λx. **election**'(x)) = **election**'
T_{INT}(ELECTION) =
ci (λy. λe. $\exists$x. **elect**'(e,x,y)) patient reading (cf. (11a))
cii. (λy. λe. $\exists$x. **elect**'(e,y,x)) agent reading (cf. (11b))

(15) a. T(PRESIDENTIEL) = **president'**
b. T(POPULAIRE) = **people'**

The rule which interprets the structure at the level of the (incomplete) NP is (16), where N_{EV} corresponds to the internal semantic representation of the eventive noun. As clause (16b) makes it clear, the rule also stipulates to what semantic translation $\mathbf{N}_{EV}$ corresponds.

(16) NP's interpretation rule 1
a. $T(N_{EV}\ A_{DEN}) = (\lambda\mathbf{N}_{EV}.\ \lambda\mathbf{A}.\ \lambda e.\ \lambda z.\ \mathbf{N}_{EV}(e,z,\ldots)\ \text{ə}\ \mathbf{A}(z))$
b. $\mathbf{N}_{EV} = T_{INT}(N_{EV})$

Insofar as ÉLECTION has two internal semantic representations, the rule yieds two interpretations. The first one is given in (17) and corresponds to the patient reading ('an event *e* such that an *y* has the property that there is an *x*, who is president, such the *y* elects *x*'); the second one is (18) and corresponds to the agent reading ('an event *e* such an *y* has the property that there is an *x* that elects *y*' and *x* is people).

(17) T(ELECTION PRESIDENTIELLE)
$= (\lambda\mathbf{N}.\ \lambda\mathbf{A}.\ \lambda y.\ \lambda e.\ \exists x.\ \mathbf{N}(e,y,\ldots)\ \text{ə}\ \mathbf{A}(x))(\mathbf{N}_{INT\text{-}A})(\mathbf{A})$
$= (\lambda y.\ \lambda e.\ \exists x.$ **elect**'$(e,x,y)\ \text{ə}$ **président**'$(x))$ (cf. (11a))

(18) T(ELECTION POPULAIRE)
$= (\lambda\mathbf{N}.\ \lambda\mathbf{A}.\ \lambda y.\ \lambda e.\ \mathbf{N}(e,\ldots,x)\ \text{ə}\ \mathbf{A}(x))(\mathbf{N}_{INT\text{-}B})(\mathbf{A})$
$= (\lambda y.\ \lambda e.\ \exists x.$ **elect**'$(e,y,x)\ \text{ə}$ **people'**$(x))$ (cf. (11b))

Cautious readers might have noticed however that interpretation (17) is flawed, since it says that an *x* who is a president has been elected, while (11a) means instead that an *x* has been elected and the result is that he became a president, i.e. something more akin to '(λx. λe. $\exists$z. **elect'**(e,z,x) ə result'(e, **président'**(z)))'. Since my point here is to show how to deal with the difference between the agent vs. patient reading more than to give an exact account of the meaning of *élection présidentielle*, I will leave this issue aside.[4]

[4] Actually, the content of the internal semantic representation (14ci) has to be changed and so has to be the rule interpreting the NP. In the reading in question, ÉLIRE 'elect' creates a juridical state (on a par with JUGER 'judge', for example) and the rule must be devised in such a way that it could cope with all verbs

As for the DA construct *carte routière* 'road map', the link between the HdN and the BseN is also provided by the HdN. The semantic translation of ROUTIER, given in (19), corresponds to the external meaning of the lexeme, while that of CARTE 'cart' in (20) involves two internal meanings in addition to the external meaning. Both of them stem from the fact that this lexeme denotes a functional-artefact. They correspond to the Origin and Telic qualia in Pustejovsky (1995). The Origin quale says that a map is an object which emerges through printing and the Telic quale tells us that it is a semiotic object, since it provides a representation of something.

(19) T(ROUTIER) = (λx. **road**'(x)) = **road**'

(20) a. SORT(CARTE) = o $\wedge$ functional-artefact (fct-art)
b. T_{EXT}(CARTE) = (λx. **map**'(x)) = **map'**
c. Q_{ORI}(CARTE) = (λx. λy. λe. **print**'(e,x,y))
d. Q_{TEL}(CARTE) = (λx. λy. λe. **represent**'(e,x,y))
condition: 'x is a geographical object'

With functional artefact of this kind, I suppose that the rule that combines the semantics of both the N and the denominal A at the NP level is something like (21). The source of the relation is, following the order of preference, the Telic or the Origin quale, and when neither proves suited, a contextual relation is sought for. Consequently, the interpretation of the NP *carte routière* is predicted to be (22), which is correct.

(21) NP's interpretation rule 2
a. $T(N_{FCT\text{-}ART}\ A_{DEN})$ = (λ**R**. λ**N**. λ**A**. λy. $\exists$x. **R**(e,x,y) ə **N**(y) ə **A**(x))
b. **R** = Q_{TEL}(**N**) > Q_{ORI}(**N**) > contextual

(22) T(CARTE ROUTIERE) = (λy. $\exists$x. **represent**'(e,x,y) ə **map'**(y) ə **road'**(x))

3.2. The spatial interpretation

If we consider now the slightly different DA construct *carte murale*, we see that the source of the relation shifts from the HdN to the BseN, inasmuch as there is no sense in which we can say that a map is designed to represent a wall. And it is so because MUR (or WALL) does not denote a geographical object. If we look at the semantic representation of WALL, we can arguably claim that it contains information relative to the constitution of its referent ('continuous vertical (brick or stone) structure…' cf. quale Aspect in Pustejovsky), its function ('… structure that encloses or divides an area or supports a load' cf. quale Telic) and, obviously, its Origin since it denotes an artefact. This gives us a partial representation like (23) for the semantics of WALL / MUR.

(23) a. SORT(MUR) = o $\wedge$ functional-artefact
b. T_{EXT}(MUR) = (λx. **mur'**(x)) = **mur'**
c. Q_{TEL}(MUR) = (λx. λy. λe. **enclose'**(e,x,y) $\wedge$ **area'**(x)…)
d. Q_{ORI}(MUR) = (λx. λy. λe. **build'**(e,x,y))

Since WALL quite regularly denotes a space corresponding to a Ground in a spatial relationship e.g. *painting on the wall*, *nail in the wall*, I would argue that its semantics may be supplemented by the spatial relation indicated in (24a). This relation can be seen as a default relation available whenever none of the internal meanings provides a suitable relation. It may appear with all

---

of this type. It should be noted that example (13) shows that argument *y* in the agent reading (18) can be saturated by a *de* PP at the NP level. However, no saturation is possible with a denominal A: (a) **élection (populaire présidentielle + présidentielle populaire)* is out and cannot mean 'election of the president by (the) people'. The situation is worse for the patient reading: (b) is ungrammatical and (c) cannot mean 'election of the president by French' but only 'election of the president (that takes place) in France': (b) **élection présidentielle des Français*, (c) *élection présidentielle française*.

nouns denoting an object that can easily be interpreted as a Ground e.g. *table*, *roof*, *tree*, etc. A general formulation of this relation is proposed in (24b), where **adloc**' notes the semantics of a locative adposition.

(24) a. $T_{\text{SPATIAL}}(\text{MUR}) = (\lambda x.\ \lambda y.\ \text{LOC}(\textbf{on'}(x),y))$
condition: 'x has a spatial extension'
b. $T_{\text{SPATIAL}}(\text{N}) = (\lambda x.\ \lambda y.\ \text{LOC}(\textbf{adloc'}(x),y))$
where *y* range over events or entities.

I suppose that whenever this relation is chosen, the N is given the subtype 'Space Noun'. I further assume that denominal adjectives based on spatial nouns are specified as such by the morphological rule that builds them (cf. (25a)) and that they receive two semantic translations. The ordinary one, which corresponds to their being denominal (cf. (25b)), and another one, qua spatial denominal, which is equivalent to the spatial representation of the BseN (cf. (25c)). This amounts to say that the denominal A inherits from its base N the capacity to function as a Ground in case the latter possesses this capacity.

(25) a. $mural_{\text{SPACE-DEN}} < \text{mur}_{\text{SPACE-NOUN}}$
b. $T(\text{MURAL}) = (\lambda x.\ \textbf{mur'}(x)) = \textbf{mur'}$
c. $T(\text{MURAL}_{\text{SPCE-DEN}}) = T_{\text{SPATIAL}}(\text{MUR})$

The semantic rule at the NPDA level is a variant of (21). Condition (26b) specifies that the relation is provided by the spatial relation associated to the base noun, if any (in the present case (24a)). Hence, the NP *carte murale* receives interpretation (27), which can be paraphrased by 'map which has the property of being located on (a) wall'.

(26) NP's interpretation rule 3
a. $T(\text{N A}_{\text{SPACE-DEN}}) = (\lambda\mathbf{R}.\ \lambda\mathbf{N}.\ \lambda\mathbf{A}.\ \lambda x.\ \exists y.\ \mathbf{R}(y,x) \ni \mathbf{N}(x) \ni \mathbf{A}(y))$
b. $\mathbf{R} = T_{\text{SPATIAL}}(\mathbf{BseN})$

(27) $T(\text{CARTE MURALE}) =$
$(\lambda\mathbf{R}.\ \lambda\mathbf{N}.\ \lambda\mathbf{A}.\ \lambda x.\ \exists y.\ \mathbf{R}(y,x) \ni \mathbf{N}(x) \ni \mathbf{A}(y))(\lambda x.\ \lambda y.\ \text{LOC}(\textbf{on'}(x),y))(\textbf{map'})(\textbf{wall'})$
$= (\lambda x.\ \exists y.\ \text{LOC}(\textbf{on'}(y),x)) \ni \textbf{map'}(x) \ni \textbf{wall'}(y))$

The DA construct *élection professionnelle* 'election (of representatives) within a profession' will be intepreted in the same way, since *profession* 'profession' can be neither the agent, nor the patient in (14c) (nor a final state, by the way cf. note 2). Insofar as locative phrases such as *(il travaille) dans cette profession* '(he works) in this profession', *au sein de cette profession* 'within this profession', etc. are possible, the suppletive semantic rule (24b) is available. The NP in question will get the interpretation 'election which has the property of taking place within (a) profession(s)'.
This account of NPs involving a spatial interpretation allows us to tackle the more complex issue of prefixed denominal adjectives.

4. Prefixed denominal adjectives

The phenomenon I would like to examine now is illustrated by the words in bold in examples (28):

(28) a. *La cartographie de la pente de la surface par le satellite ERS permet de repérer les lacs* ***sous-glaciaires*** *(…) de l'Antartique*. LE JOURNAL DU CNRS, N° 205-206, February-March 2007, p. 25.
'The cartography of the surface slope by satellite ERS allows one to spot the lakes located under the glaciers (…) of Antartica.'

b. *(…) un cru issu d'une vigne* ***préphylloxérique*** *(…)* MARIANNE, 15-21 September 2007, p. 95.
'a vine coming from a vineyard that existed before the epidemic of phylloxera' [in 1867]

c. *Association* ***Transfrontalière*** *de Protection des Chauves Souris. Vereinigung Grenzüberschreitender Fledermausschutz.* (aptcs.ciril.fr)
'cross-border association for bat protection'

I will discuss only the first case, *lac sous-glaciaire*. I suppose that the information given in (29)-(30) is part of the semantic representation associated to lexemes LAC / LAKE and GLACIER, respectively. Note that both can have the additional spatial meaning, since we have phrases like *(dans + sur) le lac* '(in + on) the lake', *sur le glacier* 'on the glacier'.

(29) a. SORT(LAC) = o ∧ natural-species (ntsp)
b. T_{EXT}(LAC) = (λx. **lake'**(x)) = **lake'**
c. $T_{SPATIAL}$(LAC) = (λx. λy. LOC(**in'**(x),y))

(30) a. SORT(GLACIER) = o ∧ natural-species (ntsp)
b. T_{EXT}(GLACIER) = (λx. **glacier'**(x)) = **glacier'**
c. $T_{SPATIAL}$(LAC) = (λx. λy. LOC(**on'**(x),y))

Contrary to what has been commonly assumed since Corbin (1987) and more specifically (Corbin 1990) (cf. Amiot 1997: 108-119), I contend that spatial / temporal prefixation directly apply to denominal adjectives (e.g. GLACIAIRE) and need not apply to a noun corresponding to the base noun (e.g. GLACIER). Very schematically, the derivation pattern for adjectives prefixed by SOUS- would be something like figure 3.

$$\left[\begin{array}{l} A - space - pfx - den \\ \text{PHON [1] su} \oplus \text{[2]} \\ \text{SYN - SEM} \left[\begin{array}{l} \text{HEAD [CAT a]} \\ \text{CONT [REL Sous - sem([3])} \end{array}\right] \\ \text{MORPH} - \text{ST} \left[\begin{array}{l} A - den \\ \text{PHON [2]} \\ \text{SYN - SEM} \left[\begin{array}{l} \text{HEAD [CAT a} \\ \text{CONT [REL [3]} \end{array}\right] \end{array}\right] \end{array}\right]$$

Figure 3. The rule schema for SOUS prefixation

At the phonological level, the segment /su/ is prefixed onto the phonology of the base e.g. /glasjɛʁ/ for *glaciaire*, which correctly yields *sous-glaciaire* /suglasjɛʁ/. At the semantic level, the semantic function associated with prefix SOUS- (noted Sous-sem in figure 3) is applied to the semantics of the denominal adjective, namely **glacier'**. This Sous-sem function corresponds to the translation of SOUS used as a spatial prefix (cf. (31)). In the case at hand, applying (31) to the semantics of GLACIAIRE yields the semantics of SOUS-GLACIAIRE (32).

(31) $T(\text{SOUS}_{\text{SPCE-PFX}})$ = (λQ. λx. λy. LOC(**under'**(y),x) ∋ **Q**(y))

(32) $T(\text{SOUS-GLACIAIRE}_{\text{SPCE-PFX-DEN}})$
= (λQ. λx. λy. LOC(**under'**(y),x) ∋ **Q**(y))(**glacier'**)
= (λx. λy. LOC(**under'**(y),x) ∋ **glacier'**(y))

As indicated in figure 3, spatial prefixations in SOUS-, TRANS-, INTRA-, INTER-, etc. confer type 'space-prefixed' (SPCE-PFX) on the derived adjective, which therefore becomes a 'space-prefixed-denominal adjective'. At the NP level, the rule combining the semantics of this type of adjective with the HdN is a variant of interpretation rule (26). The variation comes down to the lack of variable **A**, which follows from the fact that the semantic content carried by A has already been incorporated into the semantics of the 'space-prefixed-denominal' adjective e.g. SOUS-GLACIAIRE. The application of (33) to the semantics of LAC gives us the interpretation of *lac sous-glaciaire*, namely (34) 'a lake such that it is located under a glacier'.[5]

(33) NP's interpretation rule 4
  a. T(N $A_{\text{SPCE-PFX-DEN}}$) = (λ**R**. λ**N**. λx. $\exists$y. **R**(x,y) ə **N**(x))
  b. **R** = T($\mathbf{A}_{\text{SPCE-PFX-DEN}}$)

(34) T(LAC SOUS-GLACIAIRE$_{\text{SPCE-PXF-DEN}}$) =
(λ**R**. λ**N**. λx. $\exists$y. **R**(x,y) ə **N**(x))(λx. λy. LOC(**under'**(y),x) ə **glacier'**(y))(**lake'**)
= (λx. y. LOC(**under'**(y),x) ə **glacier'**(y) ə **lake'**(x))

An account of prefixed temporal denominal adjectives e.g. *préphylloxérique*, *pré-électoral* 'pre-election (N)', etc. can be conceived along the same line, provided some adjustments required by the temporal nature of the prefix are made.

5. Non derived denominal adjectives

The lexemes listed in the left column of (35) are underived relational adjectives in French, which have been either inherited from Latin, or borrowed and adapted from Greek or other language.

(35) a. TERRESTRE 'terrestrial' < Lat. *terrestris*
  b. GALACTIQUE 'galactic' < Grk *γαλακτικ ς*
  c. FRONTALIER 'border (N)' < Gasc. *frountalér*

Yet they behave exactly like derived relational adjectives as attested by the grammatical contrasts shown in (36), which replicate what we saw with typical relational adjectives (Ab) (cf. Tableau 3).

(36) a. *La sismologie terrestre* 'Earth's seismology'
  b. **La terrestre sismologie*
  c. **Cette sismologie est terrestre.*
  d. **La sismologie très terrestre*

All these cases, which are very numerous, are obvious instances of suppletion. In the phonological rubrique of the lexeme, a suppletive form is provided (*terrestre*, /tɛʁɛstʁ/; *galactique*, /galaktik/) instead of the expected suffixed form either, because no suffixation process of the type requested exists (e.g. no suffix /ɛstʁ/), or because the potential base does not belong to the stock of native French roots (e.g. °/galakt/, °/fʁɔ̃ tal/ are not correlated to a nominal word-form in French). These suppletions account for the underived character of the adjectives in question. Underived though they are, these adjectives are semantically complex since they are correlated to a noun, and this is why they pattern like normal denominal adjectives. The mismatch between the phonology (simple) and the semantics (complex)

[5] The interpretation of the denominal adjectival construct *lac glaciaire* is completely different since the latter means something like 'lake such that it exists because of glaciers'. The causal relation involved in this interpretation can be thought as an instanciation of the ORIGIN quale for objet that are not artefacts.

exhibited by these lexemes is directly encoded in the representation proposed under figure 4 for GALACTIQUE.

$$\left[\begin{array}{l} A - den \\ \text{PHON [1] galaktik} \\ \text{SYN - SEM}\left[\begin{array}{l}\text{HEAD [CAT a}\\ \text{CONT [REL [2]}\end{array}\right] \\ \text{ST - MORPH}\left[\begin{array}{l} Noun \\ \text{PHON} < > \\ \text{SYN - SEM}\left[\begin{array}{l}\text{HEAD [CAT n}\\ \text{CONT [REL [2] } \mathbf{galaxy'}\end{array}\right]\end{array}\right]\end{array}\right]$$

Figure 4. Structure of A *galactique*

From the point of view of semantics, the A is denominal (and therefore complex) insofar as its content is based upon the content of an existing noun. From the point of view of phonology, the A is simple since its phonology is not the result of adding a suffix to a base, inasmuch as there is no such base : the empty value of PHON attests that the form has been borrowed or inherited as a whole. For sure, the phonological ending may look like an existing suffix e.g. /ik/ in *basaltique* 'basaltic', *cyclique* 'cyclic, cyclical', etc., or /je/ in *langagier* 'linguistic', *princier* 'princely', etc.[6] Or it may not, as is the case for /ɛstʁ/, which shows up in 7 lexemes only, all inherited from Latin.

The fact that the adjectives of (36) are semantically complex allows us to cope with the paradigmatic effect observed in (37).

(37) a. *glacier / glaciaire / sous-glaciaire*
b. *galaxie / galactique / intergalactique*
c. *terre / terrestre / extraterrestre*

The adjectives in (37b-c) pattern in the same way as the one in (37a), even though they are not derived upon the base-noun mentioned on the left. On the model of what we had for prefix SOUS in (31), let us suppose that the spatial prefix INTER- has the semantics given in (38).

(38) $T(\text{INTER}_{\text{SPCE-PFX}}) = (\lambda Q.\ \lambda x.\ \lambda y.\ \text{LOC}(\mathbf{between'}(y),x) \ni \mathbf{Q}(y))$

GALACTIQUE is stored as a denominal adjective in the lexicon (cf. figure 4). Applying (38) to GALACTIQUE yields the interpretation given in (39).

(39) $T(\text{INTER-GALACTIQUE}_{\text{SPCE-PFX-DEN}})$
$= (\lambda Q.\ \lambda x.\ \lambda y.\ \text{LOC}(\mathbf{under'}(y),x) \ni \mathbf{Q}(y))(\mathbf{galaxy'})$
$= (\lambda x.\ \lambda y.\ \text{LOC}(\mathbf{between'}(y),x) \ni \mathbf{galaxy'}(y))$

[6] Corbin calls this type of ending *intégrateur paradigmatique* 'paradigmatic integrator' (Corbin 1990). A paradigmatic integrator has the phonological appearance of a suffix but none of its other properties. It allows a form e.g. *peuplier* 'poplar' to join up a series of full-fledged derived forms e.g. *pommier* 'apple-tree', *prunier* 'plum-tree', etc. on the sole basis of the formal similarity it shares with other forms of the series. In a lexematic framework, the phenomenon paradigmatic integrators aim at capturing can be expressed without postulating this device. Suffices it to take advantage of the surface similarity exhibited by the forms in question (cf. Fradin 2003 : 140-145).

At the NP level, rule (33) straightforwardly derives interpretation (40) for *voyage intergalactique*, exactly as it did for *lac sous-glaciaire* ('a travel such that it [takes place] between galaxies').

(40) a. $T(\text{VOYAGE INTERGALACTIQUE}_{\text{SPCE-PXF-DEN}})$ =
b. $(\lambda\mathbf{R}.\ \lambda\mathbf{N}.\ \lambda z.\ \exists y.\ \mathbf{R}(z,y) \ni \mathbf{N}(z))(\lambda x.\ \lambda y.\ \text{LOC}(\mathbf{between'}(y),x) \ni \mathbf{galaxy'}(y))$ $(\mathbf{travel'})$
c. $= (\lambda z.\ \ y.\ \text{LOC}(\mathbf{between'}(y),z) \ni \mathbf{galaxy'}(y) \ni \mathbf{travel'}(z))$

The appropriate interpretation is obtained without having to postulate any spurious structure or additional device such as the Copy Principle once proposed by Corbin (1987: 136).

6. Conclusion

This paper is a first attempt to propose an explicit compositional account of the semantics of denominal adjectives. It puts forward three main ideas : (i) the semantic representation of denominal adjectives is identical to that of their base-noun; (ii) the morphological operations by which complex lexemes are built change the semantic / categorial nature of the lexeme they apply to and the type of the latter keeps trace of this change at each step of the derivation; (iii) interpretive rules at the NP level attune to this information and adapt the interpretation in consequence.

One of the conclusions that can be drawn from the discussion carried out in this article is that the interpretation of NPs with a denominal adjective can be accounted for in a large part on the basis of the relation supplied either by the head noun or the base noun.

In comparison with previous approaches, most notably Corbin (1987, 1990), the present account shifts from a segmental analysis to a semantic one, a move which is in keeping with the interpretive nature of the phenomenon in question. This move has been possible because a lexematic framework has been adopted, which allows things to be stated in a much simpler way than in a morphemic framework. It seems fair to say that the latter does not permit to give a compositional analysis of the meaning of denominal adjectives without adding a lot of artefactual devices. In this respect, the present account allows us to definitely get rid of the Copy Principle, which raises more problems than it solves (Fradin 1996).

However many issues remain to be settled: the semantic types / sorts attributed to the lexical items have to be justified more thoroughly and the hierarchy they constitute made explicit; the interpretive rules are given piecemeal; neither the reason why they pick up the external vs. the internal meaning nor the way they combine them follows in a principled way from general properties; the conditions that trigger these rules are just postulated instead of being motivated. All these shortcomings remind us that this work is just a beginning and that many examples have to be analysed in detail before we can discover the generalisations at stake behind the phenomena in question.

Abbreviations

A = adjective, AZR = adjectivizer, DAC = denominal adjective construct, NPDA = noun phrase with denominal adjective, F = feminine, M = masculine, N = noun, SG = singulier, PFX = prefix.

References

Amiot Dany. 1997. *L'antériorité temporelle dans la préfixation en français*. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Corbin Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. 2 vols. Lille: Presses Universitaires du Septentrion. Original edition, Tübingen: Niemeyer.

Corbin Danielle. 1990. Homonymie structurelle et définition des mots construits. In *La définition*, J. Chaurand & F. Mazière (eds). Paris: Larousse. pp. 175-192.

Dowty David R. 1979. *Word Meaning and Montague Grammar*. Dordrecht / Boston / London: Reidel.

Flaux Nelly & Danièle Van de Velde. 2000. *Les noms en français. Esquisse de classement*. Paris / Gap: Ophrys.

Fradin Bernard. 1996. On morphological entities and the Copy Principle. *Acta linguistica Hungarica* 43 (1-2), pp. 111-151.

Fradin Bernard. 2003. *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.

Fradin Bernard. 2007. Three puzzles about denominal adjectives in **-eux**. *Acta Linguistica Hungarica* 54 (1), pp. 3-32.

Fradin Bernard. 2008 (to appear). Les adjectifs relationnels et la morphologie. In *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, B. Fradin (ed). Amsterdam / Philadelphia: John Benjamins. pp. 67-89.

Godard Danièle & Jacques Jayez. 1993. Towards a proper treatment of coercion phenomena. In *Proceedings of 6th EACL 1993*. pp. 168-177.

Jensen Per Anken & Carl Vikner. 2002. *The English Prenominal Genitive and Lexical Semantics*. Manuscript Engstien / Copenhagen.

McNally Louise & Gemma Boleda. 2004. Relational adjectives as properties of kinds. In *Empirical Issues in Formal Syntax and Semantics*, Vol. 5, O. Bonami & P. Cabredo Hofherr (eds). Paris: CNRS & Université de Paris 7. URL, http://www.cssp.cnrs.fr/eiss5. pp. 179-196.

Mezhevich Ilana. 2004. English Compounds and Russian Relational Adjectives. In *Proceedings of the North Western Linguistic Conference 2002*: URL, http://hdl.handle.net./1892/87. pp. 95-114.

Partee Barbara H. & Vladimir Borschev. 2001. Some puzzles of predicate possessives. In *Perspectives on Semantics, Pragmatics and Discourse: A Festschrift for Ferenc Kiefer*, I. Kenesei & R. M. Harnisch (eds). Amsterdam / Philadelphia. pp. 91-117.

Partee Barbara H. & Vladimir Borschev. 2003. Genitive, relational nouns, and argument-modifier ambiguity. In *Modifying Adjuncts*, E. Lang, C. Maienborn & C. Fabricius-Hansen (eds). Berlin / New York: Mouton / De Gruyter. pp. 67-112.

Pustejovsky James. 1995. *The Generative Lexicon*. Cambridge (Mass.): MIT Press.

Sag Ivan & Thomas Wasow. 1999. *Syntactic Theory. A Formal Introduction*. Stanford: CSLI Publications.

Thornton Anna. 1999. On Italian derivatives with antesuffixal glides. *Yearbook of Morphology* (1998), pp. 103-126.

Why do Languages Develop and Maintain Non-Concatenative Morphology?

Robert R. Ratcliffe
Tokyo University of Foreign Studies
ratcliffe@tufs.ac.jp

0. Abstract

It seems to be widely assumed that morphological apophony of the foot/feet type always arises diachronically from concatenative morphology. Comrie (1992) explicitly proposes this, concluding logically that if one goes back far enough into the history of any particular language one must reach a point at which there was no non-concatenative morphology. Meanwhile, a recurrent trend of research in synchronic morphology has been to show that at the deepest level all non-concatenative morphology can be explained by syntactic—combinatorial—principles (Lieber 1992, Stonham 1994), hence that at the level of universal grammar (the innate knowledge of linguistic principles hypothesized to underly language acquisition) there are no principles specific to morphology, much less to non-concatenative morphology. But if non-concatenative morphology was not in any language to begin with and if there are no innate principles which allow speakers to create or interpret morphology of this type, how could it ever possibly have arisen? In other words, to take a concrete example, why has the original subphonemic variation that gave rise to Germanic umlaut plurals been phonologized and then morphologized by subsequent generations rather than simply ignored?

I propose that the solution to this paradox is to be found in empirical evidence from first language acquisition and a maturationist theory of acquisition as developed by researchers in this field (Piaget 1948, Ingram 1989).The maturationist theory says that the language faculty (defined not as innate knowledge but as the cognitive abilities required to process language structures) develops in stages as an infant matures. The ability to process phonology and to memorize words appears before the ability to process syntax or to apply morphological rules. For morphology specifically, it has been shown that ablaut past tense and irregular plural patterns in English are initially learned without difficulty, and only at a later stage do children over-generalize -s plurals or –d past tense forms, before eventually settling on the standard system (Bowerman 1982, Brown 1973). I propose that the structure of normally transmitted languages reflects this developmental process. Children create non-combinatorial morphology from combinatorial input, because the former is actually easier for the immature language faculty to process. Non-concatenative systems tend to become structured in such a way as to facilitate memorization—by being organized on the principle of marking minimal functional/ semantic differences by minimal phonological differences. This hypothesis explains certain widely attested and systematic properties of non-concatenative (sub-) systems which can find no coherent explanation when such systems are analyzed as being underlyingly combinatorial. The hypothesis can also be extended to account for one of the characteristic differences between naturally transmitted languages, on the one hand, and pidgins and creoles, on the other. The latter are characteristically less rich in morphological devices than their sources. This is due not to the fact that creoles are created de novo by children using only the resources of UG (Bickerton 1981, 1990) but because creoles by definition derive from pidgins; and pidgins are created by adults, who have the full cognitive resources of the mature language faculty. Artificial languages like Esperanto, also created according to adult conceptions of simplicity, share this feature with pidgins. Finally, and quite speculatively I propose (in agreement with Carstairs-McCarthy 2005) that the principles governing non-concatenative morphology may reflect an older pre-syntactic phase of human speech.

1. Theoretical Preliminaries: Language acquisition and language change

Three major hypotheses are implicated in the paradox laid out in the first paragraph, along with several variants and implications of each. The first hypothesis is that child language acquisition is the locus of grammatical reanalysis which gives rise to grammatical change. This is a venerable hypothesis (Paul 1880, Kiparsky 1968, Andersen 1973). It is not uncontroversial. It has been challenged particularly from the side of sociolinguistics (see, for example the discussion in Hock 1991, ch.20). But I am not going to challenge it here. Rather I am going to assume this to be the case in order to focus on a paradox that arises from this assumption: Children can only be expected to change language if there is something different about the way that children and adults process language. (That is, if the initial state of the language faculty and the mature state are different.)

The second is Comrie's hypothesis regarding the historical source of non-concatenative morphology. This should be broken down into several variants. The first is that concatenative morphology is the *only* source of non-concatenative morphology. I am personally skeptical of this, and can at least imagine two other ways in which such morphology might develop. One is by analogic extension of what were originally (accidental) phonaesthetic similarities: *glow*, *gleam*, *glitter* >> *glitz*. The other is that one type of non-concatenative morphology may develop into another through re-analysis. This is more easily documented. Semitic and Afroasiatic languages provide many examples. One type of change, for example, involves the shift from a fixed syllabic/vocalic pattern to reduplication as a means of marking noun plurals in some Semitic, and possibly other Afroasiatic languages (Ratcliffe 1996). On the analogy of Classical Arabic qitˤtˤ pl.qitˤatˤ "cat" (f.), (formed on the same pattern, CiCaC, as *xirqa* pl. *xiraq* "rag"), one gets Maltese reduplicated plurals like *umbrella*, pl. *umbrelel* "umbrella" (Mifsud 1994). The fixed pattern plurals of Classical Arabic, in turn, are likely to have arisen from an earlier process of forming plurals by -a(a)- infixation (Greenberg 1955, Ratcliffe 1998). (i.e. CA *kalb*, pl. *kilaab* probably goes back to something like *kalb*, pl. **kalaab*).

A second, weaker variant of Comrie's hypothesis is that concatenative morphology is one possible source of non-concatenative morphology. This is demonstrably true (as the Germanic umlaut case shows), and as long as it is so, the paradox persists.

A third variant of the hypothesis, somewhat stronger than the previous, is that the change from concatenative to non-concatenative morphology is unidirectional. Speakers may reanalyze a concatenative opposition as non-concatenative (*fot/foti* >> *fot/føti*>> *fot/føt*) but the reverse will not occur (*f*ot/føt* >> *fot/foti*). From the evidence available to me at present this unidirectionality hypothesis does appear to hold. If so the paradox is strengthened. Of course this proposal depends upon drawing a sharp distinction in principle between *reanalysis* of an existing opposition and complete *replacement* of one term of an opposition through analogical extension (*hand/handa* >> *hand/hands*).

The third major hypothesis implicated in these issues is of course Chomsky's initial state/universal grammar theory of language acquisition. This theory asserts that the ability to acquire a language is equivalent to innate knowledge of universal language structures. What specifically these structures might be has never been explicitly specified, but the assumption has been that they are restricted to principles of syntax, since syntax supposedly shows less variation cross linguistically than lexicon or phonology (Chomsky 1993, p.1-3)[1]. It is this hypothesis which motivates word-syntax approaches to morphology—i.e. the attempt to incorporate morphological structures into the hypothetically innate and universal structures of syntax. Hence this is the ultimate source of the paradox pointed out in the first paragraph.

---

[1] *"UG is a theory of the initial state (S0) of the relevant component of the language faculty….."*
" UG is concerned with the invariant principles of S0 and the range of permissible variation. Variation must be determined by what is "visible" to the child acquiring language, that is, PLD. It is not surprising then, to find a degree of variation in the P[honological] F[orm] component, and in aspects of the lexicon…. Variation in the overt syntax or L[ogical] F[orm] component would be more problematic, since evidence could only be quite indirect. A narrow conjecture is that there is no such variation: beyond PF options and lexical arbitrariness… variation is limited to nonsubstantive parts of the lexicon and general properties of lexical items."

Here too several aspects of the hypothesis should be separated out. First the initial state theory, often termed a constructivist theory of acquisition, assumes that the language faculty is unchanging, essentially the same in infants and adults. Acquisition involves the fleshing out of a pre-existent syntactic skeleton with the language particular details of phonology and lexicon. One problem with this theory is that it offers no explanation for the developmental stages observed by language acquisition researchers (section 3, below). Another problem, relevant here, is that it makes it very difficult to attribute grammatical change to reanalysis by children. If children bring to the task of acquisition an adult level competence, there is no reason why they should analyze adult structures differently from adults. At least there is no reason why grammar variation should be any greater between adults and children than grammar variation among adults of the same community. If grammatical variation is simply endemic in speech communities, with no age or generational bias, there is no reason why change should occur over time, and certainly no reason why it should have any unidirectional bias—as seems to be the case in the change concatenative to non-concatenative (as well as in other types of change such as grammaticalization, Hopper & Traugott 1993).

A second aspect of the initial state theory which is more subtle, but ultimately more important is that it incorporates an assumption that language and the ability to acquire it belong to the same ontological domain, in other words, that they are the same kind of thing. This may turn out to be the best hypothesis, but there is no a priori logical reason why it must be true—as Chomsky and other generativists appear to assume. There may be abundant empirical support for the idea that humans have an innate ability to acquire a language, that this is a species specific ability, etc. But none of this tells us what kind of thing this ability is, nor requires us to believe that it is the same kind of thing, or some subset of, the knowledge of a language which is eventually acquired. The standard Generative formula "language is part of the mind/brain" simply obscures the crucial issue behind a slash mark. Simply to assert that something (language) is a mental object does not necessarily tell us anything about the neurological structures of the brain which support it.

The alternative view, that a language and the ability to acquire it are different kinds of things is a priori equally plausible. In normal speech when we say that something is (all) in the mind we do not mean that it is part of the neurological structure of the brain. Rather we mean precisely that it is imaginary, not located in the real world at all. I would suggest that it is in this sense too that we should interpret the notion that "language is in the mind." This viewpoint is widely attributed to the philosopher Karl Popper, who draws an ontological distinction between world one—the domain of real or objective objects, which exist regardless of whether people are there to observe them—and world three or imaginary objects, including language, which exist and can exist only in the human mind. (World two is the domain of emotions and feelings.) On this schema the ability to acquire and process a language is a world one object, specifically a neurological function of some kind. A language, on the other hand, is a world three object, that is an imaginary object—an artificial set of cultural norms which members of a speech community agree to imagine. On this view language is structured, of course, but we can not assume that the structure which linguists discover through analysis of languages directly reflect neurological structures in the brain. The sources of structure vary. The process by which a language is acquired may effect how it comes to be structured as it is.

If we adopt the hypothesis that knowledge of a language and the ability to acquire that knowledge belong to different ontological domains, are different kinds of things, it opens the way to looking at the maturation of the language faculty and the acquisition of a particular language as independent processes. Grammar does not grow in the brain. The brain grows, develops. As this is happening the child masters the linguistic norms of the community into which it is born. Normally-- in the course of normal language transmission-- the two processes occur to together. As a child matures it also learns a first language. But in the early stages a child may lack the cognitive ability necessary to process adult structures in the same way that adults do. At the same time the child may be highly sensitive to features of adult speech that are redundant or insignificant for the adult speaker. The ability to recognize the relatively subtle

difference in vowel quality between o/ø in a hypothetical pair like *fot/føti* matures before the ability to analyze the second member of the pair into a sequence of constituents: stem plus suffix. Memory also comes into play earlier than analytical ability. Hence it may be easier for young children simply to learn morphologically related words as individual non-complex words rather than to break them down into morphemes on the basis of distributional analysis as adult linguists do.

2. A Hypothesis

In short I propose that a Popperian theory of the ontology of language and Piagetian theory of acquisition offer a framework for the resolution to the paradox taken up in the introduction. The hypothesis is as follows: Non-concatenative morphology emerges and perseveres in languages because it is actually easier for the immature language faculty to process. Sensitivity to the fine points of human speech sounds and memory emerge earlier than the analytical and synthetic abilities necessary to process syntax and other combinatorial processes. Words which adults are able to analyze into components are heard and learned by young children as single items. But small phonetic differences among words are carefully noted by very young first language learners, and when such phonetic differences correlate in a systematic way with small differences in meaning (the same referent in singular and plural aspect, for example: man/men), this may facilitate the learning of such items, that is, their storage in, and retrieval from memory.

A theory of the language faculty, then, should recognize two distinct types of cognitive abilities which govern the organization of morphology. The first involves the recognition of minimal and systematic phonetic differences and similarities among language elements (distinctive features). This ability underlies the creation and processing of phonological inventories and certain types of paradigmatic morphological subsystems, including most non-concatenative morphology. The second involves the recognition that the sequence in which two or more elements occur can signify a relationship between them, and that the whole complex may indicate more than its parts. This ability underlies the creation and processing of phrases, sentences, and other syntactic structures and certain types of syntagmatic morphological subsystems, including compounding and much concatenative morphology. By hypothesis the first of these cognitive abilities develops earlier than the second as an individual matures. Thus those features of a language which require the first type of cognitive ability are learned sooner than the second and persist into adulthood.

This explains why concatenative structures may be reanalyzed as non-concatenative by very young children, while the opposite will not occur. Older children and adults, having both kinds of cognitive faculties, may simply replace non-concatenative by concatenative structures by analogy. (They may also analogically extend non-concatenative structures at the expense of concatenative ones, but actually seem to do so more rarely.) But since each generation of younger learners recreates non-concatenative paradigmatic contrasts, it takes time for analogic concatenative forms to spread through a speech community, and hence many non-concatenative patterns may persist for centuries or millennia. The exception is pidgins—where non-concatenative morphological patterns are, for the most part, simply not transmitted—as is expected if it is primarily very young children who are responsible for originating these patterns.

Thus the hypothesis explains:

1. why acquisition proceeds as it does (with production of single words proceeding the production of longer utterances)
2. why the change from concatenative to non-concatenative morphology is unidirectional
3. why non-concatenative (sub)systems share many properties cross-linguistically

4. why pidgins and creoles are so different in their morphological profiles from naturally transmitted in languages, and
5. (possibly) something about how language evolved.
Let us look at each of these areas in turn.

3.The stages of acquisition

The observation that children's linguistic abilities develop through a series of discrete stages dates from the earliest systematic research on child language acquisition at the beginning of the last century (Stern & Stern 1907, Nice 1917, Piaget 1929), and this observation has been repeatedly confirmed by studies of the acquisition of various languages over more than a century (Ingram 1989: 32-58). The hypothesis that the difference in observable linguistic behaviors reflects differences in cognitive ability, that is differences in cognitive, ultimately neurological, development in the growing infant is associated with the developmental psychologist Jean Piaget.

There is some difference among researchers as to the number of stages and how they are defined. But according to Ingram (1989), the following points appear to be uncontroversial. During the first year of life infants are unable to produce words, but they show an ability to distinguish the sounds of human speech from other types of sounds. (Jusczyk (2003: 63) in fact argues that infants' capacity for discriminating speech sounds is present from birth.) Babbling during this period reflects the infant's attempts to practice the use of the articulatory mechanism. The second stage of acquisition, roughly to the middle of the second year, is characterized as the period of single word utterances. Ingram (1989:41) proposes that this period correlates with the development of the ability to think symbolically, the principle of the linguistic sign. This period is followed by the period of the first word combinations. Most of the controversy then centers on how many stages there are between this stage and the final stage when a child has completely mastered adult syntax, and how much syntactic knowledge may or may not underly the child's production of multi-word utterances that do not conform to the rules of adult syntax. For the purpose of the present issue, however, it is the difference between the first two stages and the later ones that is significant. With regard to the acquisition of morphology in particular, it has been shown that there is an S-curve pattern, such that English irregular past tense or plural forms are initially learned without difficulty, are then replaced by analogically derived –s and –d forms, before the "standard" adult system is eventually mastered (Bowerman 1982, Brown 1973). Ingram (1989:65) observes:

"…. children do not produce overgeneralizations like 'foots' at the onset of plural acquisition, but instead, either omit the plural or else use it correctly. To account for this, we need another principle which elsewhere (Ingram 1985a) I have called the lexical principle: learn individual paradigmatic alternations as separate lexical items. This principle claims that the child first acquires paradigmatic variants like 'cat, cats', 'dog, dogs' as separate words, and only later realizes that there is a separable plural morpheme '-s'. The lexical principle predicts that the child will initially get 'foot, feet' correctly, and only later change the latter to 'foots'. (This is in fact, as we will see, true.)"

In short acquisition research consistently shows that the child's awareness of the phonological distinctions of the target language precedes acquisition of vocabulary. And the acquisition, or at least the production, of single words precedes any effort at combing words. The hypothesis is that while an adult will be able to analyze a form like *fot-i* (*ˌføti*) into constituents, abstracting out the phonetic effect of the suffix on the stem vowel, a child whose cognitive abilities do not yet allow it to process combinations of linguistic elements will hear both *foti*(*ˌføti*) and *fot* as unanalyzable lexical items, minimally differing from each other in form and meaning. An understanding of the principle of minimal contrast, necessary to the acquisition of phonology, precedes understanding of the principle of combination. Young learners first exploit the principle of contrast in the acquisition of lexicon.

4.What Happens in Diachronic Change?

The historical development of the Germanic "umlaut plurals" of the foot-feet type is probably the best-studied case of non-concatenative morphology emerging from concatenative. The processes involved in the development of this phenomenon can be classified under four headings: 1.phonetic effect, 2. phonemicization, 3. morphologization, 4. lexicalization and loss. These occur more or less in chronological sequence, but there are differences in how and to what extent each is played out in the different Germanic languages:

i. Phonetic effect rooted in articulation factors: Speakers unconsciously front a stem vowel in anticipation of a front vowel (-i) suffix (regressive long-distance assimilation). To begin with this is a subphonemic, i.e. phonetic effect. We infer this because umlaut emerges in all the living Germanic languages, indicating that it must have been present in some form in proto-Germanic. Yet the earliest written records of Germanic languages, composed after the individual branches had begun to diverge, do not record umlaut, indicating that the originators of the orthographic conventions did not perceive the vowels in *fot* and *foti* (presumably *[føti]) as phonemically distinct. (Lass 1997: 299: "Neither Proto-Germanic nor Northwest Germanic had /y, ø/; but these emerged in pre-historic old English as the result of i-umlaut of /u,o/…).

ii. Phonemicization of the vowel contrast. Speakers reanalyze the vowels in *fot/ føti* as phonemically distinct. We infer that this must have happened because in many Germanic languages, the front stem vowel is maintained in those parts of the paradigm where a front vowel suffix was originally present, even though this suffix, and hence the conditioning environment for the change, has in some parts of the paradigm been lost or replaced. This is the case for example in Modern Icelandic and Old English (adapted from Lass 1997: 342-44):

OE

	sg.	pl.
nom.	fo:t	fe:t
gen.	fo:t-es	fo:t-a
dat.	fe:t	fo:t-um
acc.	fo:t	fe:t

Modern Icelandic

	sg.	pl.
nom.	fotur	foetur
gen.	fotar	fota
dat.	foeti	fotum
acc.	fot	foetur

In these cases clearly we can not say that umlaut is a mechanical effect, nor can it be treated as the learning of a morpho-phonological rule, because there is no consistent environment which would trigger such a rule. The front rounded vowel in the plural of "foot" is a distinct phoneme in Icelandic, as it is in modern German. In Old English the front rounded vowels have merged with their unrounded front counterparts.

iii. Morphologization:

a. Vowel alternation becomes the sole marker of morphological contrast. This is also clear in the paradigms above. Thus in the Old English example above vowel alternation alone marks the number contrast in the nominative/accusative and the dative vs. nom./acc. contrast in the singular.

b. Vowel alternation is maintained in part of the paradigm (plural), but lost in others (case), thus becoming a more transparent marker of a grammatical category, namely plural. This is what obtains in modern German where umlaut no longer marks dative singular nouns-- although dative case as such still exists-- but is frequent as a plural marker. (But umlaut is nonetheless a redundant marker of number in German, since it always occurs in conjunction with suffixes.) In English where case marking is lost in nouns, umlaut survives only as a marker of number. (And is the only marker of number in the handful of nouns which retain it.)

c. Vowel alternation is extended by analogy as a plural marker to some words which did not originally have it. Trask 1996: 113: "… the old High German noun boum 'tree' had a plural bouma, with no umlaut and this should have come into the modern language as Baum, *Baume. Instead German has Baum Bäume." In English this kind of analogical extension is possible at least in a jocular sense: *moose/?meese* on the analogy of *goose/geese*.

iv. Lexicalization and Loss

Stem vowel alternation is largely replaced by segmental endings, as in English, where only a small closed class set of words retains it. These plurals might even be described as lexical or suppletive in English. In Dutch and the mainland Scandinavian languages these plurals have been replaced by suffixed forms.

Thus Germanic umlaut is initially a conditioned sound change. In classical Neogrammarian theory it is usual to regard sound change as an automatic, mechanical process that applies blindly and for this reason is capable of reeking havoc on other systems especially morphology. The ability of sound change to reek havoc on other systems is undeniable. But sound change (true phonemic change, as opposed to performance-based synchronic phonetic variation) is change in grammar—change in the way in which linguistic information is organized in the mind. In order to learn syntax, children have to learn to identify words. In order to identify words speakers must learn to identify a given word as somehow the same even though it may be pronounced differently in different syntactic contexts (depending on things like sentence intonation, stress, whether a preceding or following word occurs, how that word is stressed, whether it ends or begins with a vowel or consonant, etc.). Learners learn to ignore or "abstract away from" these minor variations in pronunciation. Germanic umlaut begins as a purely phonetic, sub-phonemic variation of this kind. Some speakers unconsciously front the quality of the vowel in a stem in anticipation of a high front vowel in a suffix. To the extent that later speakers do not abstract away from the slight change in stem vowel in the *fot-i* case (and similar cases elsewhere), this indicates that they are not processing this as a syntax-like sequence of the already known word *fot* plus a second word-like thing –i. Rather it indicates that they are learning *fot* and *foti* (~*føti*) as separate lexical items with different forms and different meanings.

Moreover once such alternations become established, they can be extended by analogy, as the example of German *Baum*, pl. *Baüme* (**Baume* expected) shows. Trask (1996:112) treats this as an example of Kuryłowicz's first law of analogy: "A complex marking replaces a simple marking." But it can also be taken as a counterexample to Kuryłowicz's second law of analogy: "a derived form is reshaped to make it more transparent and especially more similar to the simple forms from which it is derived." (*Baüme* is less transparent and less similar to *Baum* than **Baume* is.) It has often been observed that many of Kuryłowicz's "laws" are mutually contradictory. But Kuryłowicz, I think, is not be faulted for this. The first and second law simply reflect two countervailing trends in morphological change, both attested in the data. The questions this raises are: 1. Why should both trends exist? and 2. Are there distinct linguistic situations or distinct areas of linguistic structure where one or the other is more likely to prevail?

My answer to the first question is that the two contradictory trends reflect the difference between paradigmatic vs. syntagmatic principles in the organization of morphological knowledge. As regards the second question, it is clear that changes of the type covered by Kuryłowicz's first law will result in greater distinctiveness between members of a paradigm, while changes of the second type will make it easier to analyze words into syntactic atoms. By hypothesis, changes of the first type are typical of young language learners and are more likely to be initiated and accepted by them. Changes of the second type can only be initiated by cognitively more mature speakers.

As far as the area of linguistic structure where each type is likely to occur, let us start by observing that umlaut has been lost as a marker of dative (a change of the K2 type) in German, but extended as a marker of plural (K1). These two changes together could be taken as an example of Kuryłowicz's fifth law of analogy: "in order to re-establish a distinction of central significance, the language gives up a distinction of

more marginal significance" (Trask 1996:113), although what is lost or extended is not the distinction itself but only a partially redundant way of marking it. And it is not so much a matter of maintaining a central distinction (however that is defined) as maintaining a distinction at the derivational (semantic) end as opposed to the inflectional (syntactic) end of the derivation-inflection continuum. Number is externally referential: *dog* and *dogs* or *foot* and *feet* refer to different objects in the real world. In order to grasp such a difference a child need merely be able to understand the principle of the linguistic sign. Case is only internally referential. In order to understand the difference between dative and nominative case, a child needs to understand the syntactic roles that nouns can have in larger structures such as phrases and sentences. By hypothesis the cognitive ability to recognize a sign matures before the ability to combine signs into larger expressive units. By the time one is prepared to learn the syntactic function 'dative' one is also prepared to parse a word into syntagms, hence non-concatenative marking of case is less likely to arise, persevere, or be extended by analogy, than non-concatenative marking of number, which requires only that the mind be prepared to recognize that *dog* and *dogs* or *foot* and feet refer to different things.

5. Universal properties of non-concatenative and otherwise "idiosyncratic" morphology

On the theory that only syntactic principles are universal (by virtue of being innate), deviations from combinatorial principles in morphology must be explained as language-particular accretions on innate universal structure. The major problem with this, as I have argued up to now, is that it offers no way to explain how such accretions could possibly have occurred. But a second problem with the strict combinatorial approach is that it offers no way to account for similarities across languages in the forms which non-combinatorial morphology can take.

The approach outlined here, by contrast, encourages us to think that there really *should* be universal cross-linguistic properties of non-concatenative morphology, and these should be visible even in languages where such systems are small and un-productive. By hypothesis these universal properties are ultimately rooted in genetically determined cognitive faculties which come into play in the course of acquisition. Small, non-productive systems are not necessarily any less systematic or logical than large, productive ones. It may simply be that they are organized according to cognitive principles which are no longer central to adult competence.

In the absence of a comprehensive typological study any discussion of possible universals will necessarily be somewhat impressionistic. But let me begin by noting four widely found properties of natural language morphology which are not predicted by and cannot be economically explained by combinatorial principles, but which make perfect sense in the framework outlined here, which emphasizes the role of contrastive/ paradigmatic principles in the organization of linguistic knowledge, particularly in the early stages of acquisition. The first two of these refer to non-concatenative systems specifically, the other two to morphological systems generally.

i. Cross-linguistically non-concatenative morphological processes like ablaut or reduplication tend to be found in certain categories more than others. Bybee (1985:36-37) observes that the most likely categories for stem change in inflectional verbal morphology are aspect (found in at least 10 of the 50 languages in her sample), tense, and mood (found in at least three languages each) and that the least likely categories are person and number. In Classical Arabic stem change is involved in the marking of tense (vowel alternation: *kat<u>a</u>ba/yakt<u>u</u>b* "he wrote/he writes"), voice (vowel alternation: *k<u>a</u>t<u>a</u>ba/k<u>u</u>t<u>i</u>ba* "he wrote/it was written"), and valence (alternation in quantity: *k<u>a</u>taba/ kat<u>t</u>aba/ k<u>aa</u>taba* "he wrote/he caused s.o. to write/he wrote to s.o."), but not in the marking of person or number. In Germanic it is found in the marking of tense (*ride/rode*), deriving from Indo-European ablaut, and in valence (*sit/set, lie/lay*),

deriving from umlaut induced by a causative suffix which is later lost. In Japanese it is found only in the marking of valence (*hajimaru/hajimeru* "start (intr.)/start (tr.)"

I suspect that a similar pattern holds for nouns, and that non-concatenative morphology is more likely to be found in categories like number, and possibly gender, rather than, for example, case. I am not aware of any systematic study of this question. But in Afroasiatic languages (which I have surveyed in this regard, Ratcliffe 1992) internal stem change in the plural is widespread, found in dozens of languages in the Chadic, Cushitic, Berber, and Semitic subgroups (Greenberg 1955, Newman 1990, Zaborski 1986, Ratcliffe 1992). But the only example of internal stem change marking case in these languages that I have come across is in Somali, where for some nouns the distinction between absolutive, nominative, and genitive is marked solely by tonal difference (Mous ms., Saeed 1999). Stem change marking gender is slightly less infrequent, found in a few Cushitic languages, where tone is again used (Somali ínan "boy," inán "girl," Mous ms.) and very rarely and sporadically in Semitic (Classical Arabic ʕariis "groom"/ ʕaruus "bride").

What unites the categories where non-concatenative morphology appears to be most frequent is that they fall somewhere in the middle of the derivation/inflection, or semantic/syntactic continuum. Person marking on the verb is a major constituent of the verb phrase, taking the place of a subject NP. Case marking only makes sense if one understands the syntactic relationships in a sentence. These categories are thus closely connected with syntactic structure. Categories like nominal number, and verbal tense and aspect, on the other hand have external reference. Yet unlike, say, the members of a compound like *black-berry*, the inflections indicating verbal tense, aspect, nominal number, and so forth do not by themselves refer to an object or action. Rather they refer to a particular way of conceptualizing an object or action. The word /dogz/ refers to two or more individuals of the class "dog," but we can not decompose this into two signs with direct semantic reference to the objective world. The first part /dog/ clearly is a minimal sign which means "one dog." But then /z/ would have to mean "at least one other and possibly all the other dogs," which is obviously not the meaning of /z/ in /frogz/ or countless other cases. In other words /z/ does not itself directly refer to anything in the external word. Rather singular and plural refer to different ways of conceptualizing a particular referent: as an individual or en masse. The same is true for verbal categories like aspect, which distinguishes between an action occurring once or instantaneously and the same action occurring frequently or over a period of time. In short, it seems that the morphological categories which are most likely to be expressed in terms of non-concatenative oppositions are those which are least interpretable as atoms of syntactic structure or sequences of discrete signs.

ii. The phonological structures which are exploited in non-concatenative alternations seem to be drawn from the same set cross-linguistically. These include quantitative alternation with no change in quality (vowel lengthening, gemination, and reduplication) and qualitative alternation with no change of quantity (alternation of a vowel segment or of a single vowel or consonant feature, and alternation of stress or tone). McCarthy & Prince (1998) propose that alternations of the first type are based on universal prosodic categories: mora, syllable, foot, and prosodic word. Thus alternations based on vowel length or gemination, like the Arabic *kattaba*, *kaataba* opposed to *kataba*, cited above, can be described in terms of a difference of a single mora. Most examples of partial reduplication involve the addition of an empty heavy or light syllable. Apophonoy (or ablaut) is often based on alternations in a single distinctive feature (front/back in the case of Germanic umlaut plurals and Indo-European ablaut, high/low—/u/, /i/ contra /a/-- in the case of Arabic and Semitic stem-vowel alternations indicating tense or aspect). Segeral (1998:29) proposes that universally "only a very small subset of logically possible combinations of two vowels is predicted to be able to contract an apopohonic relation." I am skeptical of the details of Segeral's proposal because it relies on a non-optimal analysis of the alternations in Arabic, but the central idea that there could be cross-language and cross-family universals in apophony is promising.

These two features of non-concatenative systems can be analyzed very elegantly in the schema proposed by Ford & Singh (2003:19)[2], wherein any morphological relationship is described by a "Word Formation Strategy" having the following form:

[X]α↔[X']β where

(i) X and X' are words;
(ii) α and β are morphological categories;
(iii) ↔ indicates an equivalence relation (a bi-directional implication);
(iv) X' is a semantic function of X;
(v) ' indicates a formal difference between the two elements of the relation of the morphological operation;
(vi) ' can be null if α ≠ β

This is intended for and can be used to capture any morphological relationship. But it seems ideal for describing certain kinds of non-concatenative relationships, particularly those where a minimum phonological difference correlates with an abstract conceptual contrast like number or aspect. For example the morphological relationship between *goose* /gu:s/ and *geese* /gi:s/ could be represented as follows:

[Cu:C]sg. ↔ [Ci:C]pl.

The relationship between /dog/ and /dogz/ can also be represented in this framework.

[X]sg. ↔ [Xz]pl.

But the first case allows a further reduction into distinctive features:

[Cv:C]-pl. ↔ [Cv:C]+pl.
-front +front

Since the consonants and syllable structure are constant between singular and plural, this can be even further reduced to:

[X-front]-pl. ↔ [X+front]+pl

The point here is that the Ford/Singh schema represents one kind of ideal, which comes closest to being actualized in some kinds of non-concatenative morphology. Oppositions like *goose/geese* represent a highly economical and iconic way to relate form and function. Grammatical categories like number and aspect which are based on a basic, usually binary, conceptual contrast are easily characterized in terms of features (technically "distinguished" rather than distinctive features). It is therefore most "natural" (transparent and economical) to express these categories by minimal phonological differences which can also be characterized in terms of (distinctive) features such as difference in the quality or quantity of a vowel in otherwise identical stems. Another example of this kind of iconicity in non-concatenative morphology is provided by oppositions like Arabic *kataba* ("write," two arguments) vs. *kattaba* ("cause to write," three arguments), *kaataba* ("write something to someone," three arguments), where a minimal contrast in phonemic quantity correlates with a minimal contrast in valence (Ratcliffe 2005): The addition of a mora equals the addition of an argument, and the direction of spread correlates with the grammatical role of the new argument. (Leftward spread, gemination, means new argument in subject role. Rightward spread, vowel lengthening means new argument in object role.)

[X]val ↔ [X+1μ]val+1arg

Thus minimally contrastive non-concatenative morphology is one ideal toward which morphological systems tend to evolve. Yet this ideal is never reached, for two reasons. First such morphology tends to have inherent structural limitations which hinder or prohibit its productive extension. For example, a singular-plural opposition expressed through a contrast back-front on the unique vowel of a CvC word is transparent and easily extendable by analogy to other CvC singulars with back vowel (hence *moose/?meese* is at least imaginable). But what are speakers to do when confronted with a singular stem which has a non-back vowel or a word with more than one vowel? Concatenation of an affix may be a less economical mode

[2] Also earlier in Ford, Singh & Martohardjono (1997).

of expression, but in general does not encounter productivity restrictions of this type. (For this reason concatenative modes of expression are often available as a default even in highly non-concatenative systems—such as the Arabic noun plural (Ratcliffe 1998). The second reason is that as speakers mature and they become able to control the less economical but more powerful (productive) combinatorial morphological mechanism, there is no longer any need to try to extend the non-concatenative contrast. Thus concatenation of transparent morphemes according to syntactic principles is another ideal toward which morphological systems tend to evolve. The competition between these two contradictory tendencies is the cause of the complex and "messy" character of so many natural language morphologies.

iii. Phonological underspecification/ minimality of affixes.

One of the principal differences between words and (bound) morphemes is that the latter are often phonologically underspecified, with their phonological form partially determined by the base to which they attach. For example, in Turkish, vowels in suffixes are only specified for the feature height, not for backness or roundness. The latter features spread from the last vowel of the stem—a phenomenon known as "vowel harmony": *at* "horse" *atlar* "horses," *ev* "house" *evler* "houses." In English the productive inflectional suffixes indicating past tense (alveolar stop) and noun plural/ 3sg pres. (alveolar sibilant) can be analyzed as segments specified only for place and manner features and unspecified for voice, with the voice feature spreading from the preceding segment (the last stem segment or the epenthetic vowel that intervenes when the stem ends in a consonant at the same point of articulation as the suffix). Underspecified affixes represent a stage intermediary between the ideal of combinatorial morphology, in which morphemes, being word-like things, should be fully specified in phonological form, and the ideal of non-concatenative morphology in which a morphological contrast is expressed through a minimal phonological difference between words.

Patterns like these are often treated as morpho-phonology—the result of a phonological rule that applies only in morphologically defined environments. (Vowel harmony does not apply inside stems in Turkish: /ateʃ/ "fire" /asker/ "soldier." Voicing of sibilants after voiced segments in not a general phonological rule of English: *false* /fals/, *farce* /fars/ vs. *falls* /falz/, *bars* /barz/.) Morpho-phonology is an inherently awkward problem for combinatorial morphology. There is no problem if generally applicable phonological rules disrupt the underlying phonological integrity of morphemes, but why should a language have or preserve rules whose sole purpose is to distort the underlying phonological integrity of certain classes of morphemes? Is it possible that all morpho-phonology can be analyzed as phonological underspecification—and hence as a step in the process toward paradigmacity? I am not yet ready to venture such a proposal, given the wide variety of phenomena that have been included under the umbrella of morpho-phonology. But at the very least morpho-phonology represents a trend toward lexicalization as opposed to syntagmization in morphology.

Optimality Theory attempts to explain phonological alternations essentially as the result of competition between ease of articulation and maintaining the integrity of the signal. Assimilation, such as is seen in the Turkish and English cases above, can be interpreted as due to ease of articulation. That it is blocked in the case of stems and applies in the case of inflectional suffixes, shows that the integrity of the stem is more highly ranked than the integrity of the inflectional suffix in these languages. This is explained under the hypothesis offered here as follows: the paradigmatic contrast between words (including that between singular and plural) is significant for learners and speakers of all ages, but can be maintained with a minimum of phonological marking. The integrity of affixes marking plural (and certain other classes) as independent syntagmatic units is not significant for very young learners and hence is relatively less important.

iv. Syncretism. It is very often the case that languages use phonemically identical affixes to mark different categories, even opposite categories (polarity). One obvious example is the English /-z,-s/ as plural for nouns but third person singular for present tense verbs. In Arabic /t/ occurs in the verbal conjugation as an indicator of second person (as opposed to first) but also as an indicator of feminine (as opposed to masculine) in the third person. In verbal derivation it

indicates reduction of valence (Ratcliffe 2005). In the nominal morphology /t/ indicates feminine (as opposed to masculine), and also singulative (as opposed to collective), but occasionally (in conjunction with stem change) also plural (ʔax-"brother" ʔixwat- brothers"). Affixes like these function less like words, which should have direct external reference and in which the one form/one function principle is the norm, and more like distinctive features which only serve to mark two or more items in a system as different.

It is notable too that syncretism does not result from any shortage of the segments available to a language. Languages tend to use only a subset of their segment inventory in affixes, but to use them nonetheless for multiple functions. In English inflectional morphology only the alveolar stops and sibilants are used. In Semitic, of approximately 30 reconstructable consonants only the following seven are used in the affixal morphology of the Classical Semitic languages (Akkadian, Aramaic, Hebrew, Classical Arabic, and Ge'ez): ʔ,h,m,n,t,ʃ,y (Fox 2003:11, Anttilla 1989:196). The more "marked" segments-- pharyngeal fricatives, glottalized/ pharyngealized obstruents, and fricative laterals are not used as affixes. This is paralleled in phonology where not all phonetic features present in a language are used to make phonemic distinctions (e.g. aspiration in English).

Connected with these latter two points is the diachronic trend observed by Hopper & Traugott (1993:145): "In the process of grammatical attrition and selection that accompanies morphologization, we can identify two tendencies:

(a) A quantitative ('syntagmatic') reduction: forms become shorter as the phonemes that comprise them erode.

(b) A qualitative ('paradigmatic') reduction: the remaining phonological segments in the form are drawn from a progressively shrinking set. This smaller set of phonemes tends to reflect the universal set of unmarked segments. They tend especially to be apical (tongue tip) consonants such as [n],[t],and [s], the glottal consonants [ʔ] and [h], and common vowels such as [a],[u],[i], and [ə]."

The logical endpoint of quantitative reduction is underspecification—string of segments >> single segment >> single segment not specified in some features >> single feature. The logical endpoint of qualitative reduction is syncretism—fewer segments means the ones that remain must increasingly be used for many functions. The long term trend described here can be explained in terms of reanalysis and reshaping of morphological systems in terms of minimally contrastive paradigmatic alternations.

5. Pidgins and creoles

A Pidgin is defined as "a marginal language which arises to fulfill certain restricted communicative needs among people who have no common language" (Romaine 1988: 24, citing Todd 1974). Creoles are traditionally defined as pidgins which have become native languages, though specialists in pidgin and creole studies seem to moving toward the view that there is no sharp categorical difference between them (Lefebvre, White & Jourdan 2006). Creoles and pidgins typically have distinctive structural characteristics which distinguish them from normally transmitted languages. The most salient of these is the severe reduction of the morphological resources of the lexifier language. Non-concatenative morphology is virtually never transferred from lexifier to pidgin and even ordinary concatenative marking of inflectional categories is severely reduced. McWhorter (1988) identifies the following features as characteristic of the creole type: extreme reduction of inflectional morphology, absence of allomorphy, transparency of derivational morphology, and absence of tone as a marker of lexical contrasts or morpho-syntactic categories such as tense. All of these relate specifically to the reduction of morphology, though only one specifically relates to the absence of non-concatenative morphology (tone as tense marker). It has also been observed that the vowel systems of pidgins and creoles are often simpler than that of the lexifier, and the lexicon is generally much smaller. Pidgins and creoles typically make greater use of syntactic structures to make distinctions which are made in either the lexicon or the morphology of the lexifier. Thus Tok Pisin reverts to the lexical compounds *gras bilong hed*, *gras bilong pisin*, to express

meanings which in English are carried by distinct lexical items "hair" and "feather," respectively (Romaine 1988:35).

Since pidgins/creoles are typified both by how they are transmitted, and their structural features, it would seem reasonable that there is some connection between the two. A well-known attempt to establish such a connection is Bickerton's bioprogram hypothesis (1981, 1990). Bickerton proposes that creoles have the features that they do because they are created de novo by children using only innate resources with limited help from the "degenerate" input provided by the pidgin. This hypothesis, it seems to me, locates the source of pidgin/creole distinctiveness in the wrong place. Creoles are distinct from normally transmitted languages primarily because they share the structural properties of the pidgin from which they are derived. There is an abrupt transition from the lexifier to the pidgin. The transition from the pidgin to the creole is more subtle. Scholars of creole genesis have cast doubt on whether a clear distinction between pidgin and creole phase can be discerned at all in many cases (McWhorter 1998, Lefebvre 1998, Smith 2006). Smith (2006), for example, argues that Suriname creole was formed in its essential features in a very rapid period long before substantial numbers of children could have learned it as a first language. It also seems to me that Bickerton's theory overlooks what is clearly the most robust and uncontroversially typical feature of pidgins and creoles—namely morphological impoverishment.

My proposal is this: Creoles have the features that they do because they are derived from pidgins. Pidgins have the features that they do because they are created by adults. The structure of pidgins and creoles does not reflect some pure form of the initial state of language competence, rather it reflects the mature, adult state of competence. For an adult it is easiest to memorize a small set of symbols and string them together in complicated ways to expand the expressive repertoire. For a child it is easier to memorize many individual items, especially where such items differ in systematic ways in both form and meaning. Pidgins and creoles are not primitive, or childlike systems of communication. They rely upon cognitive processing mechanisms which are both ontogenetically and phyolgenetically advanced—that is, found only in human beings who have matured past infancy. They only appear simple to adult linguists because we too have developed the cognitive skills necessary to process them. Let me refine this position a bit in order to avoid unnecessary controversy. It is quite possible that the common structural features of pidgins and creoles reflect universal human cognitive abilities which are biologically determined. I am not arguing for or against this position here. I am arguing that pidgins and creoles provide no convincing evidence that these cognitive abilities are present at birth. My hypothesis is also not incompatible with Bickerton's proposal that certain syntacto-semantic distinctions (e.g. punctual/non-punctual) encoded in creoles reflect innate cognitive distinctions present at birth. (As Romaine 1988:260 points out, Bickerton's position differs interestingly from Chomsky's in that it emphasizes the innateness of semantic distinctions rather than formal structures.) My position is that the ability to encode such distinctions *syntactically* is not present at birth. A nice example is aspect. The marking of aspect non-concatenatively is, arguably, the most economical and iconic way to do so, as I have suggested above. Aspect is the inflectional verbal category most likely to be marked by stem change in normally transmitted languages (Bybee 1985). Aspectual distinctions are also acquired early in first language acquisition (Clark 1998:382). In pidgins and creoles aspect is typically marked exclusively syntactically—through a separable particle, usually derived from an auxiliary verb or adverb of the lexifier. This distinction nicely illustrates the hypothesized difference between the adult language faculty, in which syntax is the principal encoding device, and the early stage of the language faculty in which paradigmatic modes of encoding mature first and are gradually superseded.

6. Ontogeny and phylogeny

I speculate that non-concatenative morphology, specifically minimal contrasts between signs based on features such as tone, pitch, duration, loudness (stress), or the formant frequencies of vowels, represents a residue of pre-syntax phase of hominid communication surviving in human language. Let me be clear: it is not the particular contrasts themselves as found in particular languages (e.g. *foot/feet*) which I regard as evolutionarily primitive. Where records exist we can observe things like umlaut or tonogenesis emerging in relatively shallow historical (as opposed

to evolutionary) time. Rather it is the cognitive and neurological faculties which allow for processing of non-concatenative morphology that I propose to be evolutionarily primitive and prior to the faculties necessary for processing syntax. Such morphology is constantly emerging in normally transmitted languages, because the newborn infant has or almost immediately develops the ability to learn and process a primitive communication system based on a small repertoire of minimally contrasting signs, but only later develops the cognitive abilities necessary to process adult-like syntax.

This speculation is based on comparison with other animal communication systems and a kind of ontogeny recapitulates phylogeny argument from first language acquisition. In general, animal communication is restricted to individual signals, without anything like syntactic combination of signals. Most animal communication systems have a very small repertoire of signals. These make use of articulatory distinctions which are severely limited by the articulatory apparatus of the animal in question. Nonetheless the acoustic and auditory mechanisms used by other mammals to distinguish the different signals are similar to those used in human speech: "Acoustic analyses of chimpanzee calls show that they make use of phonetic 'features' that play a linguistic role in human speech" (Lieberman 2003:8, 1968) Differences in intensity of signaling (qualitative differences) and repetition (quantitative differences) are exploited even by bees and birds, respectively (Pinker 1994:334). The remote ancestors of human beings must have evolved brains which were prepared to learn a communication system of this kind. At some point a different sort of cognitive faculty, one which allows the processing of syntax, would have evolved. By hypothesis the evolutionarily older type of cognitive faculty is present at birth, the second type develops later.

Carstairs-McCarthy (2005:181) has reached a rather similar conclusion: "non-affixal (non-concatenative or modulatory) morphology preceded affixal (or additive) morphology in language evolution." Carstairs-McCarthy argues that (non-concatenative) morphology first arises from morpho-phonology, which in turn arises from the phonetic effect of contiguous words on each other in a syntactic string. And he argues much as I have above, that the development of morphology out of free words in a diachronic change is not merely a shift from phrase-syntax to word-syntax, but a "shift from a domain of grammar in which the syntagmatic dimension is dominant to one in which the paradigmatic dimension has at least equal importance" (2005:183). I would only add to this that for such a shift to occur there must be a predisposition on the part of some speakers (by hypothesis very young ones) to analyze (originally syntagmatic) linguistic structures paradigmatically. Carstairs-McCarthy also draws attention to the fact that it is inefficient for languages to have two distinct ways to organize grammatical information. The less efficient (or at least less productive) paradigmatic mode must have evolved earlier, because if the more productive syntagmatic mode had already been available there would have been no evolutionary advantage favoring the development of the paradigmatic mode. (The analogy I would draw here is with anaerobic respiration (fermentation) which is still available to eukaryotic cells, even though it has been superseded by aerobic respiration, which is much more efficient.) However, paradigmatic modes of organization are independently needed for organizing phonological inventories on the basis of distinctive features.

Most of the speculation on proto-language (Comrie 1992, Bickerton 1981, 1990) assumes that it started with strings of words with no morphology. But my question is what form would the words have taken? Would they have been systematically related to each other? Would the lexicon have been structured? I would suggest that the answers to these questions are probably positive and that therefore morphology, specifically non-concatenative morphology would have preceded syntax. Although it seems somewhat absurd to speculate in detail about what this system of communication would have been like, for the sake of exposition, I imagine that a simple phonological sequence perhaps a CV syllable like *pa* or *ka* would have been a signal of some kind and that a change in the signal such as lengthening *paa*, *kaa* or raising the tone or pitch *pá*, *ká* would have indicated an intensification of the signal. The next evolutionary step would have been associating the signal to a specific referent, for example an animal of some kind, with the intensified signal then used to indicate many of the animals or a particularly large

version of the animal, or some other minimally semantically different aspect of the same referent

7. Conclusion

This paper has tried to pull together results from several sub-fields of linguistics, in most of which I can claim no special expertise. The solution I propose may not be the right one. But I hope to have shown that the problem is a real one. If languages structured only according to syntactic principles with no morphology are intuitively natural, then why are virtually all naturally transmitted languages, learned by children, "unnatural"? Why are the most "natural" languages in fact artificial languages and pidgins, created by adults? Why do languages tend to become more "unnatural" over time?

What linguists do, have always done, is to try to uncover underlying order in the superficial diversity and disorder of language data. But in different eras different claims have been made about the ontological status of the underlying patterns so discovered—historically older vs. psychologically real. In essence these two types of claims are mutually contradictory. It is reasonable to equate earlier in a developmental sense with earlier in an evolutionary sense. (This is the logic of ontogeny recapitulates phylogeny). But it is, I think fundamentally wrong to equate earlier in a historical sense with either of these two. Indeed the opposite equation is what we should expect: historically late = developmentally early. If children are responsible for recreating language, then their new creations must reflect neurological functions which are developmentally early, and possibly also older in an evolutionary sense. There is no reason to think that children learn complex alternations by internalizing the whole history which has led to them. Rather they are themselves are the creators of this history and they must do it by restructuring these alternations through the use of cognitive faculties which are different from those employed by adults.

When linguists interpret non-concatenative morphology in combinatorial terms they are not observing or describing the language faculty (much less a primitive, initial state of it). Rather they exercising the faculty as it exists in adults. The fact that linguists can reanalyze non-concatenative contrasts in syntagmatic terms (Lieber 1992 or Stonham 1994) or, alternatively analyze transparent head-modifier compounds in terms of static paradigmatic contrasts (Anderson 1992 or Ford, Singh & Marohardjono 1997), does not constitute a demonstration there is only one mode of morphological organization. Rather it shows that the language faculty consists not only of two different modes of morphological organization (paradigmatic and syntagmatic), but also an ability to translate between them. This too is clear from the possibility of translation between languages with different morphological profiles and from patterns of reanalysis in diachronic change.

References

Andersen, Henning. 1973. "Abductive and Deductive Change," *Language* 49.765-93.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Anttila, Raimo. 1989. *Historical and Comparative Linguistics*. Amsterdam: John Benjamins.

Bickerton, Derek. 1981. *Roots of Language*. Ann Arbor, Mi.: Karoma.

Bickerton, Derek. 1990. *Language and Species*. Chicago: University of Chicago Press.

Bowerman, M. 1982. "Starting to talk worse: Clues to language acquisition from children's late errors." in *U-shaped behavorial growth*, S. Strauss, ed.. New York: Academic Press.

Brown, R. 1973. *A First Langauge*. Cambridge, Ma.: Harvard Univ. Press.

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Carstairs-McCarthy, Andrew. 2005. "The evolutionary origin of morphology" in *Language Origins: Perspectives on Evolution*, Maggie Tallerman, ed., pp.166-184. Oxford: Oxford University Press.

Chomsky, Noam. 1993. "A Minimalist Program for Linguistic Theory." in The *View from Building 20*, Kenneth Hale and Samuel Jay Keyser, eds. Cambridge, Ma.: MIT Press.

Clark, Eve V. 1998. "Morphology in Language Acquisition" in *The Handbook of Morphology*, Andrew Spencer and Arnold M. Zwicky, eds., pp. 374-390.Oxford: Blackwell.

Comrie, Bernard. 1992. "Before Complexity." in *The Evolution of Human Language,* J.A. Hawkins and M. Gell-Mann, eds, pp. 193-211. Santa-Fe, NM: Addison-Wesley.

Ford, Alan R., Rajendra Singh & Gita Marohardjono. 1997. *Pace Panini: Towards a Word-based Theory of Morphology*. New York: Peter Lang.

Ford, Alan and Rajendra Singh. 2003. "Prolegomena to a Theory of Non-Paninian Morphology" in *Explorations in Seamless Morphlogy*, pp. 18-42. Rajendra Singh and Stanley Starosta, eds. New Delhi: Sage Publications.

Fox, Joshua. 2003. *Semitic Noun Patterns*. Winona Lake, Ind.: Eisenbrauns.

Greenberg, Joseph H. 1955. "Internal a-plurals in Afroasiatic." in *Afrikanistische Studien Diedrich Westermann zum 80. Geburtstag gewidmet*. ed. by Johannes Lukas, 198-204. Berlin: Akademie Verlag.

Hock, Hans H. 1991. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.

Hopper, Paul J. & Elizabeth C. Traugott. 1993. *Grammaticalization*.Cambridge: Cambridge University Press.

Ingram, David. 1989. *First Language Acquisition*: Method, Description, Explanation. Cambridge: Cambridge University Press.

Jusczyk, Peter W. 2003. "The Role of Speech Perception Capacities in Early Language Acquisition. in *Mind, Brain, and Language: Multidisciplinary Perspectives*, Marie T. Banich and Molly Mack, eds., 61-86. Mahwah, New Jersey: Laurence Erlbaum.

Kiparsky, Paul. 1968. "Linguistic Universals and Linguistic Change," in *Universals in Linguistic Theory,* ed. Emmon Bach and Robert T. Harms. New York: Holt, Rinehart and Winston, Inc.

Lass, Roger. 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.

Lefebvre, Claire. 1998. *Creole Genesis and the Acquisition of Grammar: The Case of Haitian Creole*. Cambridge: Cambridge U. Press.

Lieber, Rochelle. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory*. Chicago: University of Chicago Press.

Liebermann, Philip.1968."Primate Vocalizations and Human Linguistic Ability," *Journal of the Acoustic Society of America* 44, 1157-1164.

Liebermann, Philip. 2003. "Language Evolution and Innateness," in in *Mind, Brain, and Language: Multidisciplinary Perspectives*, Marie T. Banich and Molly Mack. eds., pp.3-22. Mahwah, New Jersey: Laurence Erlbaum.

McCarthy, John & Alan Prince. 1998. "Prosodic Morphology" in *The Handbook of Morphology*, Andrew Spencer and Arnold M. Zwicky, eds., pp. 283-305.Oxford: Blackwell.

McWhorter, John H. 1997. *Towards a New Model of Creole Genesis* . New York. Peter Lang.

McWhorter, John H. 1998. "Identifying the Creole Prototype: Vindicating a Typological Class." *Language* 74/4, pp. 788-818.

Mifsud, Manwel. 1994. "Internal Pluralization in Maltese: Continuity and Innovation." in *Actes des premières journées internationales de dialectologie arabe de Paris*, Dominique Caubet and Marine Vanhove, eds., pp. 90-105. Paris. Inalco.

Mous, Maarten. ms. "Cushitic." University of Leiden, Netherlands.

Newman, Paul. 1990. *Nominal and Verbal Plurality in Chadic*. Dordrecht: Foris Publications.

Nice, M. 1917. "The Speech Development of a Child from Eighteen Months to Six Years." *Pedagogical Seminary* 24:204-43.

Paul, Hermann. 1880. *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemeyer.

Piaget, Jean. 1929. *The Child's Conception of the World.* New York: Harcourt Brace.

Piaget, Jean. 1948. *Play, Dreams, and Imitation in Childhood*. New York: Norton.

Pinker, Steven. 1994. *The Language Instinct*. New York: Harper Collins.

Ratcliffe, Robert R. 1992. *The 'Broken' Plural Problem in Arabic, Semitic, and Afroasiatic*. PhD. Dissertation. Yale University.

Ratcliffe, Robert. 1996. "Drift and Noun Plural Reduplication in Afroasiatic" *Bulletin of the School of Oriental and African Studies*. 59:2.296-311.

Ratcliffe, Robert. 1998. *The 'Broken Plural' Problem in Arabic and Compartive Semitic*. Amsterdam: John Benjamins.

Ratcliffe, Robert. 2005. "Semi-Productivity and Valence Marking in Arabic—the so-called 'verbal themes' in *Corpus-Based Approaches to Sentence Structures*, Toshiro Takagaki, Susumu Zaima, Yoichiro Tsuruga, Francisco Moreno-Fernandez, and Yuji Kawaguchi, eds. Amsterdam: John Benjamins.

Romaine, Suzanne. 1988. *Pidgin & Creole Langauges*. London: Longman.

Saeed, John Ibrahim. 1999. *Somali.* (London Oriental and African Language Library, 10.) Amsterdam: John Benjamins.

Segeral, Philipe. 1998. "A Generalized Theory of Ablaut: the Case of Modern German Strong Verbs," in *Models of Inflection*, Ray Fabir, Albert Ortmann, and Teresa Parodi, eds., pp. 28-59. Tübingen: Max Niemeyer Verlag.

Smith, N. and J. Venstra, eds. *Creolization and Contact*. Amsterdam: John Benjamins.

Stern, C. & W. Stern. 1907. *Die Kindersprache*. Leipzig: Barth.

Stonham, John T. 1994. *Combinatorial Morphology*. Amsterdam: John Benjamins

Trask, R.L. 1996. *Historical Linguistics*. London: Arnold.

Zaborski, Andrzej. 1986. *The Morphology of Nominal Plural in the Cushitic Languages*. Vienna: Afro-Pub.

Exaptation from Arabic syntax to Persian lexical Morphology

John R. Perry
The University of Chicago
j-perry@uchicago.edu

It has long been natural for linguists to invoke metaphors taken from the life sciences, in particular evolutionary biology. In recent decades it has been argued that such analogies represent not mere rhetoric or metalinguistic convenience, but actual processes in language development; after all, why should not a biologically-conditioned cultural phenomenon such as language, the collective product of a life form, replicate aspects of a form of life – and not just general processes such as evolution and extinction, but actual details of the process, such as natural selection and exaptation?

"Exaptation" is a biological term coined by Stephen J. Gould and Elisabeth Vrba in a 1982 article[1]. It may be defined as "a case where an anatomical structure that originally evolved to serve one function was later commandeered to facilitate a quite different function." For instance, feathers in the proto-avian dinosaur lineage demonstrably evolved before the capacity for flight; their purpose must have been something else, such as thermo-regulation (to keep them warm), or display (to attract a mate) – functions which additionally continued in many cases. However, this structure was fortuitously available when it was later modified to provide flight-control surfaces. Darwin anticipated this process in 1859, applying the term "pre-adaptation," and citing the example of a fish's swim-bladder, as having originally evolved for flotation, and in land animals being converted to a wholly different purpose – that of respiration, in the form of lungs[2].

In linguistics the term has been adopted by, e.g., Roger Lass in a 1990 article[3], and by Laura A. Janda in *Back from the brink: a study of how relic forms in languages serve as source material for analogical extension* (Lincom Europa, 1996). My example of the redeployment of a distinctive feature to a novel purpose (which I investigated in the1980s, before I had heard of exaptation) involves the successful hijacking, by semantic determinants of Persian lexical morphology, of a syntactically-conditioned phonological alternation in the feminine ending of Arabic nominals. The human motivation was the desire, during the development of the Eastern Islamic Kulturgebiet between the seventh and twelfth centuries (see the map, fig. 4) to incorporate useful or prestigious Arabic vocabulary into Persian, using the Arabic writing system (which had been adopted in Persian) but adhering to Persian phonotactics and lexical morphology. The junk element in Arabic feminine-ending nouns and adjectives was not so much that Persian had no grammatical gender, but that it had no use for the typically Semitic syntactic structure known as the "construct state": this requires that a feminine ending be pronounced as /-at/ with terminal *-t* when its nominal is the head of a NP modified by a following noun ("pre-juncture position"), and /-a/ in all other situations ("pre-pausal position"). Thus *dawl**at** al-Sūdān* 'the state of Sudan', but *ra'īs al-dawl**a*** 'head of state' and *dawla mustaqill**a*** 'an independent state'. In each case, the feminine marker was written with an invariant hybrid graph in Arabic. Now, Persian speakers needed definitively to lexicalize a single form of a word as either *-at* or *-a*.

The solution they devised has led to an inventory of at least 1400 Arabic Feminine Ending (AFE) loanwords in the modern Persian lexicon, in a ratio of roughly 600 *-at*: 800 *-eh*, including forty doublets – i.e., copies of the same word in each form, written with distinctive graphs (final *t*, and final *h* for spoken /-a/) and two lexically distinct meanings (c. fig. 3; I use *-eh* to represent the vocalic termination as being visually quite distinct from *-at*, and to

[1] "Exapttion – a missing term in the science of form," *Paleobiology* 8 (1982): 4-15.
[2] See *The Origin of Species* Chapter VI, under "Modes of Transition."
[3] "How to do things with junk: exaptation in language evolution," *Journal of Linguistics* 26 (1990): 79-102.

incorporate the incidental vowel change that marks this form in Standard Persian). A simple example, involving a doublet, is: *āyat* 'sign, portent' vs. *āyeh* 'verse of scripture'.

Looking for a rationale behind the distribution of these allomorphs, it struck me how utterly divorced were the current Persian function of the *-at* vs *–eh* dichotomy (as lexical markers, the significance of which was yet to be described) and the Arabic syntactic alternation *-at* ~ *-a*. The anatomical location, so to speak, was identical, but the process of recycling and the new rationale had remained opaque. The forms are a salient feature of Persian's large Arabic loanword inventory, but one that had so far been taken for granted. Neither Iranian nor foreign scholars (except for one of each, with very limited results) had even taken note of the puzzle; it was indeed the case (to paraphrase Gould & Vrba) that "Current functions cannot be used to infer past causal pressures." After collecting the two inventories, I was able first to eliminate any copying of Arabic syntagmata (such as nominals in the construct state, noun adjective NPs, prepositional phrases, etc.) as a source for the loanword patterns. Construct state NPs were quite commonly lifted into Persian *in toto*; but the pre-pausal feminine ending in, e.g., *fawq al-'āda* 'extraordinary', does not correspond to the canonical lexical form *'ādat* 'custom, norm' in Persian; nor does the pre-juncture form in the epithet *sayf al-dawla* 'Sword of the State' correspond to the lexicalized Persian *dawlat* 'state'.

Secondly, I eliminated the cognate loanwords in Malay-Indonesian, various African languages, and Spanish, and determined that this regular pattern of distribution between *-at* and *-eh* was exclusive to the Persianate world, i.e., to those languages which had received their Arabic lexical copies in pre-modern times through the medium of Persian – mainly Iranic, Turkic, and Indic languages covering the shaded area of the map (fig. 4).

Thirdly, I realized that the process had been a dynamic one, both diachronically and geographically. On the evidence of Persian literature of the 10th-11th centuries, the initial ratio of incorporation into Persian showed a preponderance of *-at* over *-eh* affiliates of approx 6:1, as against a modern preponderance of 4:3 *-eh* over *-at*. So, over the intervening 1000 years, there had evidently been a considerable shift in individual items from the *–at* inventory to *-eh*. From an areal perspective, the basic Persian pattern of distribution, and the rate of shift, were not 100 percent the same in inventories of cognate copies in other languages of the area. Turkic dialects especially. showed an idiosyncratic dynamic, with instances of shift and doublets clearly based on the same rationale as in Persian, but targeting different words. Indic languages also demonstrated some independence, but in general were more conservative than Turkish and Persian, retaining more original *-at* affiliates.

Finally, what is the rationale for the dichotomy in these two series and their patterning? We may gain an overview of it by shifting our attention between fig. 1, which summarizes the principal linguistic and sociolinguistic features of the system, and fig. 2, where these are exemplified by the location of typical AFE loanwords in the Persian semantic and lexical spectrum. This "semantic spectrogram" maps words across a range of lexical, syntactic, and sociolinguistic usage, from most abstract and least specialized to most concrete and/or specialized. The illustrations in fig. 2 are singletons, i.e., borrowings which remained true to their respective incorporation in *-at* or *-eh* without formal change.

In the case of loanwords which in Arabic were deverbal or deadjectival derivatives, "specialization" may involve re-verbalization as a compound verb (+V) and/or acquisition (whether contextual or permanent) of a particular extension of the verbal sense, such as a noun of instance (English *sleeping, sleep* are action nouns, *[a] sleep* or *nap*, pl. *naps* are instance nouns), a product noun (as Eng. *[a] collection*) or instrumentive or agentive nouns (as *cooker, cook*). Thus the same nominal may occupy progressively more tangible, imageable, and countable semantic slots so as to approach the concrete polarity of an entity noun. The hatched line separating the mean point between clusters of *-at* and *-eh* affiliates in these figures I call the "Semantic Watershed." It illustrates the comparative density and location of *-at* and *-eh* copies across the semantic spectrum.

Bearing in mind the features sketched in the three sections of fig. 1, what stands out intuitively from the meaning of the forms listed in fig. 2 is the following.

(1) From a semantic and syntactic perspective, *–at* marks words for abstract, intangible, low-imagery referents. Here the *-a*t affiliates constitute mainly quality and action nouns, as nos. 5 *ṣo'ûbat* 'difficulty', 10 *ḥokumat* 'government'. Some may have expanded their semantic range for use as nouns of instance (fourth column: 'a [particular instance of] difficulty, a problem'); or, more rarely, as product, agentive, or instrumentive nouns ('agency that governs; the persons so constituted'). These are then count nouns: 'difficulties, governments'. Though they are more 'imageable', i.e., readily visualized via a speaker's or hearer's particular experiences, they seldom extend as far as tangibilia, or entity nouns (last three columns).

Conversely*, -eh* tends to mark concrete, tangible, high-imagery deverbals such as product nouns, etc.: 1 *moqaddemeh* 'preface, introduction [to a book, etc.]' (instrumentive), 4 *mervaḥeh* 'instrument for creating a draft', 5 *raqqāṣeh* 'dancing girl' (agentive). In everyday Persian usage, the deverbal function encoded in the Arabic morphological pattern may be opaque, and nouns such as this last will fall intuitively into the class "entity noun," the same as nos. 8 'town, city' (and the name of a particular city), 9 'name of a female', 10 'type of camel'.

(2) Sociolinguistically, *-at* marks unspecialized, and *-eh* specialized terms in various ways. Stylistically, those originally incorporated in *-at* as learned words (*mots savants*) often remain in the higher, literary register; some may be characterized as "Classical," i.e., archaic or imperfectly assimilated in modern Persian (*Fremdwörter*). Those that were incorporated in the *–eh* series are (or were) common in the vernacular register (and tend, of course, to be countable and to correspond to concrete and entity nouns). Those which, we might guess, were originally transmitted orally turn up as *–eh* (such as names: no. 9 *Xadijeh*, and the toponym *Maymaneh*2 in fig. 3 no. 9; common entity nouns: no. 10 *jam(m)āzeh* 'dromedary').

(3) Questions of diachronic shift *-at* > *-eh,* the generation of doublets, and differential assimilation in different languages (as limned in the third section of this epitome), are best illustrated in condensed form via fig. 3 (Doublets). To begin with a minor lexical point that is not captured in fig. 1: contrastive affiliation can, at its simplest, disambiguate what were homonyms in Arabic (i.e., a mere coincidence of root and surface forms), such as fig. 3, no. 8 *šarārat* 'evil, wickedness' vs. *šarāreh* 'spark'. Homophones and homographs in Arabic, these non-cognate words find themselves, as Persian loanwords, appropriately differentiated in sound and form, and parked at opposite ends of the semantic spectrum in accordance with their affilation in *-at* or *-eh*. More subtly, no. 4 *mānaviyat* 'Manichaeism' and *Mānaviyeh* '(the) Manichaeans' (collectively, as a sect or community) demonstrates different uses of the feminine ending in Arabic as a quality noun ('Manichaean-ness') and as a collective noun formative; the resulting surface homonymy in Arabic is disambiguated in Persian by the assignment to, and marking of, separate semantic slots appropriate to *-at* and *-eh*.

No. 6 *mas'ala*, in Arabic, likewise represented not simply two different denotations of the same lexical pattern, but two semantically distinct lexical patterns. As discrete Persian copies, the Classical Persian (CP) action noun *mas'alat* 'asking, questioning' (a *mot savant*) soon took a back seat to the everyday instance and product noun *mas'aleh* 'question, problem, issue, matter; thingy' (a vernacular euphemism for the male sexual organ). Equally ingenious is the doublet pair no. 2 *ešāriyat/-eh*, both being late Arabicate neologisms in Indo-Persian or even Urdu. The abstract 'symbolism' contrasts with the instrumentive 'index' (a count noun, readily visualized, and even tangible when printed). These neologisms were derived from Perso-Arabic doublet pair no. 1 *ešārat/-eh* 'showing; indication, gesture, sign'. The semantic extension in the direction of specialization, imageability, tangibility and vernacularity that prompted the shift *-at* > *-eh* is evident: Classical Persian *ešārat* is a verbal abstract, 'the showing, demonstration'; as +V it is re-verbalized in conjunction with an auxiliary ('to show, indicate, gesture'), and soon shifts in this function to *ešāreh*; but it is so marked also as a count noun of instance/product, 'pointer, gesture, sign'.

No. 3 *mo'ādelat* was similarly abstract, 'equivalence, balancing' when first copied into Persian. In Ottoman Turkish it shifted to *mo'ādeleh*, in the product noun sense 'equation'. Likewise restricted to Turkish is the pair no.5 *ḥarakat* 'movement' (an action

noun) and its shifted doublet ḥ*arekeh* ‘vowel sign’ (a grammatical device and the written symbol for this; modern orthography *hareket, hareke*; Persian uses *-at*).
No. 7 *ta‘ziat* was copied into Classical Persian with the meaning ‘condolence, mourning’. In Turkish this shifted to *-eh*, in the more concrete and vernacular sense of ‘obsequies, funeral’; in the usage of Shi’i Turkmen and Iranians of the fifteenth century, it specialized as ‘Moharram mourning rites for Imam Hosayn’; returning to Persian as a *Rückwanderer*, the word was further restricted to the sense ‘Passion play’ (in Persian), and later copied into Indo-Persian with reference to a ritual object, the model of the martyr’s tomb carried in procession.

In no. 9 we encounter a neat tripartite distinction of three meanings of Arabic *maymana*, a homonymous doublet from the triliteral root *YMN* ‘right (side); being of good omen’. The word of the two that is an action noun (from a stative verb, hence in effect a noun of condition or quality, ‘wellbeing, prosperity’) is logically affiliated in Persian in the *-at* series. The other word, a locative ‘right wing (of an army)’, as a specialized count noun derivative of the plain, non-metaphorical sense of *YMN*, is allocated to the *-eh* series as *maymaneh*1. Then there is also a city in Afghanistan called Maymana (*maymaneh*2); is this in origin ‘located on the right (side, wing)’, or a ‘place of prosperity’? Whichever it be, the form is appropriate.

In the course of the next several centuries, hundreds of the *-at* class shifted to the *-eh* class, some leaving behind traces as doublets in *-at*. (Or, to rephrase this in more sociolinguistic terms, a shortened form with a more specialized, imageable connotation evolved in the vernacular, which soon complemented, or superseded, the literary word). In general, the sound (and orthographic) change recapitulates the original rationale: the resulting *–eh* words are semantically more specialized or concretized (cf. fig. 3 no. 3 *mo‘ādeleh* ‘equation’, or no. 5 ḥ*arekeh* ‘vowel sign’), and/or more firmly established in the vernacular (cf. no. 6 *mas’aleh* ‘question, matter, problem’, and no. 7 *ta‘zieh* ‘Moharram passion play; cenotaph’).

There are of course apparent exceptions to the trends observed here. *Nomina actionis* especially (by definition abstract, less imageable, in their basic meanings) seem to be more arbitrarily apportioned; even of these, however, those ending in *-eh* tend to form common compound verbs in Persian (i.e., are more imageable in context, and frequent or vernacular in usage), and have also evolved count-noun referents (no. 1 *ešāreh*). This shift appears to have peaked by the late thirteenth century, by which time the majority of the Arabic loanwords that are in general use today had been incorporated in Persian, and were being transmitted to Turkish, Hindi, etc. Moreover, as is evident from the examples, not only were individual loanwords incorporated into Persian then passed on to nearly all the other languages of mainland Muslim Asia, but the intuitive rules for this binary sorting were transmitted with them, to be used innovatively in the recipient languages.

To conclude: a fortuitous syntactically-triggered dichotomy in a portion of the inventory of Arabic substantives that were copied into Persian was reanalyzed to furnish a sub-system of semantic sorting and lexical expansion in the languages of a significant cultural area. This exaptive morphological redeployment of the Arabic feminine ending in Persian was still active in Persianate cultures at least until the first decades of the twentieth century: in evidence we may cite neologisms such as Urdu *ešāriyat* and *ešāriyeh*, Persian *e‘lāmiyeh* ‘manifesto’, Turkish *eṭfā’iyeh* ‘fire service’ and *melliyat* ‘nationalism’ (also adopted in Persian), and (Soviet) Tajik *partiaviyat* ‘party loyalty’. It has ceased to be productive, except to a limited extent in Urdu of Pakistan, since Arabic ceased to be an active source of vocabulary after the language reform movements in Turkey, Iran, and Soviet Muslim Asia during the second quarter of the twentieth century.

Further references to works by J. R. Perry on this topic

1984a “*-at* and *-a*: Arabic Loanwords with the Feminine Ending in Turkish.” *Turkish Studies Association Bulletin* 8/2: 16-25.

1984b “Differential Assimilation of some Arabic Loanwords in Tajik and Uzbek.” *Folia Slavica* Vol. 7, nos. 1-2: 268-82.

1984c "Semantic Sorting and Morphological Shift in a Loanword Category." *Journal of Historical Linguistics and Philology* (University of Maryland) Vol. I, No. 2: 31-53.

1989 "Lexical Doublets and Triplets in the Turkic Languages of the USSR," in Howard L. Aronson, ed., *The Non-Slavic Languages of the USSR: Linguistic Studies*. CLS, University of Chicago: 176-186.

1991 *Form and Meaning in Persian Vocabulary: The Arabic Feminine Ending.* Bibliotheca Persica Persian Studies Series No. 12. Mazda Publishers: Costa Mesa, CA. [254 pp.]

2004 "Lexical areas and semantic fields of Arabic loanwords in Persian and beyond," in Éva Á. Csató, Bo Isaksson and Carina Jahani (eds), *Linguistic convergence and areal diffusion: Case studies from Iranian,Semitic and Turki*c. RoutledgeCurzon: 97-110.

Fig. 1. Epitome of the *t.m.* Series Dichotomy

	-at	*-eh*
Semantic & Syntactic Features	Abstract; *and/or* intangible, insubstantial, low-imagery. Mass noun +VN (Nominalization of VP)	Concrete; *and/or* Tangible, substantial, high imagery. Count noun +N; +V (Reverbalization)
Sociolinguistic Features	Mot savant; *Fremdwort* High register; *and/or* Classical, literary; ?higher written frequency	Lexically assimilated Low register; *and/or* modern, vernacular; ?higher spoken frequency
Diachronic & Dialectal Features	Abstract, etc. Mass noun +VN Mot savant > *Fremdwort* High register, etc. *But also:* written > spoken	> Concrete, etc. > Count noun > +N; +V (reverbalization) > Lexically assimilated > Low register, etc. written < spoken

Fig. 2. Semantic Spectrogram: Single-Series Loans in Persian

	Qual. N.	Action Noun		Inst. N.	Product Noun, &c.		Entity Noun			
		+N	+V		Int.	Tan.	Mas	Cnt		
1. *oxovvat*	t				h	h		h	*moqaddemeh*	1.
2. *žarurat*	t				h	h		h	*majmu'eh*	2.
3. *morovvat*	t					h		h	*'amaleh*	3.
4. *nobovvat*	t					h		h	*mervaḥeh*	4.
5. *ṣo'ubat*	t			t		h		h	*raqqāṣeh*	5.
6. *xošunat*	t	t		t			h	h	*'olufeh*	6.
7. *kodurat*	t	t		t			h		*xārejeh*	7.
8. *sokunat*	t	t	t	t				h	*madineh*	8.
9. *'oqubat*		t		t				h	*xadijeh*	9.
10. *ḥokumat*		t	t	t	t	t		h	*jam(m)āzeh*	10.

GLOSSES: ***-at*** 1. 'brotherhood'; 2. 'necessity'; 3. 'manliness'; 4. 'prophethood'; 5. 'difficulty'; 6. 'harshness'; 7. 'turbidity'; 8. 'residence'; 9. 'punishment'; 10. 'government'.
-eh 1. 'preface'; 2. 'compendium'; 3. 'worker'; 4. 'fan'; 5. '(female) dancer'; 6. 'fodder'; 7. 'foreign parts'; 8. 'town; Medina'; 9. 'Khadija'; 10. 'dromedary'.

Fig. 3. Semantic Spectrogram: Sample Doublets in Persian, Turkish and Urdu (P, T, U)

	Qual. N.	Action Noun		Inst. N.	Product Noun, &c.		Entity Noun		
		+N	+V		Int.	Tan.	Mas	Cnt	
1. *ešārat*		t h	h	h	h	h			*ešāreh*
2. *ešāriyat* (U)	t				h	h			*ešāriyeh* (U)
3. *mo'ādelat* (T)	t	t			h				*mo'ādeleh* (T)
4. *mānaviyat*	t						h		*mānaviyeh*
5. *ḥarekat*		t	t	t h	h	h			*ḥarekeh* (T)
6. *mas'alat* (CP)		t	t		h			h	*mas'aleh* (P)
7. *ta'ziat* (P)		t	t		h			h	*ta'zieh* (T, P, U)
8. *šarārat*	t							h	*šarāreh*
9. *maymanat*	t	t						h	*maymaneh*$_1$
								h	*maymaneh*$_2$

GLOSSES: 1. 'indication'/'gesture; sign'; 2. 'symbolism'/'index'; 3. 'equivalence'/'equation'; 4. 'Manichæism'/'the Manichæans'; 5. 'motion, movement'/'vowel sign'; 6. 'asking'/'problem; thingy'; 7. 'condolence'/'obsequies' (T), 'religious drama' (P), 'cenotaph' (U); 8. 'wickedness'/'spark'; 9. 'prosperity'/'right wing; name of town'.

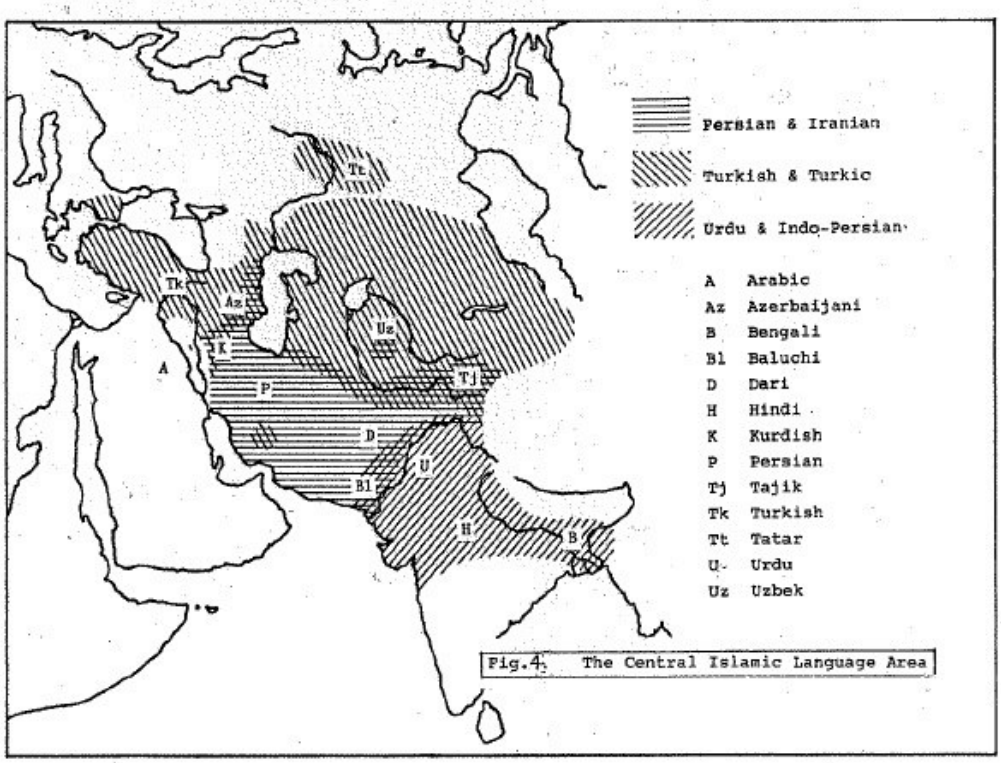


Fig.4. The Central Islamic Language Area

Analogy *vs* Rules: How Can Diachronic and Synchronic Perspectives be made to work together?

AMIOT Dany
Université d'Artois, JE 2489 Grammatica. F-Arras.
dany.amiot@univ-artois.fr

1. Introduction

The research I present in this paper is part of an on-going research project; consequently, some of the questions I raise I consider to be (relatively) solved, others not. More specifically, I will try to answer two questions:
– Firstly, how should we conceptualize and represent the evolution that leads from the use of a term as a preposition to its use as a prefix?
– And secondly, what theory is best fit to account both for this diachronic evolution and for the synchronic use of the term as a prefix?

Most researchers study either diachronic changes, either affixes from a synchronic morphological point of view; there are but a few researchers who combine both points of view. I decided to look into the work of Teresa Vallès (2004) and Geert Booij (2005 and 2008), because, although they are originally not into historical linguistics, they both work in theoretical frameworks that allow them to describe the emergence of new words or new patterns in language. Teresa Vallès does so in studying lexical creativity in the framework of the usage-based model of Joan Bybee, and Geert Booij, in studying, amongst others topics, some problematic cases between compounding and derivation in the framework of Adele Goldberg's Construction Grammar.

In this paper, I will first present some distinctions I make between the formatives of prepositional origin; then I will illustrate the problem I discuss by using the concrete example of *après*. In parts 3 and 4, the proposals of T. Vallès and G. Booij will be set out and "evaluated" in relation to the two questions raised above. In my conclusions, I will resume the results of this small investigation and show how they open some perspectives for future research.

2. Some distinctions

The questions I will raise turned up in previous diachronic and in synchronic research on French prepositions and formatives such as *sur* 'on', *après* 'after' or *sans* 'without'. These items present the particularity of having both prepositional and formative uses in contemporary French. This is shown in the examples under (1):

(1) *sur le sol* 'on the ground'
après les vacances 'after the holidays'
sans (son) parapluie 'without his /her umbrella'
surexposition 'overexposure'
après-guerre 'the post-war years'
sans-gêne 'lit without embarrassment, lack of consideration'

In fact, these data are well-known and comparable examples have already been observed in many languages.

In Amiot (2004), I claimed that these items are ambivalent: some, such as *sans-*, are still prepositions, others, such as *sur-*, have turned into full blown prefixes, and still others, such as *après-*, have some kind of intermediate status. The arguments I used to show this, were, briefly put, the following:
(i) a prefix has at least one meaning that is different from the corresponding preposition;
(ii) it is part of a paradigm that contains other items possessing beyond any doubt the status of prefix;

(iii) it can be associated to bases of different categories and be used to form lexemes that are members of different categories;
(iv) the complex lexeme is semantically endocentric;
(v) if the complex lexeme is a noun, it inherits its gender from the lexeme base.

This list of five arguments resumes the work of several researchers, in particular that of Scalise (1984) and Iacobini (1998, 2004); the arguments show that a prefix can originate from a corresponding preposition.

Sur- and *sans-* allow us to illustrate the differences that exist between French items with respect to their degree of prefixisation.

Concerning the prefix *sur-*:
(i) Its main meaning is 'excess', as can be seen in (2):

(2) *surcharge* 'overload', *surestimer* 'overestimate, over-value', *suralimentation* 'overfeeding'

However, this meaning cannot be expressed by the corresponding preposition.

(ii) *Sur-* is a member of the same paradigm as *hyper-*, which is a prefix without any doubt:

(3) *hypertension* 'hypertension', *hyperactif* 'hyperactive'

(iii) It can enter in processes that concern several distinct lexical categories:

(4) N → N: *charge* → *surcharge*
V → V: *estimer* → *surestimer*
A → A: *doué* → *surdoué* 'gifted, super talented'

(iv) The complexes with *sur-* are semantically endocentric, that is, they are more or less hyponyms of the base: a *surcharge* is a sort of charge (it's a too heavy charge), *surestimer* is estimate, but the estimation is too great, and so on.
(v) As a consequence of their endocentricity, the complexes are of the same gender as the base: *surcharge* is feminine just as *charge* is and *surpoids* 'excess weight' is masculine just as *poids* is.

Sans-, on the other hand, is really similar to the corresponding preposition *sans*:

(i) It has exactly the same meaning, they both express 'absence' ('without'):

(5) *sans-abri* 'litt: without-shelter, homeless' / *il est sans abri pour la nuit* 'he has no shelter / housing for the night"

(ii) In French, there is no paradigm it can be a member of;
(iii) It can only be involved in one categorical relation:

(6) N → N: *sans-abri, (faire un) sans-faute* 'lit without-mistake; to do something (in sport, at school for example) without a mistake'

(iv) The complex lexemes formed using *sans* are semantically exocentric: a *sans-abri* is not a sort of *abri* 'shelter', but somebody with no housing, a *sans-faute* is not a sort of *faute* 'mistake', but something (a race, an exercise, etc.) done without a mistake'.
(v) The complexes – at least in inanimate nouns – have the default gender, that is the masculine, whatever the gender of the "base": *faute* is feminine and *fil* 'line/wire' is masculine, but both *sans-faute* and *sans-fil* 'lit without-line, telegraph' are masculine.

Après- exemplifies a third case: it has some properties of the preposition and some of the prefix; consequently, it is to be situated between the two categories:
(i) It only expresses temporal posteriority whereas the corresponding preposition expresses both temporal and spatial posteriority:

(7) *l'après guerre* 'lit. the after-war, the postwar period', *l'après-match* 'the after match period'
il est né après la guerre 'he was born after the war' / *la boucherie est juste après la boulangerie* 'the butcher shop is just after the bakery'

(ii) *After-* belongs to the same paradigm as *post-*, a prefix (cf. *postdoctoral* 'ibid.');

(iii) It can only be involved in one categorical relation: N → N;

(iv) The complex lexemes are exocentric: *l'après-guerre* is not a *guerre* 'war', but the period after the war, *l'après-match* is not a match, but the period after the match, and so on;

(v) The gender of the lexemes built using *après* is not easy to determine, because the definite determiner (*le* masc. or *la* fem.) preceding them is systematically elided (*l'*) before a vocalic initial, as in *après*. Moreover, some / a limited number of lexemes are masculine and feminine (*après-midi, après-guerre*).

In the end, then, *après* presents two characteristics that are typical of prefixes and two characteristics that are typical of prepositions.

These differences can be related to the dates of the first attestation of the preposition / adverb as a formative:
– *sur-* was a prefix in Old French from early on, since it is the French counterpart of *super*, which was already a preposition, an adverb and a prefix in Latin;
– the first uses of *après* as a formative date from the fifteenth century; whereas the first uses of *sans* as a formative only date from the eighteenth century.
These items can therefore be ordered on a grammaticalization scale.

2. An example: the emergence of the first nouns formed by *après*

The first noun in which *après* appeared is *après-disner* 'after-diner", at the beginning of the fifteenth century.

This noun is the result of a gradual reanalysis of *après disner*, which was first an adverbial phrase, formed out of the adverb / preposition *après* and the infinitive verbal form *disner*.

(8) a. [Baye, II, 1411-1417, 28] Cedit jour, **après disner**, je fu mandé de par monseigneur de Bourgoigne[1] 'this day, **after diner**, […]
b. [Chartes Abb. St-Magl. T.F., t.3, [1330-1436], 650] Comme le dymanche XXVIIIe jour du moys d'avril derrenierement et nagaires passé, environ deux heures **aprés disner**, 'about two hours **after diner / after dining**'
c. [Reg. crim. Chât., I, 1389-1392, 195] elle estant seule en l'ostel dudit Hennequin, son maistre, à heure d'**après disner**
'lit at hour of **after diner**'

In (8 a) *après disner* is really an adverbial phrase, but in (8 b & c) it is a kind of complement to the temporal noun, without any preposition in (8 b), and introduced by the preposition *de* in (8 c). In these two last contexts, *après-disner* is not really an adverb anymore, and *disner* tends to be reanalysed as a noun : *après-disner* in these examples could be glossed by 'after the disner', and also by 'after having diner'.

[1] This example and the following come from the *Base de Lexiques de Moyen Français* (*DMF1*), http://www.atilf.fr

Gradually, *après-disner* evolves / turns into a noun; it can then be used in contexts where it is coordinated with a noun as in (9):

(9) [Baye, I, 1400-1410, 140] Cedit jour, à **matin** et **après disner**,
'This day, at **morning** and **after diner**'

The expression has acquired full nominal status when it is used as the head of a phrase and is preceded by a determiner, such as in (10); we find (at least) two spellings for *diner*: *disner* or *disnee*:

(10) [Ch. d'Orléans, *Chanson*, 1415-1440, 248] En gibessant toute **l'aprés disnee**
'hunting all **the after diner**'
[Commynes., III, 1495-1498, 130] à **l'après disner**,
'at **the after diner**'

The only other Old French nouns with *après* as a first constituent are *après-souper* (1502) and *après-midi* (1702). During the next centuries, but a few nouns were coined using *après-* (for example *après-minuit* by Flaubert, nineteenth century); it is only from the middle of the twentieth century that this mode of formation became more productive: *après-guerre* 'the post war years'*, l'après-match* 'the after match period'*,* and above all, many words of the form *après* + PN (proper name), this PN can notably be an anthroponym (*l'après-Ceaucescu* 'the after-Ceaucescu period'*, l'après-Bush* 'the after-Bush period', a date (*l'après-11 septembre* 'the after-eleventh of September period'), and so on.

I will now set out the framework of Teresa Vallès to see if it can account for these data.

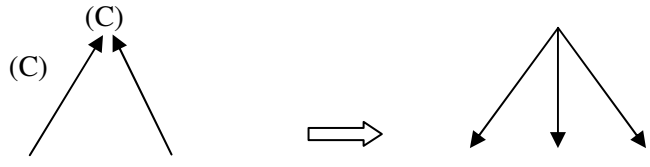
3. Teresa Vallès's proposal

T. Vallès, who is working in the framework of Joan Bybee's usage-based model, considers that neologisms are created by analogy, in its more basic sense: "any change due to the influence of one form on another" (Joseph 1998 : 362). She distinguishes between punctual and conventional analogy.

Punctual analogy is the extension from one exemplar, and it contains the following stages (in my translation):

– first, a neologism is coined as the result of the extension of one exemplar used as a model, or a prototype: A → B. Once this neologism is created, the schematization and the abstraction can produce a new pattern, C.

– the new pattern, which is a low level pattern, can be used to analyse the two words, the model and the neologism coined from this model, in such a way that A and B become actualizations of the new pattern C, that can become productive. Becoming a productive pattern often depends on the frequency of the elements in play. T. Vallès schematizes this evolution as in figure 1, in Vallès (2004: 147):



…

figure 1

In conventional analogy, on the other hand, a new pattern can become a sort of rule thanks to the ability of the language users to abstract patterns from existing words; these abstract

patterns are then conventionalized and used for regular word formation. This sort of "rules" is also based on analogy, but T. Vallès does not really explain in what sense they are different from punctual analogy.

Then, there is a continuum between the lexical innovation by analogy, the creation of low level patterns and the creation of abstract conventionalized patterns.

It is worth noting that this conception of lexical innovation and its extension to the coinage of other words is underlain by two strong principles:
(i) Language users play an active role by their ability to abstract low level patterns or more abstract ones, which allow them to create word formation rules;
(ii) The mental lexicon contains all existing words (the regular and the irregular ones, the simple and the complex ones, the affixes and the lexemes, etc.), and these constitute a network in which all the nodes share multiple and complex relations. These multiple connexions between the words in the mental lexicon allow the language users to abstract the patterns that will serve to create the neologisms or the rules.

Such a conception of the mental lexicon is frequently found in the work of linguists working in a diachronic perspective that is inspired by cognitive linguistic theories; in the present case, it is based on the usage-based model of Joan Bybee (1985), which goes back to the theories of Langacker (1987, among others).

How can Vallès'model be of interest for the problems we are dealing with ? Such a connexionist model is able to account for the first uses of *sans-, après-* and *sur-* (for example) as formatives. Indeed, what was said before about *après* can be transposed to this theoretical frame.

As was pointed out above, the first nouns *après* attaches to are *disner* at the beginning of the fifteen century, *souper* one century later, and *midi* two centuries after *après souper*. There is a great temporal distance between the coinage of these nouns.

At this point, two facts deserve to be pointed out:
– before their lexicalization, *après souper* and *après midi* were used in contexts similar to those of *après disner*: for both of them, the adverbial forms were attested, but they were not very frequent and the contexts in which they appeared showed less variation;
– the noun *after disner* had a very frequent use when *after souper* was coined.

Thus, it seems that once *après disner* was coined (by reanalysis from specific contexts of high frequency; on these general topics, see among others, Hopper & Traugott (1993), Heine *& al.* (1991), Lehmann (1995), etc), it could serve as a model for the formation of *après souper* and *après midi*: indeed, the former – *après souper* – involves an infinitive form, just like *après disner*, and the latter is formed using a particular kind of temporal noun that Van de Velde (2000) has called "temporal proper name" and that is generally used without a determiner, even in modern French.

Consequently, the assumption can be made that a low level pattern comparable to those described by Vallès has been set up, based on punctual analogy. The possibility of analogical processes relies on the notion of paradigm, underlain by the notion of network.

It is however more difficult to account in the same way for the extension to regular formations, above all those in *sur-*, when *sur-* is a prefix, because, in my opinion
(i) the relation between the notions of conventional pattern and that of the emergence of abstract patterns is not sufficiently clarified; these two related notions are not made sufficiently explicit and it is difficult to understand to what extent there is "abstraction".
(ii) the notion of affix / prefix is insufficiently developed. Indeed, the only real criterion that is used is that of dependence: a prefix is phonologically and semantically dependent, while a stem is characterised by its independence. A second criterion, which is a semantic one, appears when T. Vallès writes:

> "The préfixoïdes *euro-* and *eco-* can be considered prefixes, because they have no autonomy […]; but they are not prototypical prefixes on account of their semantic content, which is that of a lexical unit" (Vallès 2004: 191) [my translation]

However, T. Vallès does not give more detailed information concerning the semantic value of the prefixes. One would expect identification criteria and a precise definition of what is a prefix.

Moreover, many researchers have attempted to show that items such as *euro-* and *eco-* are not prefixes but bound-stems (or bound-roots), and that they enter in compounding processes (on this topic, see for example Corbin 1992, Bauer 1997, 1998, Lüdeling & *al.* 2002, Iacobini 2004, Namer & Villoing 2005 & 2006, Ralli 2007 & 2008, Dal & Amiot 2008).

Let us now have a look at the analysis put forward by Geert Booij.

4. Geert Booij's proposal (2005 and 2008)

In the two papers I refer to, G. Booij adapts the principles of Goldberg's Construction Grammar to the domain of morphology. This theory also originates from Cognitive Grammar. The fundamental assumptions concerning the existence of immanent abstract patterns, the role of the language user and the lexicon are more or less similar to those T. Vallès adheres to. I quote:

> "Language users acquire knowledge of these abstract morphological schemas on the basis of their knowledge of a set of words that instantiate this pattern. Once they have come across a sufficient number of words of a certain type, they can infer an abstract scheme, and will able to extend that class of words." (Booij 2005: 124)

But, contrary to Vallès, Booij does not focus on this question, what he wants is to

> "develop a constructional theory of word formation that makes use of some basic ideas of Construction Grammar, in particular constructional schemas, the idea of a hierarchical lexicon, multiple linking, and intermediate levels of generalization for an adequate account of word formation. These ideas have also been developed in the framework of Cognitive Grammar (Langacker 1988, 1998)." (Booij 2008)

According to Booij, patterns of word formation can be considered as constructions, that is, in the terms of Goldberg (1995), who studies verbal constructions:

> "Form-meaning correspondences that exist independently of particular verbs. That is, it is argued that constructions themselves carry meaning, independently of the words in the sentence" (Goldberg 1995: 1)

According to Booij, there exist general morphological patterns for each process of word formation; here is that of compounding:

(11) $[[a]_X\ [b]_{Yi}\quad]_Y$ 'Y_i with some relation R to X'
$\qquad\qquad\qquad\quad\ \ |\qquad\ \ |$
$\qquad\qquad\qquad\ [\alpha F]\ \ [\alpha F]$

Such a schema is in fact a form-meaning correspondence, it accounts for the fact that the compounds of Dutch are right-headed endocentric; the subscript [αF] indicates that the pertinent features (gender, declension class for nouns, and so on) percolate from the right constituent to the complex word. The relation R between X et Y is not specified at this level.

And here is the general pattern of prefixation:

(12) $[a\ [b]_Y\quad]_Y$ 'Y modified by *a*'
$\qquad\quad\ \ |\quad\ \ |$
$\qquad\ \alpha F\quad \alpha F$

The principles are similar but there are some differences:
– a prefix, like a suffix, has no lexical category;
– prefixes are normally category-neutral, so the properties of the base are transferred to the complex word;
– a prefix has the semantic function of a modifier.

Booij makes the assumption that in the morphological domain there also exist constructional idioms (in the sense of Jackendoff 2002); a constructional idiom in morphology is an abstract construction in which one position is occupied by a specific lexical item. For example, in French, there exist constructional idioms involving the prepositions *sans* or *après*:

(13) $[[sans]_P\ [y]_V\ ]_V$ 'something / somebody without Y'
$[[après]_P\ [y]_V\ ]_V$ 'period after Y'

These schemes are to be interpreted as follows: when the preposition *sans / après* and a noun are conjoined to form a compound noun, this noun means, respectively, 'something / somebody without Y' such as in *sans-abri, sans-papiers* or *sans-faute* / 'period after Y', such as in *après-guerre, après-match* or *après-Ceaucescu*.

Constructional idioms can also be set up for affixation; here is the example of the prefix *sur-*:

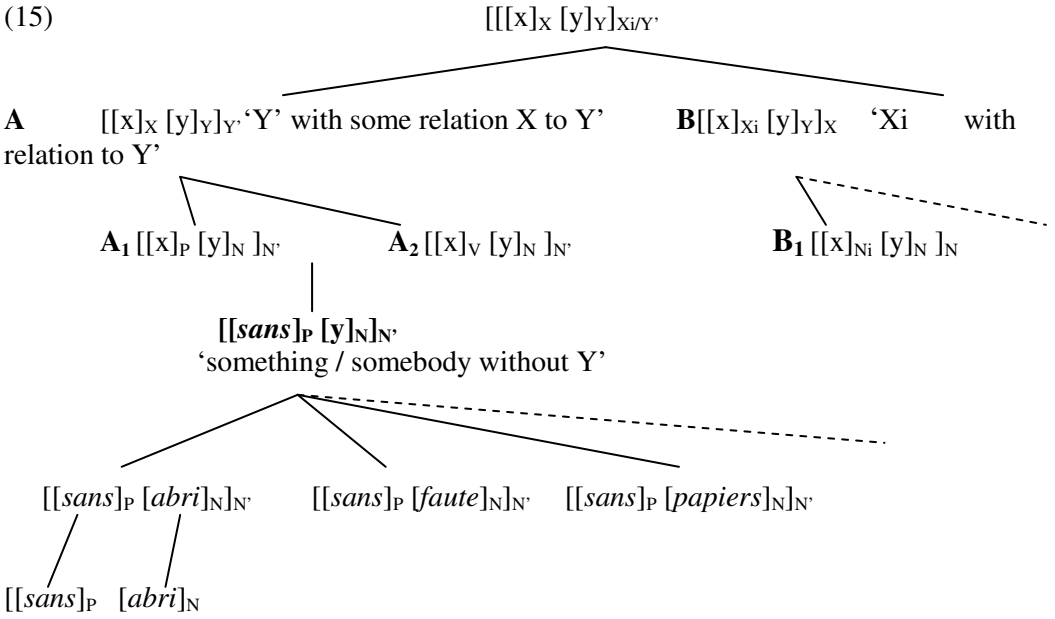
(14) $[sur\ [y]_y\ ]_y$ 'Y in excess'[2]

The two *x* subscripts in (14) indicate that, as was pointed out under 2., the lexemes built by *sur-* belong to several categories (nouns, verbs and adjectives) and the complex lexeme is of the same category as its base.

Constructional Idioms are intermediate sub-schemes between the general scheme and the individual words; they can be represented as follows, for example for *sans* (it is an adaptation, and an extension, from Booij 2005 and 2008):

[2] This is not the only meaning a lexeme built with *sur-* can express, but it is indeed the most frequent, expressed by the bulk of types of bases.

(15)



This representation needs some clarification:
(i) The Constructional Idiom (in bold type in (15)) is in fact situated on an intermediate level between the most general levels and the individual items. This representation takes the shape of a multiple inheritance tree, where the lower nodes inherit the properties of their dominating nodes.
(ii) I make the assumption that, in French and possibly in other Romance languages, there are two general schemes of composition, one for the formation of exocentric lexemes (A), and another for the formation of endocentric ones (B). The scheme for exocentric lexemes subsumes two other sub-schemes, one in which the first constituent is a preposition (A_1) and another in which the first constituent is a verb (A_2), such as, for example, in *brise-glace*: 'lit. break-ice; icebreaker'. The scheme that accounts for the formation of endocentric lexemes can be exemplified by a lexeme such as *requin-marteau* 'lit. shark-hammer; hammerhead'.
(iii) As is also the case in other Romance languages, French compounds are left-headed when they are endocentric and when they are exocentric, their interpretation beginning at the left. For the exocentric compounds, the prime in Y' or N' indicates that, although the compound is of the same category as the second constituent, it is semantically different (it is a consequence of exocentricity). Conversely, the subscript *i* indicates the semantic head for the endocentric compounds.
(iv) The constructional idiom of *sans-* is an instantiation of the schema of exocentric compounds, and more precisely, of A_1.
(v) The last line indicates that each constituent of a compound is linked with its corresponding syntactic counterpart, and also inherits properties from it.

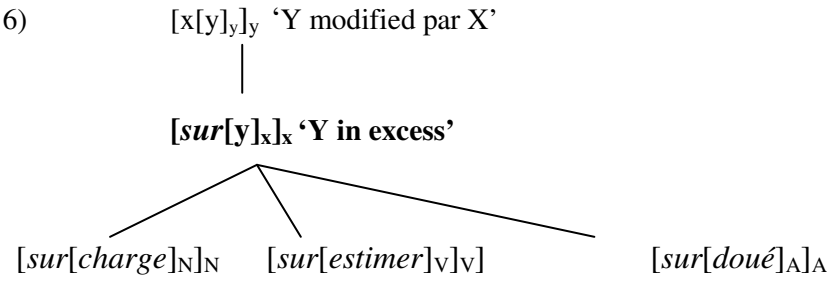
I also want to insist on the fact that constructional idioms cannot be hypothesized for all sorts of compounding; it is possible to do so when the first constituent is a preposition, but not in the two other schemes of the representation under (15), A_2 et B_1. For these, there is no intermediate level between the pattern and the individual lexemes because no regular series of lexemes are formed from one of the two constituents[3]; therefore, the

[3] In the best case, some words are created with the same first verb (*garde-malade* 'home-health aide', *garde-feu* 'fireguard', *garde-manger* 'larder', etc.) or with the same first noun (*requin-marteau, requin-baleine* 'whale shark', *requin-tigre* 'tiger shark', etc.).

emergence of constructional idioms can be seen as the first step toward the path that leads from composition to derivation.

As for the prefix *sur-*, its constructional idiom is a part of the general scheme of affixation, and more precisely the scheme of prefixation, that can be simplified in the following way:

(16)



Booij never considers a formative as a prefix as long as it corresponds to a preposition; thereby, such a representation could only hold, in the analysis of Booij, for a prefix such as *hyper-,* a prefix that belongs to the same paradigm as *sur-*, but without any correspondent preposition in French. But this difference in the analysis is not very important here.

Booij's model is interesting in that it offers a general overview of the architecture of word formation processes, and of the sub-regularities which characterise it, these sub-regularities being represented by the constructional idioms, on which accurate constraints are exerted. But, from our perspective, which is diachronic, the following question arises: how is it possible to conceptualize in this framework the evolution from the creation of the first complex words in which the left constituent is a preposition to the morphological operation of prefixation *via* the stage of the formation by an operation of compounding?

When the first lexemes occur, a low level pattern (cf. Vallès) can be created (cf. the emergence of the first words with *après*, § 3.). Such a pattern can correspond to an existing pattern, as a matter of fact $[[x]_P\ [y]_N\ ]_{N'}$. When the process becomes regular, a constructional idiom can emerge, for example $[[après]_P\ [y]_N]_{N'}$, in which *après* is still a preposition, and the complex lexemes are exocentric. The difficulty lays in the evolution toward a derivational constructional idioms: if, for example, *après* become a real prefix, as *sur-* is, the evolution would be:

(17) $[[après]_P\ [y]_N]_{N'} \rightarrow [après[y]_x]_x$

As a prefix, *après* would be a modifier and would form endocentric lexemes (of several categories), just as *sur-* does. Such an evolution is not easy to conceptualize: what are the operations allowing it?

A way to conceptualize it is to consider that there has been a "leap" in the evolution: when the formative has acquired enough autonomy with respect to the preposition it comes from (= when it has become "light" enough, cf. the criteria at the beginning of the paper), it can be integrated in the general pattern of derivation ($[x[y]_y]_y$; in this way, it "becomes" a modifier able to form endocentric lexemes; then, it can carry on its evolution (in, gradually, combining with lexemes belonging to other categories than the nominal category).

5. Perspectives

Booij works in the theoretical framework of Construction Grammar, but also in the framework of Item & Arrangement, which, according to Aronoff and Fudeman (2005) "proceeds from a picture of each language as a set of elements, and the patterns in which those elements occur" (Aronoff and Fudeman 2005: 47). In such a framework, the affixes are morphemes (that is pairing form-meaning) like the lexemes they attach to. Considering that affixes are morphemes allows one to conceptualise quite easily the path that leads from the domain of composition to that of derivation, which is an important result when one works in a diachronic perspective.

However, the Item & Arrangement perspective has two disadvantages:

– it is not very easy to account for the paradigmatic relations between affixes in this frame;

– it isolates affixal derivation from non affixal word formation processes, such as back-formation, shift stress, duplication, and so on.

Consequently, what needs to be examined now is whether an Item and Process perspective, which, in the words of Aronoff & Fudeman "gives no independent status to the items" and where "complex words result from the operation of processes on simpler words" (*ibid.*) is compatible with the analysis developed here (notably the evolution from compounding to derivation). Some researchers begin to adopt this frame for research in a diachronic perspective, for example Amanda Pounder, who studied the evolution of denominal adjectives in German.

Bibliography

D. Amiot, Préfixes ou prépositions? Le cas de *sur-, sans-, contre-* et les autres, *Lexique* 16, Presses du Septentrion, Villeneuve d'Ascq, 2004, pp. 67-83.

D. Amiot, and W. De Mulder, Les préfixes *avant-* et *sur-* en français et les chemins de grammaticalisation, In M. Grossmann, and A. M. Thornton, (eds), *La formazione delle parole,* Bulzoni, Roma, 2005, pp. 31-51.

D. Amiot, Prépositions et préfixes, *Modèles Linguistiques* 53/1, Université du Sud Toulon-Var, Toulon, 2006, pp. 19-34.

M. Aronoff, and K. Fudeman, *What is Morphology?*, Blackwell Publishing, Oxford, 2005.

Bauer, L., *English Word-formation*, Cambridge University Press, Cambridge, 1994.

L. Bauer, Is there a Class of Neoclassical Compounds, and if so, is it Productive?, 1998, *Linguistics* 36/3, pp. 403-422.

G. Booij, Compounding and Derivation: Evidence for Construction Morphology, In W.U. Dressler, D. Kastovsky, and F. Rainer (eds.), *Morphology and its Demarcations*, 2005, John Benjamins. Amsterdam/Philadelphia, pp. 109-132.

G. Booij, Composition et morphologie des constructions, In D. Amiot (éd.), *La composition dans une perspective typologique*, 2008, Artois Presses Université, Arras, pp. 49-74.

Bybee, J., *Morphology. A Study of the Relation between Meaning and Form*, John Benjamins, Amsterdam/Philadelphia, 1985.

D. Corbin, Hypothèses sur la frontière de la composition nominale, *Cahiers de grammaire* 17, 1992, Université de Toulouse Le Mirail, Toulouse, pp. 189-218.

D. Corbin, Locutions composées ; unités polylexématiques: lexicalisation et mode de construction, In M. Martins-Balter (éd.), *La locution entre langue et usage*, 1997, ENS editions, Fontenay/Saint-Cloud, pp. 53-101.

G. Dal, and D. Amiot, Integrating Combining Forms into a Lexeme-Based Morphology, In G. Booij *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5*), 2007, University of Bologna. URL http://mmm.lingue.unibo.it/

G. Dal, and D. Amiot, Composition néoclassique en français et ordre des constituants, In D. Amiot (éd.), *La composition dans une perspective typologique*, 2008, Artois Presses Université, Arras, pp. 89-114.

W.U. Dressler, and M. Ladányi, Productivity in Word Formation (WF): a Morphological Approach. *Acta Linguistica Hungarica* 47/2, 2000, Budapest, pp. 103-144.

B. Fradin, L'approche à deux niveaux en morphologie computationnelle et les développements récents en morphologie. *T.A.L.* 35/2, 1994, Hermès Sciences, pp. 9-48.

A. Goldberg, *A Construction Grammar Approach to Argument Structure*, The University of Chicago Press, Chicago, 1995.

Heine, B., U. Claudi, and F. Hünnemeyer, *Grammaticalization: a Conceptual Framework*, University of Chicago Press, Chicago, 1991.

Hopper, P. and E. C. Traugott, *Grammaticalization*, Cambridge (UK), Cambridge University Press, 1993.

C. Iacobini, Distinguishing Derivational Prefixes from Initial Combining Forms. In G. Booij, A. Ralli and S. Scalise (eds.), *Proceedings of the First Mediterranean Conference of Morphology*, 1998, University of Patras, Patras, pp. 132-140.

C. Iacobini, Prefissazione. In M Grossmann & F. Rainer (eds.), *La formazione delle parole in italiano*, 2004, Max Niemeyer, Tübingen, pp. 97–161.

Jackendoff, R.S., *Foundations of Language*, Oxford University Press, Oxford, 2002.

B. Joseph, Diachronic Morphology, In A. Spencer and A.-M. Zwicky (ed.), *The Handbook of Morphology,* 1998, Blackwell Publishers, Oxford/Malden Mass, pp. 351-373.

M. Ladányi, Productivity as a Sign of Category Change, In W. U. Dressler, O. F. Pfeiffer, M. Pöchtrager and J. Rennison (eds), *Morphological Analysis in Comparison*, 2000, John Benjamins, Amsterdam/Philadelphia, pp. 13-141.

Langacker, R., *Foundations of Cognitive Grammar. Vol. I: Theoretical Prerequisites*, Stanford University Press, Stanford, CA, 1987.

R. W. Langacker, A Usage-Based Model, In B. Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, 1988, John Benjamins, Amsterdam/Philadelphia, pp 127-161.

R. W. Langacker, Conceptualization, Symbolization, and Grammar, In M. Tomasello (ed.), *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure*, 1998, Lawrence Erlbaum Ass. Publ., Mahwah NJ./London, pp. 1-39.

Lehmann, C., *Thoughts on Grammaticalization*, Lincom Europa, München/Newcastle, 1995.

A. Lehrer, Are Affixes Signs?, In W. U. Dressler, O. F. Pfeiffer, M. Pöchtrager and J. Rennison (eds), *Morphological Analysis in Comparison*, 2000, John Benjamins, Amsterdam/Philadelphia, pp. 143-154.

A. Lüdeling, & *al.*, Neoclassical Word Formation in German. In G. Booij and J. van Marle (eds.), *Yearbook of Morphology 2001*, 2002, Kluwer, Dordrecht, pp. 253-283.

F. Namer, and F. Villoing, Assigning Category to Non-autonomous Bases in Neoclassical Compounding, *12th International Morphology Meeting*, May 25-28, Budapest, 2006.

Plag, I., *Morphological Productivity. Structural Constraints in English Derivation*, Mouton de Gruyter, Berlin/New York, 1999.

Pounder, A., *Processes and Paradigms in Word-Formation Morphology*, Mouton de Gruyter, Berlin/New York, 2000.

A. Ralli, Compound Marking in a Cross-Linguistic Approach, In N. Hathout and F. Montermini (éds), *Morphologie à Toulouse*, 2007, Lincom Europa, Munich, pp. .

A. Ralli, Composés déverbaux grecs à radicaux liés. In D. Amiot (éd.), *La composition dans une perspective morphologique*, 2008, Artois Presses Université, Arras, pp. 189-210.

S. Scalise, *Generative Morphology*, Foris, Dordrecht, 1984.

P. Ten Hacken and D. Smyk, Le rôle de l'analogie et des règles dans la formation des mots, *Travaux linguistiques du Cerlico* 16, 2003, pp. 11-26.

Vallès, T., *La cretivitat lèxica en un model basat en l'ús*, Publications de l'Abadia de Montserrat, 2004.

D. Van de Velde, Existe-t-il des noms propres de temps?, *Lexique* 15, 2000, pp. 35-45.

Van Goethem, K. *La grammaticalisation comme paramètre en linguistique comparative. Le cas de l'emploi préverbal des prépositions du français et du néerlandais*, Thèse de Doctorat, Université de Leuven, 2006.

J. Van Marle, Rule-Creating Creativity: Analogy as a Synchronic Morphological Process, In W.U. Dressler and *al.* (eds). *Contemporary Morphology*, 1990, Mouton de Gruyter, Berlin, pp. 267-273.

J. Van Marle, Paradigmatic and Syntagmatic Relations, In G. Booij, Ch. Lehmann and J. Mugdan (ed.), *Morphology. A handbook on Inflection and Word Formation*, 2000, Walter de Gruyter, Berlin/New York, pp. 225-234.

New insights into the rivalry of suffixes

Carola Trips
Universität Stuttgart
carola.trips@ling.uni-stuttgart.de

Abstract

This paper deals with the claim that parallel derivations act as rivals and compete with each other. This claim is based on the assumption that word-formations with the same base and diffirent suffixes are synonymous. I will show with a study on the development of the Modern English suffixes -hood, -dom and -ship that this is not the case. In fact, these elements which used to be nouns and developed into suffixes via a stage in which they act as morphological heads of compounds bear salient meanings they had when they had the status of free morphemes as well as meanings that developed from metonymic shifts. These meanings differ and therefore the suffixes are not synonymous and do not compete with each other. I will propose a lexical-semantic analysis of my findings based on Lieber (2004) that takes into account the diachrony of these elements.

1 Introduction

In Modern English (ModE) so-called parallel derivations like kinghood, kingdom and kingship occur examples of which are given in (1) to (3). Here we find the same base and different suffixes with the same function of building abstract nouns.

(1) Panchatantra is a collection of ancient Indian tales written by a wise man to teach the king's children about **kinghood**.
(http://a.parsons.edu/ radhika/thesis/briefcase.swf, 25/4/07)
(2) Introduction to the Plant **Kingdom**, their morphology & life cycles.
(scitec.uwichill.edu.bb/bcs/bl14apl/bl14apl.htm, 25/4/07)
(3) We beseech you, your **kingship**, to institute a system of hereditary peerage based upon merit and loyalty (i.e., campaign contributions) so that we peasants will have someone to look up to other than the tawdry celebrities on TV.
(http://blog.peakdems.org/2005_12_25_peakdems_archive.html, 25/4/07)

The elements in bold – *kinghood, kingdom, kingship* – can syntactically be identified as elements of the type noun and morphologically as free morphemes. Moreover, as mentioned above, we can clearly identify the base king and the bound derivational sufies *-hood, -dom* and *-ship*. With respect to the semantics of the word-formations found in these contexts, *kinghood* denotes the state of being a king, *kingdom* the realm of plants and *kingship* the form of address of a king. We will see below that the meaning of these formations is composed of the meaning of the base and the suffix, i. e., it will be claimed that suffixes contribute to the meaning of formations.

In the literature (see Aronoff 1976, Plag 2003) it has been observed that wordformations with the same base and different suffixes occur and seem to denote the same type category, e.g., abstract nouns. Therefore, it has been claimed that they act as rivals and compete with each other (e.g. Tschentscher 1958, Dalton-Puffer 1996, Plag 2003). This similarity, or better the semantic relatedness between formations with -hood, -dom and -ship is often stated the following:

> *The native suffix -dom is semantically closely related to -hood and -ship, which express similar concepts*. (Plag 2003, 88)

However, it is never exactly stated what "similar concepts" really means and which differences the term "similar" also actually implies. Does it mean that parallel derivations are semantically differentiated sufficiently to call them "different words" or does it mean that they are synonymous? According to Martin (1906, 71) the latter applies to parallel derivations: "... in der grossen Mehrzahl der Fälle keine Verschiedenheit der Bedeutung". In line with Martin,

Dalton-Puffer (1996, 128) comes to the conclusion that "... in the absence of any positive evidence pointing towards systematic meaning differences between parallel formations in Middle English, I assume the suffixes involved to be synonymous". Tschentscher (1958, 180) also assumes competition ("TUM und seine Konkurrenten") which is based on shared meanings between the three German suffixes *-tum, -heit* and *-schaft* but also on differences. The fact that e.g. *-tum* and *-heit* compete is based on the meaning "Würde, Rang" ('dignity, rank'). The semantic relatedness between formations with -schaft and -heit is due to their active reading.

On the other hand, there are semantic differences, e.g., the difference between *-heit* and *-tum* which is based on the static character of the latter element. The difference between -schaft and *-heit* is that the former takes nominal bases that denote offices and ranks the bearers of which have to become one instance of this office or rank and were not born with it. What all three elements have in common is, according to Tschentscher, the meaning 'power', and that is also why they act as rivals (she illustrates this point with the formations *Aposteltum, Apostelheit* and *Apostelship*). In the following, I will show with diachronic data that suffixes do not act as rivals in parallel derivations. My claim is that suffixes are semantically different enough to coexist because they bear salient meanings they acquired in the course of time as well as metonymies resulting from these salient meanings. This leads to the assumption that suffixes have meaning that contributes to the meaning of word-formations with these suffixes, implying that there is an interplay between meaning of base and suffix.

The starting point of my diachronic study is Old English (OE) where the three elements under investigation had the status of abstract nouns with a number of salient meanings:

(4) a. *hād*: 'status, office, rank'
b. *dōm*: 'authority, judgement'
c. *scipe*: '(resultant) state, condition'

On their way to becoming suffixes, further meanings arose from metonymic shifts. Hence, the salient meanings given in (4) as well as meanings that arose from metonymic shifts are part of the meaning of the ModE suffixes:

(5) a. *-hood*: 'a distinguishing feature of one's personal nature' (salient meanings 'status, office, rank', metonymies 'state, place, time')
b. *-dom*: 'possession of the qualities required to do something or get something done' (salient meanings 'authority, judgement', metonymies 'territory, realm')
c. *-ship*: 'result of a process of creating' (salient meanings 'created thing, '(resultant) state', metonymies 'function, forms of address, skill/art')

The examples in (6) to (8) from Early Modern English (EME) illustrate that the three suffixes also bear these metonymic meanings:

(6) *A ploughman of your **neighbourhood** that has never been out of his parish.* (LOCKE,75.150)

The formation *neighbourhood* clearly has a local reading. The example shows that the meaning 'surrounding area or district' a metonymy. This also applies to the formation kingdom in the following example:

(7) *He said the pope had declared that England was his **kingdom**, and that he had sent over commissions to several persons.* (BURNETCHA,2,166.172)

In Old English (OE) and Middle English (ME), the salient meaning of *kingdom* was 'dominion, authority of a king', in this example it denotes 'the territory over which a king's power extends, realm, country' (see the Middle English Dicitionary (MED)). Thus, this meaning arose due to a metonymic shift from the authority of a king to the territory over which a king has the authority. Example (8) with the formation horsmanship also shows a metonymic meaning:

(8) *The end of **Hors-manshippe**.* (MARKHAM,1,86.112)

The formation used to denote "chivalrie ... also a tenure by Knights service" (Lexicons

of Early Modern English (LEME)) and developed into 'the skill involved in riding horses' (Dictionary of Contemporary English (DCE)). Thus, what these examples show is that the ModE suffixes include salient meanings (see (4)) as well as a number of metonymies. When we compare the three suffixes, we find similar meanings like 'rank', 'office' and 'authority', but there are also a number of meanings that only occur with one of the suffixes like 'skill/art'. As claimed above, these clear differences in meaning result from different salient meanings and metonymic shifts of these meanings. The semantic similarity that has been observed in the literature is that diachronically the three nouns *hād, dōm* and *scipe* denote abstract states and refer to persons:

(9) *þæt he þǣr onfenge* ***ærcebisceopes hāde.***
that he there received archbishop's office
(Bede_3:21.248.11.2540)

(10) *&* *þis folc ne oncneow* ***Godes dōm.***
and this people not knew God's judgement
(ÆCHom_I,_28:412.64.5503)

(11) *Hæbbe ic mīnes* ***cynescipes*** *gerihta swa mīn fæder hæfde, and*
have I my kingly position's rights as my father had and
mīne þegnas hæbben heora ***scipe.***
my servants have their position
(BT, L.Edg. ii. 3; Th. i. 266, 15-18)

In (9) *hād* denotes the office of an archbishop (both with a referential and non-referential reading), in (10) *dōm* denotes the judgement of God, and in (11) *scipe* denotes the position of servants.

For ModE, we can assume that the three suffixes building abstract nouns form a paradigm with a general feature 'state of N' (that is where they overlap, compare the meanings given in (5) again), but each of them also shows meanings different from the meanings of the other suffixes in the paradigm. In the following, I will show that these semantic differences play a crucial role in the distribution of derivations with the three suffixes today and actually can explain a) why sometimes a derivation is only possible with one of the suffixes b) why parallel derivations with different suffixes never denote the same meaning.

In the following section, I will discuss Aronoff & Cho's (2001) analysis of ModE *-ship* suffixation because some of my theoretical assumptions (sections 4 and 5) will be based on their analysis.

2. *-ship*-suffixation in ModE: Aronoff & Cho's (2001) analysis

Aronoff & Cho (2001, hence A & C) present an account for part of the observations made above by exploring the semantic conditions of *-ship*-suffixation in ModE. They propose that the type of predicate a base belongs to explains the occurrence or nonoccurrence of such a derivation, drawing on the distinction between stage-level and individual-level predicates (Carlson 1977). A stage-level predicate applies to temporary states and denotes properties of states. The noun friend is one example. The temporary nature of these predicates allows them to occur with an adjective like longtime as in John's longtime friend. An individual-level predicate, on the other hand, applies to an individual without taking into consideration time, it denotes stable properties of individuals. That is why it cannot be modified by longtime: **John's longtime parents*. A & C claim that this distinction plays a crucial role in ModE *-ship*-suffixation and illustrate this with the following examples (2001, 168):

(12) a. airmanship, friendship, kingship, penmanship, priestship, sponsorship
b. ??parentship, ??wifeship, ??nieceship, ??womanship

They note that the base nouns *airman, friend, king, penman, priest* and *sponsor* are all stage-level predicate since they denote instable, transient stages. In contrast, the base nouns *parent, wife, niece* and *woman* are all individual-stage level predicates since they denote stable

properties of individuals[1]. Strikingly, it seems to be the case that –ship can only occur with stage-level predicates, and that is why a formation like childship does not exist whereas childhood does. Therefore, A & N propose the following semantic condition:

(13) $X[_N$ -ship$]_N$

Condition: X is a stage-level predicate.

They further claim that "the semantics of the base selects the specific meaning of -ship" (2001, 169). For example, if a base like *friend* is a relational noun (it denotes a relation between friends) the formation with *-ship* denotes that relation. If, on the other hand, the noun is not relational like e.g. *penman* suffixation with *-ship* is still possible although it has a meaning different from friendship; the formation *penmanship* denotes the skill or art of a penman, thus showing that according to the meaning of the base the specific meaning of the suffix is selected.

A further example is priestship where the nominal base denotes the office of a priest and therefore the derived word also denotes the office of a priest. According to A & C, all the different meanings found with *-ship*-formations can be reduced to one general meaning, the meanings of the individual *-ship*-formations being determined by the combination of the base and the suffix[2]. More precisely they assume

> ... the meaning of a -ship word selects the stage-level property that is most salient in the meaning of the base. In particular, if the base is relational, the output denotes that relation; if the base denotes someone who has a skill, the output denotes that skill; if the base denotes someone who occupies a position in a hierarchy, the output denotes that position of period of office.
>
> (Aronoff & Cho 2001, 171)

Coming back to the observation that ModE formations with -hood, -dom and –ship are semantically related, implying similarities and differences, A & N's assumptions can explain the different behaviour of ModE suffixation with -ship and -hood. They give the following word-formations based on Webster's Third New International Dictionary comparing derivatives with the suffixes -hood and -ship (see Aronoff & Cho 2001, 172):

Comparison of formations with –hood and –ship

-ship	*-hood*
apprenticeship	apprenticehood
bachelorship	bachelorhood
_____	**childhood**
doctorship	doctorhood
fathership	fatherhood
_____	**girlhood**
kingship	kinghood
ladyship	ladyhood
_____	**manhood**
_____	**motherhood**
neighborship	neighborhood
_____	**parenthood**
priestship	priesthood
queenship	queenhood
_____	**sisterhood**
_____	**wifehood**

[1] A & C introduce a subclassification of these predicates into left-side individual predicates and right-side individual predicates. The former type denotes properties that individuals have at birth until a certain point in time (e.g. *child*), whereas the latter type denotes properties that individuals have from a certain point in time up to death (e.g. *mother*).

[2] A & C actually talk about the combination of the base and the context, but since it is clear from their argumentation that with context they do not mean textual context they seem to mean the elements around the base, i.e., the suffix.

Table 1:Comparison of formations with -hood and -ship

The comparison shows that whenever a base denoting an individual-level predicate occurs there is no formation with *-ship*. Thus, this finding corroborates A & C's claim that -ship is sensitive to the semantics of the base. However, this does not seem to be a problem for *-hood*-suffixations: the suffix occurs with stage-level predicates as well as with individual-level predicates (see the formations *childhood*, *girlhood*, etc.). This implies that the semantic restriction on the *-ship* suffix to allow only bases denoting properties of stages is part of the lexical semantics of the suffix. What A & N have not included in their list are -ship-formations with bases denoting individual-level predicates, although parallel derivations like fathership and fatherhood do exist. How can they be explained? A & C assume that a formation like fathership can occur, although *father* is an individual-level predicate because the formation does not denote stable properties of individuals: *fathership* denotes 'the condition or state of being the oldest member of a community' and thus has stage-level properties, whereas *fatherhood* denotes 'the state or condition of being a father' denoting stable, enduring properties of individuals[3]. What is crucial for their analysis is that *-ship*-formations are determined by the semantics of the base. I claim that the synchronic facts (parallel derivations) can only be explained if we take into account that suffixes bear meaning and that this meaning results from their history. This will become evident in the following section, which surveys the lexical-semantic history of *-hood*, *-dom* and *-ship*.

3. Data: the diachrony of -hood, -dom and -ship

As noted in the introduction, *hād*, *dōm* and *scipe* are free nouns in OE with the following
salient meanings ((4) is repeated here as (14)):

(14) a. *hād*: 'status, office, rank'
b. *dōm*: 'authority, judgement'
c. *scipe*: '(resultant) state, condition'

In my study, the semantic status of the bases was defined according to the data and categorised as being either of the stage-level or of the individual-level predicate type, in line with A & C (since a clear-cut classification of nouns in this respect does not exist, see Kratzer 1995, Maienborn 2001, Geist 2006). The result of my survey on the lexial-semantics of the nouns found as bases in OE is the following definition for the two types of predicates[4]:

> *Stage-level predicates apply to a social state which is assigned to a person by society. From this state an activity can be inferred. Individual-level predicates apply to inherent properties of persons, which cannot be determined externally (e.g. from society).*
>
> (Trips 2007, 260)

This difference can be nicely illustrated with the nouns *priest* and *child*: a *priest* denotes the office of a person and as such a state which has been assigned to this person by society. A person can hold an office and resign an office. As soon as a person holds an office he has acquired a certain social status. In contrast, the noun *child* denotes the state of a young person or the period of time of being a child. This is an inherent property of human beings that cannot be determined externally, i.e., society cannot assign this state to somebody, it is inherently given.

The three elements *hād*, *dōm* and *scipe*, functioning as heads of compounds, predominantly occur with nominal bases that are of the stage-level predicate type. Further, they predominantly bear salient meanings and match the meanings of the bases. This implies that both the base and the suffix bear meaning, and the combination of these meanings results in the meaning of a

[3] Note that A & C do not assume (as Lieber 2004, 160 claims) that according to their theory these formation are not possible, they only claim that when they occur they can never denote enduring properties.

[4] The adjectival bases were classified accordingly. For a full account see Trips 2007.

formation. Since *had*, *dom* and *scipe* bear different salient meanings, parallel derivations also bear different meanings:

(15) *Ic ÆLFRED geofendum Criste mid* ***cynehādes*** *mærnysse geweorðod,*
I A. giving Christ with king-dignity's honour honoured
habbe gearolice ongyten & þurh hāligra bōca gesægene oft
have readily learned and through holy books prophecy often
gehȳred, …
heard
(GDPref_1_[C]:1.1.2)

In the context of (15), the formation *cynehād* denotes 'kingly state or dignity', and this state is granted by God according to the context. Thus, *cyne* denotes a stage-level predicate.

(16) *Ge habbað nū gehȳrod hū ða hālgan cyningas heora* ***cynedōm***
you have now heard how the holy kings their king-rule
forsawon for Cristes gelēafan and heora āgen līf forlēton for
rejected for Christ's belief and their own life abandoned for
hine.
him
(ÆLS_[Abdon_and_Sennes]:76.4771)

In (16), the formation *cynedōm* denotes the authority or rule of a king. The difference between cyneh ad and *cynedōm* can only be explained by the fact that *dōm* has a salient meaning different from *hād*, namely 'rule, authority', that is also part of the formation.

The same applies to formations with *scipe*:

(17) *Ne mæg nān eorðlic cyning cynelic lybban. būton he hæbbe*
NEG may no earthly king kingly live but he have
ðegenas. & swā gelōgodne hired swā his ***cynescipe*** *gerīsan*
servants and so arranged household as his king-status rise
mæge.
may
(ÆCHom_I,_8_[App]:533.3.1562)

In (17) *cynescipe* occurs with a possessive pronoun and the predicate "to rise" and denotes 'kingly power resulting from kingly dignity'. In the next example, a further meaning occurs:

(18) *Ealra manna hlāford geond þas wīdan worulde, we biddað þīnne*
all men's lord throughout the wide world we bid your
cynescipe *þæt þu nān ðingc ne bēo drēorig oððe sārig for*
majesty that you no thing NEG be cruel or sorrowful for
ðan geongan cnihton, …
the young warrior
(LS_34_[SevenSleepers]:254.195)

In this example, *cynescipe* is used as a title which can be seen by the fact that it is preceded by the second person possessive and the predicate "bid". This meaning is a metonymy arising from the salient meaning of *scipe* 'dignity'. As illustrated above with the EME example in (8), a further meaning 'skill, art' is added in the course of time, and here again we see why this special meaning could arise with *scipe* but not with the other two elements: if something is the result of being created, it can be assumed that the creator has special skills (e.g. special riding skills as in *horsemanship*). This meaning does not occur with *hād* and *dōm* because they have salient meanings that would not allow this metonymic shift.

These examples illustrate that all three elements can occur with the same base and are in this respect parallel derivations. However, contra Martin (1906), Dalton-Puffer (1996) and

others, these derivations never denote the same. This can only be explained with their different salient meanings and with their diachronic development.

Parallel derivations for ME *-hōd* and *-ship* with a nominal base of the individual-level predicate type are given below:

(19) *hit bytokenyth also our ladyys* ***modyrhode*** *and maydynhede, lightnet*
it betokens also our lady's motherhood and maidenhood, lightened
wyth þe fyre of loue;
with the fire of love
(MIRK,60.1641)

The formation *mōdyrhōde* 'denotes the state of being a mother' (here referring to the virgin Mary), but according to the MED it could also denote 'the state or fact of being a mother'. The ModE formation with *-ship*, however, bears another meaning: 'the duties (function) of a mother' (which can be abandoned). This finding is in line with A & C's assumptions.

(20) *Even had self-defence impelled Claire to abandon her* ***mothership****.*
(1905, Blackw.Mag.Feb 239/1)

To give a brief interim summary, so far it was shown that parallel derivations with different suffixes (-hood, -dom and -ship) never denote the same. What they have in common is that they all build abstract nouns predominantly referring to persons. Since the salient meanings of *hād*, *dōm* and *scipe* (as well as their metonymies) differ, the meanings of parallel derivations also differ. Therefore, these derivations cannot be called rivals in the sense that they are synonymous and compete with each other.

In section 4 I will discuss the paradigmatic nature of the three suffixes in more detail before I will present a lexical-semantic approach of formations with *-hood*, *-dom* and *-ship* that takes into account the diachrony of these elements.

4 The paradigmatic nature of ModE -hood, -dom and –ship derivations

Based on the salient meanings of *hād*, *dōm* and *scipe* in OE, I assume that the three elements were all relational nouns that required an argument (additionally to the referential argument R), and that this property allowed them to become suffixes in the first place. It is also this property that required bases of the type stage-level and individual-level predicate, depending on the salient meanings of *hād*, *dōm* and *scipe*.

The combination of a relational noun requiring a predicate as argument leads to the semantics of the formations with these elements. According to Löbner (1985, 292) relational nouns describe objects that are in a certain relation to other objects. A noun like wife is a relational noun because it describes the relation between a wife and a man the wife is married to. Therefore, relational nouns are used predicatively (wife is a two-place predicate containing the referential argument and the argument for 'wife of N'). Coming back to our three elements *hād*, *dōm* and *scipe*, I will therefore assume that they are two-place predicates:

(21) a. *hād*: predicate (x,y) 'office of N' etc.
b. *dōm*: predicate (x,y) 'authority of N' etc.
c. *scipe*: predicate (x,y) 'state of N' etc.

All three elements contain the referential argument R[5] referring to 'the office, rank, status, person' (*hād*), 'authority, judgement, office, rank' (*dōm*) and 'state, condition, rank' (*scipe*), and the argument that holds the office, that has the authority or renders a judgement, and that is in a state, holds a rank, etc. Since scipe is a deverbal form we would have to assume for this element that it inherited the argument structure from its verb ('to create something') so that scipe originally contained these arguments.

[5] The structure predicate (x,y) could also be described as predicate (R,y) where R stands for referential argument.

It was shown above, that in OE the three elements under investigation had the status of nouns, and since they were relational they required this relation to be fulfilled e.g. in a syntactic phrase as in (22):

(22) *þæt he þær onfenge* ***[ærcebiscopes hāde]***.
that he there received archbishop's office
(Bede_3:21.248.11.2540)

In this example, the genitival modifier ærcebiscopes functions as the argument required by *hād* 'office of N'. Since the three elements have meanings that generally refer to persons (only a person can hold an office or rank or render a judgement), most modifiers found with these elements are nouns denoting persons. These phrases were found quite frequently in OE implying that these relations were highly salient at that time.

Another possibility to express the required relation is a morphological structure where the relational noun acts as morphological head and is modified by a noun or adjective:

(23) *Þa he nolde for his* ***biscop hāde*** *hi aweg adrifan, …*
then he not-wanted for his bishop office they away drive
(GDPref_and_3_[C]:7.188.19.2363)

A compound like *biscophād* satisfies the selectional restrictions, the argument required by the head noun appears as first member of the compound. Semantically, the first member restricts the extension of the meaning of the second member. But why should we assume that the property of being relational (used predicatively) is a prerequisite for elements to become suffixes? Because they require a relation between two elements and therefore they require the co-occurrence of elements. Under these assumptions, the observation that there is an intermediate stage where the elements reaching suffixal status function as second members of compounds seems only natural. As free elements in syntactic phrases *hād*, *dōm* and *scipe* require an argument that appears as preceding modifier, as stems in compounds they also require an argument that appears as preceding modifier on the word-level, so this property is retained throughout the change and still found in ModE formations with *-hood*, *-dom* and *-ship*.

Coming back to the brief sketch of the lexical-semantic development based on the salient meanings of -hood, -dom and -ship (section 1), we can say that the ModE suffixes build a paradigm denoting abstract nouns and display the following semantic differences:

(24) a. *-hood*: [state] 'a distinguishing feature of one's personal nature' (salient meanings 'status, office, rank', metonymies 'state, time')
b. *-dom*: [process] 'possession of the qualities required to do something or get something done' (salient meanings 'authority, judgement', metonymies 'territory, realm')
c. *-ship*: [achievement] 'result of a process of creating' (salient meanings 'created thing, (resultant) state', metonymies 'function, forms of address, skill/art')

The polysemy of the suffixes partly derive from the metonymic shifts that were also briefly mentioned in section 1. For *-hood* the salient meanings 'status, office, rank of N' have been assumed. Since the element mainly occurred (and still occurs) with nouns denoting persons, and since some of these nouns inherently contain the time dimension (as *boy* or *child*), the meaning 'time, period' arose. As concerns *-dom*, the salient meanings 'authority, judgement of N' were assumed and the meanings 'status, territory, realm' can straightforwardly be derived from it via metonymic shifts (if somebody has the authority, she has the power to judge; the authority is generally restricted to a territory or realm, see also above). For *-ship* the salient meanings '(resultant) state, condition of N' were assumed. A certain state can imply an office where duties are performed, and a person holding such an office or rank can be addressed with a title. Moreover, there are bases that denote persons with a certain skill like *penman* or *craftsman*, and these nouns inherently denote a professional agent. In combination with the suffix these nouns may thus denote 'skill/art' of these persons.

The three features [state], [process] and [achievement] also given in (24) include all differences in meaning between the suffixes and result from their diachronic development. They are based on an abstract dimension, the dimension of aspect. These features will be integrated into the analysis proposed in the next section which takes into consideration the diachrony of the suffixes *-hood*, *-dom* and *-ship*.

5. An extension of Lieber's (2004) analysis

Lieber's (2004) lexical-semantic framework provides a systematic way to comprehensively describe the semantics of simplexes and complexes, i.e. including all types of word-formation. Her framework has the following properties: it is decompositional, it is cross-categorial and it accounts for polysemy. It should be noted, that her theory has so far been applied to the synchrony of English word-formation but, as has been shown in this paper, questions raised for the synchrony equally matter to the diachrony. Since she assumes a process during which derivations are built up with their lexical-semantic representations, her approach qualifies well to analyse the development of the suffixes under investigation. Therefore, an extension of Lieber's approach will be proposed to explain and analyse the diachronic aspects of word-formation discussed above. In the following, her main assumptions will be introduced.

First, Lieber assumes that a lexical-semantic representation consists of a Semantic/ Grammatical Skeleton and the Semantic/Pragmatic Body (see also Hovav & Levin 1992 and Mohanan & Mohanan 1999. In line with Jackendoff (1990), she assumes that the skeleton consists of a function and two or more arguments of that function (1) a. As concerns derivation, skeletons are hierarchically layered, i.e., functions can take functions as their arguments (1) b:

(25) a. [F1 ([argument])]
b. [F2 ([argument], [F1 ([argument])])]

The structures in (1) a. and b. show skeletons of morphological complexes that are built up compositionally. Apart from these structures, Lieber proposes a number of features that are contained in the skeleton. Some of these features6[6] are functions and take arguments. All the features proposed by Lieber are binary and privative (present/absent). The most basic categories are those comprising SUBSTANCES/THINGS/ESSENCES and SITUATIONS and are characterised by the features [+/- material] and [+/- dynamic][7]:

> [**+/- material**]: The presence of this feature defines the conceptual category of SUBSTANCES/THINGS/ESSENCES, the notional correspondent of the syntactic category Noun. The positive value denotes the presence of materiality, characterizing concrete nouns. Correspondingly, the negative value denotes the absence of materiality; it defines abstract nouns.
>
> [**+/- dynamic**]: The presence of this feature signals an eventive or situational meaning, and by itself signals the conceptual category of SITUATIONS. The positive value corresponds to an EVENT or process, the negative value to a STATE.

To illustrate the structure (skeleton/body) and the features proposed, let us look at the lexical-semantic representation of the noun *leg*:

(26) leg: [+ material ([ ], [ ]),]

[6] Only those features that will be used for analysing the three su_xes under investigation will be discussed.

[7] Lieber points out (2004, 23) that for her all features have two properties: first, they are equipollent (positive or negative value) and second, they are privative (present or absent; if features are absent they are irrelevant for the item in question.)

The noun *leg* is a concrete noun but since it is a relational noun it has more than one argument (apart from the referential argument, it needs another argument: the object it is a leg of), for example *table* in 'leg of the table').

Two further features, [B] and [CI], are suggested by Lieber (and relevant for my analysis) to capture aspects of quantity of lexical items, especially for derivations building abstract nouns:

> **[B]**: This feature stands for "Bounded". It signals the relevance of intrinsic spatial or temporal boundaries in a SITUATION or SUBSTANCE/ THING/ESSENCE. If the feature [B] is absent, the item may be ontologically bounded or not, but its boundaries are conceptually and/or linguistically irrelevant. If the item bears the feature [+B], it is limited spatially or temporally. If it is [-B] it is without intrinsic limits in time or space.
>
> **[CI]**: This feature stands for "Composed of Individuals". The feature [CI] signals the relevance of spatial or temporal units implied in the meaning of a lexical item. If an item is [+ CI], it is conceived of as being composed of separable similar internal units. If an item is [- CI], then it denotes something which is spatially or temporally homogeneous or internally undiffierentiated.

Lieber further claims that all major categories (i.e. A, N, V) take a referential argument R (see Williams 1981 and Higginbotham 1985) which is the external argument of a noun. So, for example a noun like *Londoner* denoting 'person who lives in London' receives a personal interpretation and has a referential use.

In line with Lieber's argumentation, I assume that OE *hād*, *dōm* and *scipe* have skeletons with two arguments and the feature [-material], since they denote abstract entities. The referential argument (R) bears the index (R_i)[8].

(27) a. *hād*: [-material ([R_i], [ ])]
b. *dōm*: [-material ([R_i], [ ])]
c. *scipe*: [-material ([R_i], [ ])]

As discussed above, Lieber proposes the feature, [dynamic] to distinguish between events and states. She observes that among the class of SUBSTANCES/ THINGS/ ESSENCES there are those which are processual denoting states, events, actions, or even a relation of some kind, and those that lack this processual flavour. As examples for the former type she gives nouns like *author* or *habit* because these nouns have a processual flavour ('writing a book', 'doing something over and over again'). Since *hād*, *dōm* and *scipe* are all relational nouns and all denote states or events and intuitively involve doing something (have an office, have authority (to judge), create something), they are all processual. Therefore, all three elements also contain the feature [dynamic]. The latter feature need not be further specified because the fact that it is present is essential. The two features [-material] and [dynamic] make them similar in meaning.

Moreover, the feature [dynamic] is also inherent in all nouns (and adjectives) that combine with *hād*, *dōm* and *scipe*, because these nouns are required by these elements and are part of doing something (a bishop holds an ofice, a king has the authority to judge, somebody makes friends, etc.). This semantic property of both modifier and modified (in syntactic phrases and compounds) explains why these elements combine so easily in the first place. For nouns and adjectives that occur with *hād*, *dōm* and *scipe*, positive and negative values for the feature [dynamic] are tantamount to the distinction between stage-level and individual-level predicates. Apart from these features, I also assume Lieber's features [+B] and [+CI] for the lexical-semantic representations for the three suffixes under investigation because they predominantly denote animate objects which have the potential to occur singularly or in groups, and hence a collective meaning is always possible (*boyhood, Christendom, readership*, see OED). The

[8] Contra Lieber, I assume that for all word-formations the head is always the name of the function, i.e. the head is always the leftmost element.

diachronic aspect is represented with the features [state], [process] and [achievement] which were motivated in section 4. The list below summarises all the features I assume for the three suffixes:

Features
- [-material]: abstract entity
- [dynamic]: situational meaning (processual _avour)
- [+B]: bounded, limited spatially or temporally
- [+CI]: composed of individuals, composed of units
- [state], [process], [achievement]: abstract features (aspect) resulting from development of the three suffixes

For the Modern English suffixes -hood, -dom and -ship I propose the following lexicalsemantic representations:

(28) a. [-material, dynamic, +B, +CI, state ([Ri], [ ], <base>)
hood
b. [-material, dynamic, +B, +CI, process ([Ri], [ ], <base>)
dom
c. [-material, dynamic, +B, +CI, achievement ([Ri], [ ], <base>)
ship

First, note that the referential argument R is always explicitly marked in the structure, and it is the highest argument in a lexical- semantic representation of a lexical unit. Second, all the features motivated above are included as well as the features [state], [process] and [achievement] which express all differences in meaning found between *-hood, -dom* and *-ship*. Since they result from their diachronic development they can be defined as a diachronic imprint: they represent a distinctive influence on the lexicalsemantics of these elements.

Moreover, the nature of underdetermination attributed to suffixes is also nicely accounted for since they only mark the semantic frame but not a specified concrete semantic outcome. What happens in the process of derivation is that the meanings of the base and the suffix are matched. The meaning of the suffix that best matches the meaning of the base will be the meaning of the derivation.

Apart from the main parts of lexical-semantic representations, Lieber proposes a mechanism for juxtaposing and concatenating simplexes leading to the creation of complex words. By creating a complex word, two referential elements amalgamate into one referential element which is projected into syntax. This mechanism is based on the following principle (Lieber 2004, 61):

(29) Principle of Co-indexation: In a configuration in which semantic skeletons are composed, co-index the highest nonhead argument with the highest (preferably unindexed) head argument. Indexing must be consistent with semantic conditions on the head argument, if any.

Lieber's principle accounts for the fact that a compound has only one referent although it is build up of (at least) two stems with one referent each. In the process of compounding arguments that share indexes also share reference and interpretation, and are linked to a single constituent in the syntactic structure (actually she assumes complete identification of reference). Hence, the principle accounts for the linking of arguments within a lexical semantic structure as well as for the referential properties of complex words. This principle applies to compounding as well as to derivation.

In the following, this mechanism will be applied to the diachronic data: for the first stage where *hcd*, *dōm* and *scipe* occur as phrasal heads I assume the following structure (the several stages will be illustrated with *hād*) :

(30) [-mat.,dyn.,+B,+CI,state ([Ri], [ ])] [+mat.,-dyn.,+B,+CI ([Rj])
hād *bisceopes*

The phrasal head *hād* requires an argument which is satisfied by the genitival modifier in the syntax. The referential argument of *office* stands in a salient relation to the referential argument of *bishop* (in most cases it is the relation that modifies the office). Co-indexation is marked with different indices, i (head) and j (non-head), and in the process of co-indexation the referential argument of the non-head is deleted (this is indicated with underlining the non-head). At this stage in the development, co-indexation takes place at the phrasal level. The next stage of the development is the stage where *had* is the morphological head of a compound. In line with Lieber, I assume the mechanism of co-indexation to explain (complete) identi_cation of reference of the two elements. For the compound *bisceophad* I propose the following representation:

(31) [-mat.,dyn.,+B,+CI, state ([Ri], [ ])] [+mat.,-dyn.,+B,+CI ([Rj])
hād *bisceop*

Two requirements have to be met: a grammatical one and a semantic one. The first one is the requirement of the noun to obtain an argument and to assign a theta-role to that argument. In line with Meyer (1993, 111) this requirement is satisfied here because *bisceop*, functions as argument (office of bishop). Co-indexation of the highest head argument with the highest non-head argument takes place resulting in identification of reference (the lower argument cannot be a candidate for co-indexation since it is the argument required by the noun, and as such the lower argument). Apart from this process, the semantics of *hād* and *bisceop* need to be matched: since an office is an abstract entity pertaining to persons, the argument should be a noun denoting persons. This requirement is also fulfilled. Moreover, both elements contain the feature [dynamic] since they both take part in the process of holding an office, and hence we gain an interpretation where the first element most suitably matches the semantics of the second element[9]. The same applies to compounds with *dōōm* and *scipe*.

According to Lieber, technically the difference between a compound and a suffix is hierarchical ordering, i.e., the difference between juxtaposition and subordination. Since suffixes are bound they are dependent on the base. Nevertheless, they determine all grammatical specifications of the whole element. Lieber claims that they are hierarchically superior to the base, and the lexical semantic structure will show that hierarchy.

Thus, the difference between a compound with had and a derivation with -hood in a lexical-semantic representation would look as follows:

(32) [-mat.,dyn.,+B,+CI,state ([Ri], [ ])] [+mat.,-dyn.,+B,+CI ([Rj])
hād *bisceop*

(33) [-mat.,dyn.,+B,+CI,state ([Ri], [ ] [+mat.,-dyn.,+B,+CI ([Rj])])
hood *monk*

As can be seen from the representations given above, the only difference between the two structures is the hierarchical organisation. This difference is also marked morphologically: juxtaposition is the concatenation of stem + stem, and subordination is the concatenation of a base and a bound element. We could also say that this manifests itself in the requirement of affixes to attach to a base which is part of the lexical entry of affixes. Therefore, there is a correlation of lexical-semantic representation and the morphological shape of the word.

Coming back to the alleged rivalry of suffixes, the lexical-semantic differences listed in (24) are illustrated with the parallel derivations *doghood, dogdom* and *dogship* in ModE. It will be shown that what was said above for formations with nominal bases that denote persons also applies to nominal bases that do not have this property.

(34) a.
[-mat., dyn., +B, +CI, state ([Ri], [ ], [+mat.,-dyn. +B, -CI, state ([Rj])])]
hood *dog*
b.

[9] Of course, speakers may also create a compound like monkey o_ce but here only the selectional restrictions and co-indexation are ful_lled but not the semantic requirement, at least not at its best.

[-mat., dyn., +B, +CI, process ([Ri], [ ], [+mat.,-dyn. +B, -CI, state ([Rj])])]
dom *dog*

c.

[-mat., dyn., +B, +CI, achiev. ([Ri], [ ], [+mat.,-dyn. +B, -CI, state ([Rj])])]
ship *dog*

The noun *dog* clearly does not denote a person but an animate entity, it has the meaning 'a member of the genus Canis'. The only meaning that matches a meaning of *-hood* is 'status' and thus the derivation *doghood* denotes 'the status of a dog'. An example from the OED is: "The world calls it manhood, it is doghood rather". The base also has the potential to denote collectivity: "A lap-dog would be necessarily at a loss in framing to itself the motives and adventures of doghood at large" (OED). As noted above, the difference between *doghood* and *dogdom* lies in the matching of features of *dog* and *dom*. At first sight, it seems that no meaning of -dom matches the meaning of dog, since all meanings imply persons (authority, judgement). What we actually find is the meaning 'world of dogs' (realm of dogs), which is also found for other productively built formations with *-dom*. An example from the OED is "A graduate in horse-management and dogdom". It can be observed that metonymic shifts, being the most recent meanings of derivations, occur more frequently with productively built formations. As concerns the formation *dogship*, the suffix *-ship* derives from the same root as ModE *shape* and originally denoted 'created thing' also including the state this created thing has as a result of the process of creation. On the basis of this meaning, the other meanings given for -ship could develop. Now if this element is attached to a noun like *dog*, the meaning that matches the meaning of the derivation best is 'state of N'. Strikingly, the meaning that does occur is the metonymy 'title', which can be nicely illustrated with the following example from the OED "Yes, when your Dogship's damn'd". This meaning is predominantly found for other productively built formations with *-ship* (it could be assumed that the meaning 'state of being a dog' is blocked by the derivation *doghood*).

The analysis of the parallel derivations *doghood, dogdom* and *dogship* reveals that all meanings that have developed in the course of time are part of the lexical-semantic representation of a suffix. Moreover, the suffix is sensitive to the meaning of the base, regardless of whether the base denotes a person and matches the salient meanings of *-hood*, *-dom* and *-ship*, or whether the base denotes different properties and as a result matches other meanings (mainly metonymies) of the suffixes. Under these assumptions, it is more than evident that parallel derivations are not synonymous, they form a paradigm building abstract nouns with a general meaning 'state/condition of N' but they also bear meanings different from each other, due to metonymic shifts. Thus, although they share meaning, they are semantically different enough to coexist.

6. Conclusion

This paper has dealt with the assumption that parallel derivations built with the suffixes *-hood*, *-dom* and *-ship* are synonymous because they are semanically related. It was shown that if the development of these suffixes is taken into consideration we have to come to the conclusion that they never denote the same. These suffixes are polysemous, they bear the salient meanings they had when they had the status of free morphemes as well as meanings that developed from metonymic shifts. What they have in common is that they all build abstract nouns predominantly referring to persons, and in this respect they build a paradigm. But since their salient meanings as well as their metonymies differ, the meanings of parallel derivations also differ. Therefore, these derivations cannot be called rivals in the sense that they are synonymous and compete with each other. In the analysis which is based on Lieber (2004) and partly on Aronoff & Cho (2001) it was assumed that suffixes bear meaning and that the base is sensitive to their meaning. Thus, the base selects the meaning that matches best the meaning of the suffix implying that suffixes are underspecified. The features [state], [process] and [achievement] can be defined as diachronic imprint of the development of these suffixes which express all the semantic differences between formations with *-hood*, *-dom* and *-ship* and

therefore differences between parallel derivation. This study shows once again that important new insights can be gained if the diachronic perspective is taken into account.

References

Aronoff, Mark (1976): Word formation in generative grammar. Cambridge, Mass.: MIT Press.

Aronoff, Mark; Cho, S. (2001): 'The semantics of -ship suffixation'. Linguistic Inquiry, (32), 167-173.

Carlson, Greg (1977):, Reference to kinds in English. Amherst, Ph.D. thesis, University of Massachussetts.

Dalton-Puffer, Christiane (1996): The French In_uence on Middle English Morphology: A Corpus-based Study of Derivation. Berlin, New York: Moton de Gruyter.

Geist, Ljudmila (2006): Die Kopula und ihre Komplemente. Zur Kompositionalität in Kopulasätzen. Tübingen: Niemeyer (= Linguistische Arbeiten 502).

Higginbotham, James (1985): 'On semantics'. Linguistic Inquiry, (16), 547-594.

Hovav, Malka Rappaport; Levin, Beth (1992): '-ER nominals: implications for the theory of argument structure'. Tim Stowell & E. Werhli, ed., Syntax and Semantics, XXVI: Syntax and the Lexicon. New York: Academic Press, 127-53.

Jackendoff, Ray (1990): Semantic Structures. Cambrige Mass.: MIT Press.

Kratzer, Angelika (1995): 'Stage-level and individual-level predicates'. Gregory Carlson et al., ed., The Generic Book. Chicago: University of Chicago Press.

Kroch, Anthony; Taylor, Ann, ed. (2000): The Penn-Helsinki Parsed Corpus of Middle English, Second Edition (PPCME2). Philadelphia: University of Pennsylvania.

Lieber, Rochelle (2004): Morphology and lexical semantics. Cambridge: Cambridge University Press.

Löbner, Sebastian (1985): 'Definites'. Journal of Semantics, (4), 279-326.

Maienborn, Claudia (2001): Die logische Form von Kopula-Sätzen. Humboldt-Universität zu Berlin.

Martin, Friedrich (1906):, Die produktiven Abstraktsu_xe des Mittelenglischen. Ph.D. thesis, Straÿburg.

Meyer, Ralf (1993): Compound Comprehension in isolation and context. Tübingen:Max Niemeyer Verlag. 18

Mohanan, Tara; Mohanan, K. P. (1999): 'On representations in grammatical semantics'. Tara Mohanann & Lionel Wee, ed., Grammatical semantics: Evidence for Structure in Meaning. Stanford, CA: CSLI publications, 23-76.

Plag, Ingo (2003): Word-Formation in English. Cambridge: University Press.

Taylor, Ann, Warner, Anthony, Pintzuk, Susan; Beths, Frank, ed. (2003): The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE). University of York.

Trips, Carola (2007): Lexical Semantics and diachronic morphology: the development of the derivational suffixes -hood, -dom and -ship in the history of English. Univ. Stuttgart: Habilitationsschrift.

Tschentscher, Christhild (1958):, Geschichte der germanischen Bildungssilbe TUM. Stuttgart, Ph.D. thesis, Universität Stuttgart.

Williams, Edwin (1981): 'Argument Structure and Morphology'. The Linguistic Review, 1, 81-114.

Italian VeV Lexical Constructions*

Francesca Masini *Anna M. Thornton*
Università dell'Aquila *Università Roma Tre*
fmasini@uniroma3.it thornton@cc.univaq.it

1. Introduction

In this paper we will describe a process of lexical enrichment found in Italian, that yields nouns and adjectives taking two verbs as input.

Guevara & Scalise (forthcoming) have shown that verbs are a disfavoured input category for compounds in Indo-European languages (Guevara & Scalise forthcoming: table (25)), and that VV compounds are quite rare in Romance languages (Guevara & Scalise forthcoming: table (32)).

However, some lexical items that have been described as V-V compounds do exist in Italian. At a purely descriptive level, we can recognize at least two groups. The first, in (1), comprises words formed by a sequence of two different verbs; the second, in (2), comprises words formed by the reduplication of a single verb.

(1) **Italian VV compounds**

a.	Persons	*portareca*	lit. carry.bring	'delivery boy'
		pappataci[1]	lit. eat.keep_quiet	'happy cuckold'
b.	Animals	*pappataci*[2]	lit. eat.keep_quiet	'*Phlebotomus papatasi*' (a kind of insect)
c.	Instruments	*saliscendi*	lit. go_up.go_down	'latch'
		toccasana	lit. touch.cure	'cure-all'
d.	Places	*bagnasciuga*	lit. wet.dry	'water-line; foreshore'
		montascendi	lit. go_up.go_down	'sloping path'
e.	Games	*stacciaburatta*	lit. sieve.sieve	'name of a game'
f.	Processes	*dormiveglia*	lit. sleep.wake	'drowse'
g.	Actions	*giravolta*	lit. turn.turn	'pirouette; about face'
		parapiglia	lit. shield.take	'hubbub'

(2) **Italial VV reduplicative compounds**

a	*fuggifuggi*	lit. run_away.run_away	'rush, stampede'
b.	*pigiapigia*	lit. push.push	'rush, stampede'
c.	*copiacopia*	lit. copy.copy	'generalized copying'

Group (1) contains lexemes that belong to many different semantic categories, including action nouns, while group (2) only contains action nouns.

The lexemes in group (1) do not result from a productive lexeme-formation rule; the words listed in (1) nearly exhaust the stock of VV compounds in Italian, and have been formed over a period extending from the 14th century to the first half of the 19th century.

The lexemes in group (2), instead, appear to be the product of a lexeme formation rule, that took off at the middle of the 19th century and is still productive today. This rule, which is described in Thornton (2007b), forms action nouns which name actions performed by a plurality of agents; the input appears to be restricted to verbs with disyllabic stems.

Thornton (2007b) found that both the action nouns in group (1), such as the ones in (1g), and the lexemes in group (2) have originated from the lexicalization of utterances containing a sequence of two verbs in the imperative singular form. This kind of utterance is a stylistic

* This article is the result of close collaboration between the two authors; however, for academic purposes, Francesca Masini is responsible for sections 2.1, 2.2.3, 2.3.4, 3, 5, 6 and Appendix; Anna M. Thornton for sections 2.2.1-2.2.2, 2.3.1-2.3.3, 4. Section 1 is to be ascribed to both authors. Anna M. Thornton's research for this paper has been funded by MIUR, within the PRIN 2005 project *CompoNet*.

device found in most or all Romance languages; it has been studied by Spitzer (1918; 1951-52), who has called it "descriptive imperative". It consists in using a series of two or more imperatives to describe actions, or, in Spitzer's own words:

(3) Descriptive imperative ("impératif descriptif")
"l'action exécutée en fait est présentée sous forme de l'ordre qui aurait été nécessaire dans la situation" [*the action actually performed is presented by means of the order which would have been necessary in the situation*] (Spitzer 1952: 22; our translation)

Spitzer observed that, often, instances of descriptive imperative are commands or exhortations used to direct sailing or battlefield maneuvers. These utterances have then been lexicalized, becoming action nouns that refer to the situation in which it would be appropriate to utter the command, or name the maneuver itself.

Nouns such as the ones in (1g) and (2) do not exhaust the supply of action nouns based on two verbs in the imperative form in contemporary Italian. A third kind exists, exemplified in (4a-b):

(4) **VeV lexical constructions / VeV lexemes**

Action nouns			
a.	[tira e molla]$_N$	lit. pull and let_go	'see-saw, playing fast and loose'
b.	[va e vieni]$_N$	lit. go and come	'coming and going, toing and froing'
Adjectives			
c.	[usa e getta]$_{ADJ}$	lit. use and throw_away	'disposable'
d.	[mordi e fuggi]$_{ADJ}$	lit. bite and run_away	'very quick'
Concrete nouns, *names of games*			
e.	[gratta e vinci]$_N$	lit. scratch and win	'instant scratch lottery / instant scratch lottery ticket'

At the descriptive level, the Action nouns in (4a-b) differ from those in (1g) only because the conjunction *e* 'and' appears between the two verbs. But some of the items in this type originate from descriptive imperatives as much as the lexemes in (1g) and (2). For example, *tira e molla* (4a) was originally a call used on ships to order a specific maneuver, first attested in Venetian and hence passed to several Mediterranean languages (cf. Kahane 1938 on Modern Greek τιραμόλα, τιραμολάρω; Spitzer 1941 on Catalan *tira y amolla*; cf. also Spanish *tira y afloja* attested since 1787 (CORDE)).

A sequence such as *tira e molla* is definitely not a prototypical word: it is spelled as three words, it contains a conjunction (and conjunctions are found as input to lexeme formation processes even less than verbs, if at all: cf. Guevara & Scalise forthcoming), and appears to contain inflected verb forms (although this is not necessarily the only possible analysis, cf. section 2.3.3.2 below). However, items like the ones in (4) fulfill the same task as lexemes created by means of lexeme formation rules or morphological constructions as intended in Booij (2005a,b,c): they enlarge the lexical stock of Italian.

These multiword lexemes have been analyzed in Masini (2006, 2007) as *phrasal words*, i.e., constructions (in the technical sense of Construction Grammar and Construction Morphology, cf. Fillmore, Kay & O'Connor 1988, Goldberg 1995, 2006, Booij 2005a,b,c) which are lexical in nature and function as multiword formation patterns that enrich certain word classes.

In this paper, we will study what we will call the **VeV lexical constructions** or, in short, **VeV lexemes** in contemporary Italian. As can be seen from the examples in (4), VeV lexemes are not limited to action nouns: they also include adjectives (4c-d), concrete nouns and names of games (both exemplified by (4e)).

We will present a sketch of how the lexical constructions in (4) have become independent from the first lexicalized occurrences and have given rise to a new productive type, i.e., – in constructionist terms – how the VeV lexeme construction has become entrenched enough to produce new formations.

Indeed, it is important to realize that by no means all of the VeV lexemes we were able to collect originated as descriptive imperatives: many have been created in a way similar to that used to create neologisms. Therefore, we will start by describing our VeV lexical constructions in the same way as one would describe a lexeme formation rule, looking for constraints on both the input and the output.

2. VeV lexemes

2.1. The corpus

2.1.1. Sources

The corpus on which we based our analysis has been collected from different sources. On the one hand, we used dictionaries and lexicons (GRADIT, DISC, the Eulogos *Lista delle poliremathiche* [List of multi-word expressions]) plus occasional observations (Masini 2007). On the other hand, we used *la Repubblica 1985-2000* corpus (henceforth *laR*), which contains 16 years of issues of the daily newspaper *la Repubblica*, for a total of 380 million tokens (http://sslmitdev-online.sslmit.unibo.it/corpora, cf. Baroni *et alii* 2004). The token count throughout the paper is based only on the corpus *la Repubblica*, to give an idea of the distribution of these forms in a corpus.

2.1.2. Essential quantitative data

Our corpus consists of **66 types**, each of which may display different lexical categories (either noun or adjective, cf. section 2.3.4 for further details) and different spellings (cf. section 2.3.2): *V1 e V2*, *V1-e-V2*, and *V1-V2*[1]. The whole corpus with all this information (the types with their variants and the tokens based on the corpus *la Repubblica*) is listed in the Appendix.

In our corpus, there are few very frequent VeV lexemes, the ones listed in (4): *tira e molla* (cf. 4a) (589 tokens), which is also the oldest item attested[2], *usa e getta* (cf. 4c) (462 tokens), *mordi e fuggi* (cf. 4d) (261 tokens), *gratta e vinci* (cf. 4e) (194) and *va e vieni* (cf. 4b) (97).

Then there is another small group of fairly well-represented items[3]: *taglia e cuci* (28), *tassa e spendi* (17), *sali e scendi* (14), *apri e chiudi* (9), *gratta e sosta* (8), *scappa e fuggi* (7), *prendi e fuggi* (6), *leggi e getta* (6), *tocca e fuggi* (5).

All other types have a frequency of fewer than 5 tokens, and 25 are hapaxes (*accendi e spegni*, *appila e spila*, *apri-e-gusta*, *compra e fuggi*, *compra e scappa*, *compra e vendi*, *consuma-e-getta*, *copri e scopri*, *dai e prendi*, *gratta e spara*, *lascia-e-piglia*, *lava e indossa*, *mangia e bevi*, *metti e butta*, *parla e fuggi*, *porta il riso in cascina e piangi miseria*, *prega e fuggi*, *prendi e scappa*, *prevedi e previeni*, *sbatti e butta*, *scarta e inghiotti*, *scatta (nel senso di foto) e porta a casa*, *tocca e corri*, *tocca e scappa*, *tura e stura*).

The number of verbal bases used in only one lexeme is rather high, but there are also some verbs which tend to occur in more than one type.

V1 types amount to 45. Out of these, 32 occur as V1 in only one lexeme (*accendi*, *appila*, *attacca*, *bacia*, *cerca*, *consuma*, *copia*, *copri*, *corri•*[4], *cuci•*, *dai*, *fotti*, *gioca*, *lascia•*, *lava*, *leggi*, *mordi*, *piglia•*, *porta il riso in cascina*, *prevedi*, *radi*, *sali*, *sbatti*, *scappa•*, *scarta*, *scatta (nel senso di foto)*, *stira*, *tassa*, *togli•*, *tura*, *usa*, *va*), whereas 9 items occur in two types (*apri*, *guarda*, *mangia*, *metti•*, *parla*, *prega*, *taglia*, *tira•*, *vedi*). Then we have a very small group of verbs which occur as V1 in more than two lexemes: *compra•* and *tocca* occur as V1 in 3 lexemes, *prendi•* occurs as V1 in 4 lexemes and *gratta* occurs as V1 in 6 lexemes.

The picture regarding V2 is quite similar. The total number of V2 types is 46. This group displays a slightly higher number of verbs occurring as V2 only once in the corpus of VeV

[1] Henceforth, V1 stands for the first member of a VeV lexeme and V2 for the second member.

[2] *Tocca e sana* is attested as far back as 1812, but nowadays only the variant without conjunction, *toccasana* 'cure-all, panacea', is in use. A diachronic study of the development of this and similar words is necessary, but for reasons of space we must leave it for further research.

[3] The literal translations of all the examples given in this section are to be found in the Appendix.

[4] The items followed by the symbol • occur both as V1 and as V2.

lexemes, i.e., 38 (*ammira*, *bevi*, *chiagni*, *chiudi*, *compra•*, *corri•*, *cuci•*, *gusta*, *impara*, *indossa*, *inghiotti*, *metti•*, *molla*, *passa*, *perdi*, *piangi*, *piglia•*, *prendi•*, *previeni*, *racconta*, *scendi*, *scopri*, *scuci*, *segui*, *sniffa*, *sosta*, *spara*, *spegni*, *spendi*, *spila*, *stacca*, *stura*, *tira•*, *togli•*, *torna a casa*, *trova*, *vendi*, *vieni*). Therefore, only 7 types occur as V2 in more than one lexeme: *butta*, *incolla*, *lascia•*, *stravinci* and *vinci* occur as V2 in 2 lexemes, *scappa* occurs as V2 in 3 lexemes, *getta* occurs as V2 in 7 lexemes, and finally, *fuggi* occurs as V2 in 8 lexemes, being the most frequent V2 base.

Now we will move on to analyze in detail the nature and properties of Italian VeV lexemes. We have looked for phonological, morphological and syntactico-semantic restrictions on the input and the output of these lexical constructions, as is normally done when describing a lexeme formation rule.

2.2. Input

As for the input, we have checked the following properties:

- the number and kind of syllables in the input verb stems (phonology);
- the inflectional class of the input verbs and their morphological complexity (morphology);
- the syntactic and semantic classes of input verbs and the semantic role of the subject of the verbal bases (syntax-semantics).

The results of the analysis of the input are discussed in detail in the following sections and then summarized in section 3 (Table 24).

2.2.1. Phonology

2.2.1.1. Disyllabic bases

As far as phonological restrictions on the input are concerned, we observe a strong tendency towards having verbs with disyllabic bases as input, both in V1 and in V2 (but stronger in V1). Thornton (2007a) has observed a tendency towards disyllabic first members in all sorts of Italian compounding processes. For VN compounds this tendency had already been observed by Dardano (1978: 149) and Ricca (2005). This tendency is stronger in the compounding processes that employ only verbs as bases, such as VV reduplicative compounds and VeV lexemes. To appreciate the strength of this tendency, one must compare the data given for the different sorts of deverbal compounds with the distribution of bases of different lengths in the Italian verbs overall, which is shown in the rightmost column in Table 1.

Table 1 **Length in syllables of the verbal base in several sets of data**

length of verbal base	**V1 in VeV lexemes**	**V2 in VeV lexemes**	**verbs in VV reduplicative compounds***	**verbs as first members of compounds listed in GRADIT****	**verbs in the Italian Basic Vocabulary*****
1 syllable	1,5% (1)	--	--	--	0,3%
2 syllables	90,9% (60)	84,8% (56)	95,7%	82,3%	30%
3 syllables	7,6% (5)	15,2% (10)	4,3%	17,7%	52,8%
4 syllables	--	--	--	--	15,1%
5 syllables	--	--	--	--	1,8%

* Data from Thornton (2007b)
** Data computed on Ricca's (2005) collection
*** Data computed using BDVDB (Thornton, Iacobini & Burani 1997)

In the first members of VeV lexemes, 3 of the 5 trisyllabic bases begin with a vowel. Plénat (1994) has observed that in several prosodically conditioned processes of French, an initial vowel may behave as if it were in some sense "extrametrical", i.e., as if it didn't count in

assessing the number of syllables of a word. Thornton (2007a) has found similar effects of "extramétricité des voyelles initiales" in Italian: most of the trisyllabic verbal first members listed by GRADIT (collected by Ricca 2005) begin with a vowel, and many verbs have two variants, a trisyllabic one with the semantically empty[5] vowel-initial prefix *a-* (which triggers gemination of a following consonant) and a disyllabic one without it. The three vowel-initial first members in our sample (*accendi* 'switch on', *appila* 'plug', *attacca* 'attach') contain this prefix (only etymologically in *accendi*, but still segmentable in *appila* and *attacca*, particularly in the context of the lexemes *ap+pila e s+pila* and *at+tacca e s+tacca*). The other two trisyllabic first members also contain either a prefix (*prevedi* 'foresee') or a string homophonous with a prefix and etymologically corresponding to a prefix (*consuma* 'consume'). All the trisyllabic second members also begin with prefixes or sequences which are derived from prefixes; six of them are also vowel-initial. The complete set of trisyllabic members is given in Table 2.

Table 2 **Trisyllabic members of VeV lexemes**

V1	*accendi*, *appila*, *attacca*, *consuma*, *prevedi*
V2	*incolla* (2 types), *impara*, *indossa*, *inghiotti*, *ammira*, *stravinci* (2 types), *previeni*, *racconta*

2.2.1.2. Bases with heavy penultimates?

Most of the verbs used in both members have a heavy penultimate syllable. To assess whether this is due to a phonological restriction on the input, or is simply a reflection of the distribution of heavy penultimates in Italian verbal bases overall, one should compare the data about the VeV lexemes with data about Italian verbs in general.

Such data are not easily available. As a benchmark of comparison, we computed the percentage of verbs with heavy vs. light penultimate among the disyllabic verbs of the Italian Basic Vocabulary. Table 3 shows some data about the quantity of the penultimate syllable in verbal bases in several sets of data.

Table 3 **Quantity of the penultimate syllable in verbal bases in several sets of data**

quantity of the penultimate syllable	**V1 in VeV lexemes**	**V2 in VeV lexemes**	**verbs in VV reduplicative compounds***	**verbs as first members of compounds listed in GRADIT****	**disyllabic verbs in the Italian Basic Vocabulary*****
heavy (closed)	64,6%	77,3%	75%	65,9%	61%
light (open)	35,4%	22,7%	25%	34,1%	39%

* Data from Thornton (2007b)
** Data computed on Ricca's (2005) collection
*** Data computed using BDVDB (Thornton, Iacobini & Burani 1997)

As the data show, a tendency towards verbal bases with a heavy penultimate is already present in the disyllabic verbs of the Italian Basic Vocabulary; it is slightly stronger in all sorts of verbal bases used in compounding, and it is strongest in the second members of VeV lexemes, while the distribution among the first members of the VeV lexemes is closer to the one expected in the null hypothesis.

In conclusion, the tendency to employ verbal bases with a heavy penultimate cannot be considered a genuine phonological restriction on the input.

[5] Cf. Crocco Galéas & Iacobini (1993).

2.2.2. Morphology

2.2.2.1. Inflectional class of the base

Italian verbs belong to different inflectional classes. It is a matter of some debate whether one should recognize three main classes, based on the three different thematic vowels that appear between the verb's root and the inflectional ending in some forms of the paradigm (such as the second person plural of the present indicative, all forms of the imperfect indicative, and most infinitives (the regular ones)), or just two macroclasses, i.e., the traditional first conjugation and a second class which comprises all other classes and subclasses (this position is taken by Dressler & Thornton 1991). This is not the place to defend one or the other position. In the following analysis, we will adopt the traditional tripartition, to make the analysis comparable to other published analyses and to use categories more familiar to most readers.

Table 4 **Italian verbal inflectional classes (conjugations)**

traditional number	**infinitive ending (citation form)**	**Main characteristics of the class**
1st	*-are*	regular and productive (adapted loanwords and all productive verbal suffixes belong here)
2nd	*-ere*	unproductive and extremely irregular (Dressler *et alii* 2003) (sometimes subdivided in two subclasses, according to whether the stress on the infinitive form falls on the penultimate or the antepenultimate syllable)
3rd	*-ire*	has two subclasses: a minority of verbs do not take the *-isc-* infix; the majority of verbs take the *-isc-* infix in certain cells of the paradigm; this subclass is slightly productive (neologic parasynthetic verbs can be assigned to it) and regular

Table 5 shows the inflectional classes of members of the VeV lexemes, of the verbal elements in other compounding processes of Italian, and of Italian verbs overall.

Table 5 **Inflectional classes of verbal bases in several sets of data**

inflectional class	**verbs in Zingarelli minore***	**verbs marked FO+AU+AD+CO in GRADIT****	**verbal element in VN compounds (Ricca 2005)**	**verbs in VV reduplicative compounds (Thornton 2007b)**	**verbs in first member of VeV lexemes**	**verbs in second member of VeV lexemes**	**16 top rank verbs in LIP and LIF (Giordano & Voghera 2002) *****
-are	79,2%	80%	81,4%	65,2%	66,7% (44)	50% (33)	43,75%
-ere	10,6	9,8%	18,6%	26,1%	25,7% (17)	27,3% (18)	37,50%
-ire	10,2	10,2%		8,7%	7,6% (5)	22,7% (15)	12,50%

* *Zingarelli minore* is a usage dictionary containing about 45.000 types; data computed from more detailed data in Thornton, Iacobini & Burani (1997: 75).
** GRADIT is the biggest Italian usage dictionary. The lemmata marked FO+AU+AD+CO are over 50.000 lexemes in common usage; we have excluded technical or obsolete lexemes listed in the dictionary. Data computed from data in Ricca (2005: 470-471).
*** LIP and LIF are frequency dictionaries of written and spoken Italian, based on a 500.000-token corpus each.

It is immediately apparent that the 1st conjugation, which has by far the largest type-frequency in the overall lexicon of Italian, is underrepresented among verbs used in VV

reduplicative compounding and in VeV lexemes (while verbs used as first members of VN compounds distribute exactly as in the overall lexicon). Conversely, the 2nd conjugation, which is the totally unproductive one, is highly overrepresented in these sets of data. The 3rd conjugation is overrepresented in the second members of VeV lexemes, but this is partly due to the high type-frequency of a single second member, *fuggi* 'run away', appearing in 8 types (12% of the total).

A tentative interpretation of these data is the following: both VV reduplicative compounds and VeV lexemes are closely connected to syntax and discourse, since, originally, they were born as lexicalizations of sentences (cf. 1). In speech and writing, 2nd conjugation verbs are very frequent[6] , and their high token frequency is probably reflected in the high rate in which they appear as members of lexicalized chunks of speech. Therefore, the distribution of inflectional classes of the input verbs corresponds to that found in token frequency in speech and writing rather than to that found in type frequency in the dictionary.

It is also worth noting that the 3rd conjugation verbs that appear as members of VeV lexemes (as well as of VV reduplicative compounds) all belong to the *-isc*-less subclass of this conjugation, which is the smallest subclass and is totally unproductive. The two 3rd conjugation verbs (*sentire* 'feel', *venire* 'come') among the 16 most frequent Italian verbs (cf. Giordano & Voghera 2002: 292, footnote 2) also belong to this subclass. This fact is probably related to the fact that both VV compounds and VeV lexemes are formed almost exclusively from verbs with disyllabic bases, and no *-isc-* verb can have a disyllabic base, as *-isc-* provides a syllable in itself and there are no monosyllabic verb stems in the 3rd conjugation.

In VN compounds, verbs from the *-isc-* subclass do appear as first members. In VN compounds, the tendency to have disyllabic first members is less strong than in constructions with two verbs (cf. Table 1 above), and *-isc-* can be accommodated in two ways, either by actually using an *-isc-* stem, yielding a VN compound with a "long" first member, or by using in compounds an *-isc*-less stem of the verb, the same stem that is used in derivation. The second strategy is somewhat more common, but often both strategies apply to the same compounds, as shown by the data in Table 6.

Table 6 **VN compounds doublets with *-isc-* and *-isc*-less stems**[7]

pulisciorecchie	11	~	*puliorecchie*	35	lit. clean-ears 'cotton swab, Q-tip'
pulisciscarpe	47	~	*puliscarpe*	3	lit. clean-shoes 'shoe-shining machine"
pulisciunghie	16	~	*puliunghie*	1	lit. clean-nails 'tool for cleaning nails"
puliscivetri	38	~	*pulivetri*	45	lit. clean-glass 'window cleaner"

Quite to the contrary, the second strategy would not apply to VeV lexemes (and VV reduplicative compounds), which strictly require the stem homophonous to the imperative (cf. section 2.3.3.2). Therefore, the only way to comply with a restriction demanding disyllabicity of (first) members is to select verbs that already have a disyllabic stem, which excludes *-isc-* verbs: if an *-isc*-verb were to be used in one of these constructions, it would have to appear in its imperative form, which invariably displays *-isc-*, therefore being at least trisyllabic.

2.2.2.2. Morphological complexity of the bases

Verbs used as bases in VeV lexemes can be morphologically complex, but only in certain ways. Suffixed (denominal and deverbal) verbs are excluded, very likely because they would

[6] Giordano & Voghera (2002: 292 footnote 2) list the 21 most frequent verbs in both written and spoken Italian (LIF and LIP corpora; the lists drawn from each of the two corpora coincide in these high ranks). Excluding two auxiliaries (*essere*, *avere*) and three modal verbs (*potere*, *dovere*, *volere*) (which, by the way, all belong to the 2nd conjugation), among the remaining 16 top rank content verbs, 7 (43,75%) belong to the 1st conjugation, 6 (37,5%) to the 2nd conjugation, 2 (12,5%) to the *-isc*-less subclass of the 3rd conjugation, and one, *fare* 'to do /make', the top rank verb, has a mixed conjugation, displaying a few forms belonging to the 1st (e.g., 3rd person singular) but most forms belonging to the 2nd conjugation (e.g., imperfect indicative, present subjunctive, etc.).

[7] The number following the lexeme is its frequency in Google. Data from Thornton (2007a).

invariably have a stem longer than two syllables. Prefixed verbs, on the contrary, are not excluded. There are several kinds of prefixed verbs among members of VeV lexemes (18 in total):

(5) Truly prefixed verbs[8]

prevedi	*prevedere*	'to foresee'	<	*vedere*	'to see'	(V1)
previeni	*prevenire*	'to prevent'	<	*venire*	'to come'	(V2)
stravinci	*stravincere*	'to over-win'	<	*vincere*	'to win'	(V2)
sbatti	*sbattere*	'to fuck'	<	*battere*	'to beat'	(V1)
scopri	*scoprire*	'to uncover'	<	*coprire*	'to cover'	(V2)
scuci	*scucire*	'to unstitch'	<	*cucire*	'to stitch, to sew'	(V2)
stura	*sturare*	'to uncork'	<	*turare*	'to cork'	(V2)

(6) Parasynthetic denominal verbs

incolla	*incollare*	'to stick'	<	*colla*	'glue'	(V2)
scarta	*scartare*	'to unwrap'	<	*carta*	'paper'	(V1)

(7) Verbs formed on verb stems that do not occur by themselves, but only preceded by prefixes (often by at least two prefixes, to form pairs of verbs of opposite meaning)

appila	'plug'	~	*spila*	'unplug'
	**pila(re)*			
attacca	'stick, attach'	~	*stacca*	'detach'
**tacca(re)*				

(8) Verbs which are not synchronically analyzable as prefixed, but that are etymologically prefixed and contain an initial string homophonous with existing prefixes

accendi	'switch on'	(V1)
ammira	'admire'	(V2)
consuma	'consume'	(V1)
impara	'learn'	(V2)
indossa	'wear'	(V2)
inghiotti	'swallow'	(V2)
racconta	'tell, narrate'	(V2)

Among these prefixed verbs, 8 (44,4%) are vowel-initial, and therefore they may not count as trisyllabic bases if we follow Plénat's principle of extrametricality of initial vowels; 5 (27,8%) have the prefix *s-*, which is asyllabic and thus doesn't add a syllable to the base, preserving its disyllabicity. The remaining five prefixes do add a syllable to their base. However, two of these (*prevedi* 'foresee' and *previeni* 'prevent') appear in the same lexeme (*prevedi e previeni*) and create an alliterating effect, whereas another (*stravinci* 'over-win') is the second member of the lexemes *gratta e stravinci* lit. scratch and over-win 'a competition with prizes' and *gioca & stravinci* lit. play and over-win 'a competition with prizes', which are formed by analogy to the very frequent *gratta e vinci* lit. scratch and win 'instant scratch lottery / instant scratch lottery ticket' (*stravincere* is obtained by adding the intensifying prefix *stra-* 'over' to the base *vincere* 'to win'). Further, *consuma* 'consume', in the VeV lexeme *consuma-e-getta* lit. consume and throw away 'disposable', is clearly formed by analogy along the lines of *usa e getta* lit. use and throw away 'disposable'.

Therefore, there is only one true exception, i.e., *racconta* 'tell, narrate'. *Racconta* occurs in *bacia e racconta* lit. kiss and tell, a lexeme which is clearly a calque on English *kiss and tell*. Some other lexemes containing trisyllabic bases are calques: *copia e incolla* < *cut and paste*, *lava e indossa* < *wash and wear*. In sum, trisyllabic bases (all of which contain a prefix in our

[8] In (5) and (6) the forms of prefixed verbs used in VeV lexemes are listed, followed by their citation (infinitival) form and by the citation form of their base.

sample) are peripheral with respect to disyllabic bases, and occur mainly in analogical formations and calques.

2.2.3. Syntax and semantics

2.2.3.1. Syntactic classes

Among the syntactic properties of the verbal bases that form VeV lexemes, we have investigated the kind of syntactic classes involved. The classification of verbs based on argument structure is notoriously problematic. For our current purposes, we have decided to use the classification of Italian verb classes developed by Ježek (2003). We have taken into account three major verb classes[9]:

- **transitive-only** verbs, including verbs with at least two arguments (e.g., *compra* 'buy', *togli* 'remove', *dai* 'give');
- **transitive/intransitive** verbs, including verbs such as *gioca* 'play' (*i bambini giocano nel giardino* 'the children play in the garden' *vs. giocare una partita* 'to play a game'), but also transitive verbs that may present the "null object alternation" (cf. Lo Duca 2000, quoted in Ježek 2003: 98; see also Levin 1993: 33, who speaks of "unspecified object alternation") (e.g., *mangiare* 'to eat': *sto mangiando* 'I am eating' vs. *sto mangiando una mela* 'I am eating an apple');
- **intransitive-only** verbs, including one-argument verbs, either unaccusative (e.g., *corri* 'run', *fuggi* 'run_away') or unergative (e.g., *sosta* 'park/stop', *chiagni* 'cry').

Intransitive (but also transitive) pronominal verbs (e.g. *alzarsi* 'to get up', *arrabbiarsi* 'to get angry') never occur as bases of VeV lexemes. The same can be said of zero-argument verbs including impersonal and atmospheric verbs (e.g. *sembra* 'it seems', *piove* 'it rains'), that do not take a dummy subject in a pro-drop language such as Italian.

Table 7 quantifies the presence of the three syntactic classes in our corpus.

Table 7 **Syntactic classes of the verbal bases in our corpus**

Syntactic verb classes	**V1**	**V2**
transitive-only verbs	41%	36,4%
transitive/intransitive verbs	56%	42,4%
intransitive-only verbs	3%	21,2%

Ježek (2003) claims that, in the database she used, the transitive-only group is the most numerous, even though – she states – this datum might be biased by transitive verbs that present the null object alternation, which are not distinguished from transitive-only verbs in a systematic way. This factor might also be relevant for our analysis. As Table 7 shows, most input verbs for VeV lexemes belong to the transitive/intransitive class and not to the transitive-only class. Indeed, this can be partially biased by the fairly high presence of transitive verbs with null object alternation (e.g. *bevi* 'drink', *leggi* 'read', *mangia* 'eat', *taglia* 'cut'), which are not counted as transitive-only. In addition, the transitive/intransitive class also includes transitive verbs that display an alternation with pronominal verbs with *-si* (e.g. *accendere/accendersi* 'to switch on': *accendi la luce* 'switch on the light' *vs. la luce si accende* 'the light goes on'). However, these verbs should indeed be considered as fully transitive in our case, since their pronominal variant could not occur in VeV lexemes.

As for intransitive-only verbs, they are much more frequent as V2. This is due to the high type frequency of unaccusative verbs such as *fuggi* 'run away' (8 types) and *scappa* 'run away' (3 types) in this position. In this respect, it is worth noting that the two intransitive-only

[9] Ježek (2003) distinguishes 15 classes of verbs. However, three main groups can be identified: transitive-only verbs, intransitive-only verbs (which are then divided into intransitives with the auxiliary *avere* 'to have', named unergative, intransitives with the auxiliary *essere* 'to be', named unaccusative, and pronominal intransitives with the auxiliary *essere* and *-si*) and verbs that display transitive/intransitive alternations of various kinds.

(unaccusative) verbs occurring in V1 position (*sali* 'ascend' and *va* 'go') are followed, in V2 position, by other intransitive (unaccusative) verbs that are semantically opposite, namely *scendi* 'descend' and *vieni* 'come'.

Summing up, zero-argument and pronominal verbs are banned from VeV lexemes. In V1 position there is a remarkable tendency to have intransitive/transitive or transitive verbs (the intransitive-only verbs form an oppositive couple with V2), whereas the three major classes are more equally distributed in V2 position.

2.2.3.2. Semantic role of the subject

In this section we analyze the subject of V1 and V2 from a semantic viewpoint and try to determine if there is any preferential semantic role involved. We are aware that semantic roles are a hotly debated question in contemporary linguistics, and we do not want to enter into the theoretical discussion about them. Therefore, we have based our investigation on the classification given in Blake (1994: 68-71), which is reported in Table 8.

Table 8 **Semantic roles according to Blake (1994: 68-71) (partial)**[10]

Semantic role	Definition	Example
Patient	an entity viewed as existing in a state or undergoing change an entity viewed as located or moving an entity viewed as affected or effected by an entity	***The flame*** *grew bright* ***The stone*** *moved* *The bird sang* ***a song***
Agent	an entity that performs an activity or brings about a change of state	***The robots*** *assembled the car* ***The sun*** *melted the ice*
Instrument	the means by which an activity or change of state is carried out	*She squashed the spider with* ***a slipper***
Experiencer	the creature experiencing an emotion or perception	***They*** *love music*
Beneficiary	the animate entity on whose behalf an activity is carried out	*She did the shopping for* ***her mother***

We have noticed a general tendency to select verbs whose subject has the semantic role of Agent as input to VeV lexemes. This is particularly true of verbs in V1 position, as can be seen from the data in Table 9[11].

Table 9 **Semantic role of the subject of V1 and V2**

Semantic role	V1	V2
Agent	97.8 %	87.0 %
Patient	-	10.9%
Experiencer	2.2 %	2.1 %

The only base in V1 with a role which we can classify as Experiencer is *vedi* 'see'. However, *vedi* in the actual examples of our corpus is used more with the meaning of 'watch' than of 'see', and therefore we may say that the subject of this verb has an Agent-like role.

As for V2, there is again a strong tendency towards the Agent role, but there is a slightly larger variety than for V1: besides one base whose subject has the Experiencer role (*ammira*

[10] Blake's (1994) list of semantic roles is more complete. Table 8 includes only the semantic roles that are relevant for our analysis.

[11] For the sake of completeness, we must add that we also counted as Agent the subjects of motion verbs such as *corri* 'run_away' (V1 and V2), *fuggi* 'run_away' (V1 and V2), *sali* 'ascend' (V1), *scappa* 'run_away' (V1 and V2), *scendi* 'descend' (V2), *torna a casa* 'come back home' (V2), *va* 'go' (V1) and *vieni* 'come' (V2). This apparently clashes with Blake's (1994) classification, in which located or moving entities are viewed as Patients. Notwithstanding this, we believe that the subjects of these verbs are Agents.

'admire'), we also have some subjects with the Patient role (e.g., *perdi* 'lose', *stravinci* 'over-win', *trova* 'find', *vinci* 'win')[12].

2.2.3.3. Aktionsart

In analyzing the actional properties of the verbal bases, we have made use of the classification by Bertinetto (1986) (based on Vendler 1967), since this classification is based on Italian data.

Table 10 **Actional classification adapted from Bertinetto (1986)**

Actional class	**durative**	**telic**	**dynamic**
risultativo (accomplishment)	+	+	+
continuativo (activity)	+	-	+
risultativo-continuativo (accomplishment-activity) [13]	+	±	+
trasformativo (achievement)	-	+	+
puntuale (punctual)	-	-	-
stativo (stative)	+	-	-

The analysis, whose results are displayed in Table 11, shows that all Aktionsart classes are represented in the corpus but one, i.e., stative verbs. This is quite expected if we consider that the verbal bases tend to have an Agent as subject. This seems to be the only strong restriction as far as Aktionsart is concerned, even though punctual verbs are also underrepresented.

Table 11 **Actional classes of V1 and V2**

Actional class	**V1**	**V2**
risultativo (accomplishment)	6.7 %	6.5 %
continuativo (activity)	26.7 %	23.9 %
risultativo-continuativo (accomplishment-activity)	22.2 %	-
trasformativo (achievement)	44.4 %	67.4 %
puntuale (punctual)	-	2.2 %
stativo (stative)	-	-

As for other classes, there are two facts worth mentioning: the higher presence of classes with the [+durative] feature in V1, and, conversely, the higher presence of classes with the [-durative] feature in V2. Once again, this is probably due to the high type-frequency of certain bases in V2 position, and in this case to the numerous types with the structure [V1 *e fuggi* 'run away' / *getta* 'throw (away)'].

2.2.3.4. Semantic classes

As a final investigation of the syntactico-semantic properties of the input, we have looked to see whether we could find any regularities in the semantics of the verbal bases. We have based our analysis on Levin's (1993) classification of verb classes[14]. The categorization of the verbal bases in our corpus according to Levin's classes produced quite fuzzy results.

[12] Ježek (2005: 116, Table 4.5) uses the Italian verb *vincere* 'win' as an example of verb whose Subject has the Beneficiary role, which she defines as the entity that benefits from the event. However, Blake uses a different definition of the role of Beneficiary (cf. Table 8) that does not seem to include the verb *vincere* 'to win' (and *perdere* 'to lose').

[13] This class refers to verbs that can be either telic or atelic depending on whether they are used generically or not (e.g., *ho letto un libro in un giorno* 'I have read a book in one day' vs. *ho letto tutto il giorno* 'I've been reading all day long').

[14] A methodological note is needed here. Of course, Levin's classification was developed for English verbs, so we have to be careful in interpreting the data. However, it is not uncommon to use Levin's classes to categorize verbs belonging to other languages (cf. for instance Xu, Aronoff & Anshen 2007).

As for V1, the verbal bases belong to 22 classes. The bases are well distributed among these classes, so there does not seem to be a predominant class. However, we do notice a tendency to use verbs that denote change, causation and motion: 6 bases belong to the class "Change of state" (labelled as 45 in Levin 1993), 5 to the class Motion (labelled as 51), 4 to the class "Change of possession" (labelled as 13), 3 to the class "Verbs of creation and transformation" (labelled as 26) and 3 to the class "Verbs of removing" (labelled as 10).

The results for V2 are in line with the other analyses carried out: in this position there is less differentiation of classes with respect to V1, since we registered 16 classes only. As with V1, the most represented classes are verbs of change, causation and motion: 10 bases belong to the class "Change of possession" (labelled as 13 in Levin 1993), 9 bases belong to the class "Motion" (labelled as 51)[15], 4 bases belong to the class "Throwing" (labelled as 17), and, finally, 4 belong to the class "Change of state" (labelled as 45).

To conclude, we notice that the tendency to have verbs denoting change, causation and motion as a base for VeV lexemes is in line with the results of the Aktionsart analysis, according to which Achievements are the most represented category, followed by Accomplishments and Activities.

2.3. Output

We proceed now to illustrate the properties of the output VeV lexemes. We have checked the following properties:

- the quality of the stressed vowels, the quantity of the penultimate syllable, the length in syllables and the quality of the word onset for V1 and V2 in each VeV lexeme (phonology);
- the spelling variants of VeV lexemes;
- their inflection and the form of the verbal elements (morphology);
- and, finally, the output lexical categories, the semantic relationship between the two members and whether the two verbal bases share the object and the subject (syntax-semantics).

The results are discussed in detail in the following sections and then summarized in section 3 (Table 25).

2.3.1. Phonology

2.3.1.1. Stressed vowels

Figure 4 shows the distribution of stressed vowels in the two members of the VeV lexemes.

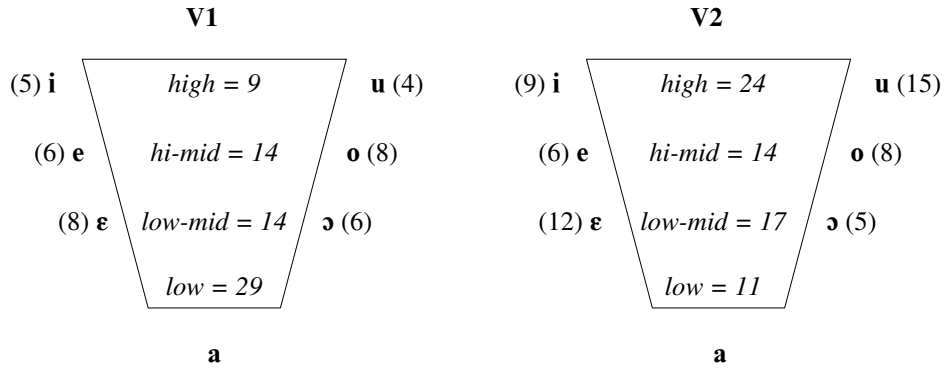


Figure 4 **Stressed vowels in first and second members of VeV lexemes (on all 66 types)**

There is a clear tendency for first members to display a low stressed vowel (the mode is /a/, appearing in 43,9% of the first members) and for second members to display a high stressed

[15] The high incidence of this class is due to the high type frequency of *fuggi* 'run away'.

vowel (the mode is /u/, appearing in 22,7% of the cases – but this figure must be considered carefully, because it derives mostly from the frequent occurrence of a single second member, *fuggi* 'run away', in several different VeV lexemes).

The figures in Figure 4 refer to all members of all the VeV lexemes in the sample, so if a single verb appears in more than one VeV lexeme it has been counted twice.

If we count each verb-base type only once, the results do not change much, as shown in Figure 5: the tendency to have /a/ in the first member remains stable (42,2%), while the tendency to have a high vowel in the second member is slightly reduced (30,4% vs. 36,3%).

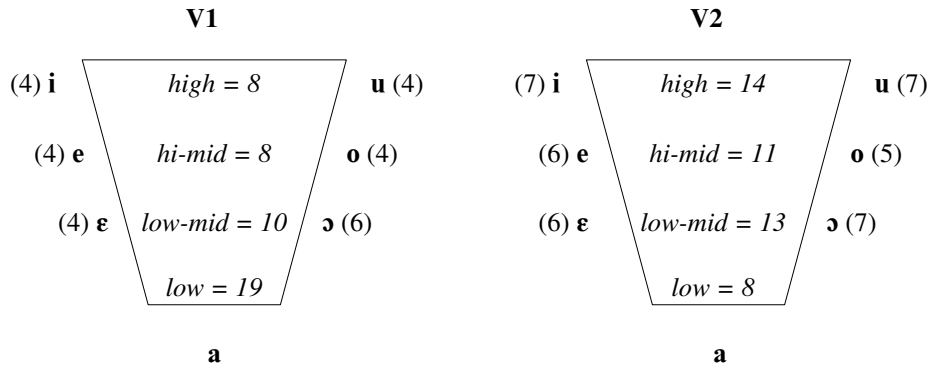


Figure 5 **Stressed vowels in first and second members of VeV lexemes (on member types)**

The raw distribution of stressed vowels in the two members suggests that there might be a tendency to build VeV lexemes with two members differing as much as possible in their stressed vowels[16]. This tendency should be particularly strong in the semantic subclass of VeV lexemes which we call opposites (cf. section 2.3.4). In fact, all lexemes in this semantic class have different vowels (differing mainly for backness and highness) in the two members, except for the five lexemes that are built with the same verb stem preceded by prefixes of opposite meaning, or whose second member is formed from the first one by prefixing the reversative prefix /s/ (e.g., *appila e spila* lit. plug and unplug, *attacca e stacca* lit. attach and detach, *copri e scopri* lit. cover and uncover, *cuci e scuci* lit sew and unsew, *tura e stura* lit. cork and uncork).

Overall, most VeV lexeme have disharmonic stressed vowels, as shown in Table 12.

Table 12 **Disharmonic stressed vowels in VeV lexemes**

same vowel in both members	14	21,2%
different vowels in the two members	52	78,8%

2.3.1.2. Syllabic make-up

We have checked whether the VeV lexemes display harmony in the number of syllables of the two members, and in the weight of their stressed syllabes. The data are shown in Table 13 and Table 14.

[16] This is compatible with the studies on conjoined word pairs (often called "binomials") in general: most analyses of the phonological properties of binomials revealed a tendency towards having words with this or that vowel or consonant in first rather than in second position. For instance, Salvioni (1902: 372), working on Italian alliterative pairs, notes that words with stressed *a* and *o* are found preferably in first position, whereas words with stressed *e*, *i* and *u* are found in second position (e.g., *modo e manièra* 'manner, way', *farne di cotte e di crude* '(be up to) all sorts of tricks'). This finding is replicated in our data. Interestingly, Behagel (1928) found exactly the opposite tendency for German: words with stressed *i* and *u* precede those with stressed *a*.

Table 13 **Penultimate syllable weight harmony?**

Quantity of the penultimate in V1	**Quantity of the penultimate in V2**	**Number of instances**	**Penultimate syllable weight harmony**
no penultimate (open monosyllable)	open	1	68,2% (45) harmonic
open	open	8	
closed	closed	36	
open	closed	15	31,8% (21) disharmonic
closed	open	6	

Table 14 **Syllable length harmony?**

Number of syllables in V1-V2	**Number of instances**
1-2	1 (*va e vieni*)
2-2	51
2-3	9
3-2	4
3-3	1 (*prevedi e previeni*)

As for the weight of the stressed syllable, 45 types (i.e., 68,2%) present a harmonic pattern (closed-closed and open-open), whereas 21 types (i.e., 31,8%) present a disharmonic pattern (closed-open and open-closed).

As for the number of syllables, there is a clear tendency to syllable length harmony: 78,8% of the VeV lexemes in the sample have members of equal length in syllables. In those lexemes in which syllable length is not equal in the two members, in most cases the second member is longer than the first. It should also be noticed that of the four cases with a first member longer than the second, three (*accendi e spegni* lit. switch on and switch off, *appila e spila* lit. plug and unplug, *attacca e stacca* lit. attach and detach) have a vowel initial first member, that could be considered in fact disyllabic by Plénat's principle of extrametricality of initial vowels, and so make the syllable length of the two members balanced.

This datum is in line with the *Gesetz der wachsenden Glieder* ('law of increasing members') formulated by Behagel (1909) for frozen binomials in German, Latin and Ancient Greek (and independently found by Jespersen 1905 for English). Subsequently, Morawski (1927) claimed, on the basis on Spanish data, that the shorter member precedes the longer one only in case of rhyme or alliteration and added that, in rhymed pairs with an equal number of syllables, there are a number of phonetic factors ruling the order of the members. One of these factors requires that words beginning with a vowel or an *h* precede those beginning with a consonant. We investigate word onset harmony in the following section.

2.3.1.3. Word onset harmony?

We checked for harmony or disharmony in the nature of the initial phoneme of the two members. The data are summarized in Table 15.

Table 15 **Word onset harmony**

Word onset in V1 and V2	**Number and percentage of cases**
C initial – C initial	55 (83,3%)
C initial – V initial	6 (9,1%)
V initial – C initial	5 (7,6%)
V initial – V initial	--

Clearly, there is a tendency for both members to start with consonants, and so to start with the same kind of phoneme (C vs. V). However, there are no V initial – V initial lexemes. The items that do not have harmonic word onsets are for the most part disharmonic in other respects

too: 81,8% of the types with different word onsets also differ in syllabic length. All the 6 C initial – V initial lexemes have a disyllabic first member and a trisyllabic second member; 3 of the 5 V initial – C initial lexemes have a trisyllabic first member and a disyllabic second member; the remaining two V initial – C initial lexemes (*apri e chiudi* lit. open and close, *apri-e-gusta* lit. open-and-enjoy) have two disyllabic members.

These data also point to a strong correlation between being vowel-initial and being trisyllabic, which lends further support to Plénat's hypothesis that initial vowels are extrametrical in some respects in prosodically constrained morphological or lexeme formation processes.

Morawski's finding that V-initial words tend to precede C-initial words in Spanish is not replicated in our Italian data.

2.3.1.4. Phonological words

The two members of the VeV lexemes are distinct phonological words according to the tests proposed by Nespor (1993: 173-174): they can both display low-mid vowels, that only appear under main word stress in Italian; the stressed vowels in open penultimates lengthen in both members.

This is expected, as it is normally the case with all sorts of Italian novel native compounds (cf. Peperkamp 1997: 123-129).

2.3.2. Spelling

As our data come from written corpora, a word on spelling is in order. The three words making up the VeV lexemes are usually spelled separately, with a blank between them. An alternative spelling, with hyphens dividing (or connecting?) the three words, is attested in a minority of cases. Table 16 gives figures for the different spellings found in the *la Repubblica* corpus for the most frequent VeV lexemes.

Table 16 **Spelling of the most frequent VeV lexemes in the *la Repubblica* corpus**

tira e molla	537
tira-e-molla	38
tira-molla	16
tiramolla	26
tirammolla	2
tiremmolla	1

usa e getta	413
usa-e-getta	60
usa-getta	0
usagetta	0
useggetta	0

mordi e fuggi	237
mordi-e-fuggi	23
mordi-fuggi	1
mordifuggi	0
mordeffuggi	0

gratta e vinci	190
gratta-e-vinci	4
gratta-vinci	0
grattavinci	0

grattevvinci	0

va e vieni	84
va-e-vieni	13
va-vieni	0
vavvieni	0
vaevieni / vaevvieni	0

Spellings with hyphens are common for other sorts of verb-based compounds in Italian. The VV reduplicative compounds in the *la Repubblica* corpus are spelled both as two separate words and with hyphens, with a ratio of about 50/50 (Thornton 2007b). Hyphenated spellings are also common for VN compounds, particularly low-frequency ones and adjectival ones (Ricca in prep.), although prescriptive norm would have them spelled as a single word with no blanks or hyphens. So, if anything, our data stand out for the low frequency of hyphenated variants. This might be interpreted as pointing to the fact that the separateness and distinctness of the two members must be preserved, even by means of spelling; in the case of VeV lexemes, a hyphen, as opposed to a blank space, is felt as closer to a way of linking than to a way of separating, and is therefore little used in this type.

It can be observed that a few lexemes are also attested without the conjunction *e* between the two members, namely *apri-chiudi* lit. open-close, *cerca-trova* lit. look_for-find, *cuci-scuci* lit. sew-unsew, *mordi-fuggi* lit.bite-run_away, *sali-scendi* lit. ascend-descend, *tassa-spendi* lit. tax-spend, *tira-molla* lit. pull-let_go. However, most lexemes are never attested without the *e*, and never in synthetic spellings. This orthographic characteristic points to the fact that both the conjunction and spelling the two verbs separately are felt as necessary, and so suggests an inherently additive and double nature of the semantics of the type (cf. section 2.3.4).

2.3.3. Morphology

2.3.3.1. Inflection

The VeV lexemes are invariable, both in their nominal and in their adjectival use. Examples (9)-(10) show two couples of nominal VeV lexemes (singular-plural), whereas examples (11)-(12) present VeV lexemes in adjectival function accompanying both singular and plural nominal heads. No lexeme presents an overt plural marker[17].

(9) a. *…non comprerò mai più* ***un gratta e sosta*** *…*
…not buy.1SG.FUT never anymore a scratch and park
'… I will never buy a scratch and park ticket again …'

b. *…sono già due volte che vengo multato,*
…be.3PL.PRES already two times that come.1SG.PRES fine.PTCP.PST.M,
nonostante abbia pagato ***i gratta e sosta****…*
although have.1SG.SBJV.PRES pay.PTCP.PST the.PL scratch and park
'… I have been fined already twice, although I had paid the scratch and park tickets …'

(10) a. *… dopo* ***un estenuant-e tira e molla*** *…*
… after an exhausting-SG pull and leave …
'…after an exhausting see-saw …'

b. *… dopo* ***estenuant-i tira e molla*** *…*
… after exhausting-PL pull and leave …
'… after exhausting see-saws …'

(11) a. *… dopo* ***un trattament-o "stira e ammira"*** *…*
… after a treatment-SG iron and admire …
'…after a perfect ironing treatment …'

[17] All examples are taken from *la Repubblica* corpus.

b. … *figli* *di* ***casalingh-e*** ***"stira*** ***e*** ***ammira"***…
… children of housewife-PL iron and admire …
'… children of perfect housewives …'

(12) a. … *si chiama* ***"bottigli-a*** ***usa*** ***e*** ***getta"*** …
…call.3SG.PRES.PASS bottle-SG use and throw_away …
'…it is called "disposable bottle"…'

b. … *produrremo* *solo* ***bottigli-e*** ***usa*** ***e*** ***getta*** …
…produce.1PL.FUT only bottle.PL use and throw_away …
'… we will produce only disposable bottles …'

Following Hopper & Thompson (1984), the invariability of VeV lexemes, i.e., the lack of morphological markers typical of nouns and adjectives, denotes the low prototypicality of these forms as nouns and adjectives.

2.3.3.2. The form of the verbal element

As in all other verb-based compounds of Italian, the verbs in VeV lexemes appear in a form which is homophonous to the 2nd person singular imperative. Table 17 shows that different hypotheses (both 3rd person singular and the "Scalisian" stem, i.e., the verb stem as defined by Scalise 1983, that is, the infinitive form of the verb without its inflectional marker *-re*) fail to yield the correct output in most of the cases (the bases of the lexemes in the first column are representative of different inflectional classes).

Table 17 **The form of the verbal bases**

VeV lexeme	**imperative**	**3rd ps.sg.**	**Scalisian stem**	**input V citation form (infinitive)**
mordi e fuggi	***mordi!***	***morde***	***morde-***	*mordere*
	fuggi!	***fugge***	***fuggi-***	*fuggire*
gratta e vinci	***gratta!***	***gratta***	***gratta-***	*grattare*
	vinci!	***vince***	***vince-***	*vincere*
va e vieni (also *vai e vieni*)	***va!/ vai!***	***va***	***anda-***	*andare*
	vieni!	***viene***	***veni-***	*venire*
mangia e bevi	***mangia!***	***mangia***	***mangia-***	*mangiare*
	bevi!	***beve***	***be-***	*bere*
corri e tira	***corri!***	***corre***	***corre-***	*correre*
	tira!	***tira***	***tira-***	*tirare*
cuci e scuci	***cuci!***	***cuce***	***cuci-***	*cucire*
	scuci!	***scuce***	***scuci-***	*scucire*

The question of course is why this homophony between the verbal members of VeV lexemes and imperatives holds. Romance verb-based compounds have been an object of investigation for almost two centuries. Research has focused mainly on VN compounds, while compounds containing two verbs have been less studied, probably because they are much less numerous than VN ones (Thornton 2007b).

Scholars have been divided into three main parties over the nature of the verbal element in VN compounds:

a) those maintaining that the verbal element is the imperative, as diachronic evidence shows that the compounds originate in imperative sentences;
b) those maintaining that the verbal element is the 3rd person singular of the present indicative (a minority, including the italian scholars Tollemache and Merlo);
c) those maintaining that the verbal element is a verb stem (Pagliaro).

The whole question was settled about half a century ago, with balanced positions such as the ones expressed by Migliorini and Folena and quoted in (13):

(13) a. "Formalmente, si tratta d'imperativi; concettualmente, oggi questo carattere è molto meno avvertito." [*Formally*, *they are imperatives; conceptually*, *today this property is much less felt*] (Migliorini 1957[2]: 82 [1935: 42]; our translation)
b. "la valutazione sincronico-semantica va […] distinta da quella diacronico-etimologica" [*the synchronic-semantic analysis is to be distinguished from the diachronic-etymological one*] (Folena 1958: 104; our translation)

We share the views expressed in (13), especially in the light of the diachronic data reported in section 1.: as we have seen, all VV Italian lexical constructions, including VeV lexemes, originate in discourse from a particular stylistic device that Spitzer (1951-1952; cf. also Spitzer 1918) named "impératif descriptif" (descriptive imperative). Therefore, the homophony between the verbal elements in VeV lexemes and imperatives is due to the fact that the earliest attested lexemes are in fact lexicalizations of imperative sentences. When the construction became entrenched enough to behave as a more abstract pattern for the creation of new lexemes (without a discoursal source), the construction retained a restriction to the effect that the verb base used in it is homophonous with the singular imperative. To characterize the morphological make-up of the output of the currently productive construction, we could say that it employs the verb stem used in compounding, as defined in works such as Rainer (2001) and Thornton (2005: 157-160), i.e., a specific morphomic stem which has a place in Italian morphology (inflection and lexeme-formation)[18].

It is worth noting that postulating that the VeV lexeme construction makes use of a stem like other compounding phenomena is not trivial, since VeV lexemes, being phrasal words, should be made up of free forms. On the other hand, the compounding stem is always completely homophonous with the imperative singular free form, which is the verb form at the origin of VeV lexical construction, so it is impossible to decide whether these constructions display the compounding stem or the imperative form.

2.3.4. Syntax and semantics

2.3.4.1. Subject and object

As a first syntactic property, we have checked whether the main arguments (i.e., subject and direct object) of the verbal bases are shared. Of course, the calculation for the object involved only pairs with transitive verbs (cf. section 2.2.3.1).

As shown in Table 18, the large majority of VeV lexemes are formed by verbs that share the same object (e.g., *apri e chiudi* lit. open and close, *copia e incolla* lit. copy and paste, *leggi e getta* lit. read and throw_away). However, there is also a remarkable number of cases that present distinct objects, such as for instance *radi e getta* lit. shave and throw_away, *tassa e spendi* lit. tax and spend, *gratta e vinci* lit. scratch and win[19]. Finally, there are a few lexemes in which the two verbs may but also may not share the same object, e.g., *compra-e-vendi* lit. buy and sell, *dai e prendi* lit. give and take. It is worth noting that these two verbs belong to a specific class of opposites that Cruse (1986: 234) calls "indirect converses".

Table 18 **Shared objects**

Shared object	%
+	71.0 %
±	4.5 %
-	24.5 %

[18] The form of the verbal element used in VeV lexemes is homophonous to the imperative in a stricter way than for verb-based compounds, such as VN compounds, where the morphome used in derivation may sometimes be preferred to the one used in compounding to achieve disyllabicity of the compound's first member, as the data in Table 6 above show.

[19] We should add that the pairs that do not share the object are mostly made up of one transitive verb and one transitive/intransitive verb. Since the latter may simply be read as intransitive, we may say that there is actually a strong tendency to share the direct object.

In conclusion, the shared object does not seem to be a restriction for the formation of VeV lexemes, but rather a mere tendency. On the contrary, the tendency to share the subject is much stronger. Data are given in Table 19 (based on the whole corpus).

Table 19 **Shared subjects**

Shared subject	**%**
+	74.25 %
±	25.75 %
-	0 %

V1 and V2 definitely tend to share the same subject: in the examples in (14), the person who buys and runs away is the same person, and the person who speaks and throws away is the same person. There are lexemes in which one can have a disjunct subject interpretation (as in (15), where a reading is possible in which there are two people performing the actions alternatively, or different people V1-ing and V2-ing at different times), but in those very cases it is also possible to have a shared subject interpretation. A disjunct subject interpretation only is never possible.

(14) Shared subject
a. … *epoca* *del* *compra* *e* *fuggi* …
… age of.the buy and run_away
'… age of frenetic and superficial buying …'
b. … *cellulari* *parla* *e* *getta* …
… mobile_phones speak and throw_away
'disposable mobile phones'

(15) Optionally shared subject
a. … *continuo* *accendi* *e* *spegni* …
… continual switch_on and switch_off
'a continual switching on and switching off'
b. … *poltroncine* *apri* *e* *chiudi* …
… armchairs:DIM open and close
'back stalls that fold up'

In conclusion, the shared subject condition seems to be a strong restriction on the formation of VeV lexemes. In this respect, it will be useful to remember that the verbal bases forming the lexemes preferably have subjects with the semantic role of Agent (cf. section 2.2.3.2). Therefore, so far we can say that VeV lexemes are made up of two verbal bases which share an agentive subject.

2.3.4.2. Lexical categories of the output

The first feature we explore is the kind of lexical category of the output. VeV lexemes are either nouns or adjectives.

Some adjectival (Adj) VeV lexemes are presented in (16).

(16) **Adjectives**

a.	*leggi e getta*	lit. read and throw_away	'disposable (said of books)'
b.	*lava e indossa*	lit. wash and wear	'wash and wear'

Nominal VeV lexemes belong to different subclasses. Action nouns (henceforth Naction) denote an event that is described by the (sequence of the) two input verbs (cf. (17)).

(17) **Action nouns**

a.	*dai e prendi*	lit. give and take	'giving and taking'
b.	*compra e fuggi*	lit. buy and run_away	'a frantic and quick buying'

Concrete nouns (henceforth Nconcrete) denote an inanimate entity that has some role in the event that is described by one or both of the input verbs: in most cases, the noun denotes a patient of the action described by both verbs (as in the examples in (18a-f)); in a few cases the noun denotes the patient of V1 where V2 is intransitive (18g) or intransitively used (18h); in a smaller number of cases (those in (19)), the noun denotes the instrument used to perform the actions described by both verbs or by V1 only; in a couple of cases the noun indicates the place where the action denoted by the two verbs takes place (as in (20)); three of these concrete nouns are proper names (a magazine (20a), a restaurant (20b), and a product (19d)):

(18) **Concrete nouns** (patient)

a.	*consuma-e-getta*	lit. consume and throw_away	'disposable item'
b.	*leggi-e-getta*	lit. read and throw_away	'worthless book'
c.	*vedi e getta*	lit. see and throw_away	'worthless movie'
d.	*usa e getta*	lit. use and throw_away	'disposable item'
e.	*gratta e sniffa*	lit. scratch and sniff	'scratch and sniff card/sticker'
f.	*mangia e bevi*	lit. eat and drink	'a sort of sundae'
g.	*gratta e sosta*	lit. scratch and park	'scratch and park ticket'
h.	*gratta e vinci*	lit. scratch and win	'instant scratch lottery ticket'

(19) **Concrete nouns** (instrument)

a.	*copri e scopri*	lit. cover and uncover	'convertible truck roof'
b.	*radi e getta*	lit. shave and throw_away	'disposable razor'
c.	*parla e getta*	lit. talk and throw_away	'disposable mobile phone'
d.	*stira e ammira*®	lit. iron and admire	'name of an ironing starch marketed by Johnson'

(20) **Concrete nouns** (place)

a. *cerca-trova* – lit. look_for-find – as name of a magazine that publishes only classified ads

b. *mordi e fuggi* – lit. bite and run_away – as name of a fast-food restaurant

Finally, names of games (henceforth Ngame) are a kind of action nouns; they may be also ascribed to proper names, like some of the concrete nouns mentioned above. In half of these nouns V1 and V2 describe actions to be performed in playing the game (e.g., (21a)), in the other half V1 describes an action to be performed to play and V2 describes the outcome (e.g., (21b)).

(21) **Names of games**

a.	*mangia e passa*	lit. eat and pass	'Double Bughouse (a variant of the game of chess)'
b.	*gratta e vinci*	lit. scratch and win	'instant scratch lottery'

Out of the 66 types, 32 (48.5%) display only one output category, with the following distribution:

- action nouns = 16 (*accendi e spegni*, *appila e spila*, *compra e fuggi*, *compra-e-vendi*, *cuci e scuci*, *dai e prendi*, *fotti e chiagni*, *lascia-e-piglia*, *metti e togli*, *parla e fuggi*, *piglia e lascia*, *porta il riso in cascina e piangi miseria*[20], *prega e segui*, *tocca e corri*, *tura e stura*, *va e vieni*);

[20] In our corpus, *porta il riso in cascina e piangi miseria* is the gloss of *fotti e chiagni*; therefore, the former has been categorized like the latter, i.e., as an action noun.

- concrete nouns = 1 (*consuma-e-getta*)[21];
- names of games = 4 (*gioca & stravinci*, *gratta e stravinci*, *mangia e passa*, *tira e vinci*);
- adjectives = 11 (*apri-e-gusta*, *attacca e stacca*, *compra e scappa*, *gratta e spara*, *guarda e compra*, *lava e indossa*, *metti e butta*, *prevedi e previeni*, *sbatti-e-butta*, *scarta e inghiotti*, *scatta (nel senso di foto) e torna a casa*).

The remaining 34 types present more than one category. The combinations found are the following:

- adjectives/action nouns = 15 (*apri e chiudi*, *bacia e racconta*, *copia e incolla*, *corri e tira*, *guarda e impara*, *prega e fuggi*, *prendi e fuggi*, *prendi e getta*, *scappa e fuggi*, *taglia e cuci*, *taglia e incolla*, *tassa e spendi*, *tocca e fuggi*, *togli e metti*, *vedi e fuggi*);
- adjectives/concrete nouns = 6 (*gratta e sniffa*, *leggi-e-getta*, *parla e getta*, *radi e getta*, *stira e ammira*, *vedi e getta*);
- adjectives/concrete nouns/action nouns = 6 (*copri e scopri*, *gratta e sosta*, *mangia e bevi*, *mordi e fuggi*, *sali e scendi*, *usa e getta*);
- adjective/action nouns/names of games = 4 (*prendi e lascia*, *prendi e scappa*, *tira e molla*, *tocca e scappa*);
- concrete nouns/action nouns = 1 (*cerca e trova*);
- adjective/names of games = 1 (*gratta e perdi*);
- adjectives/concrete nouns/action nouns/names of games = 1 (*gratta e vinci*).

If we split the total number of types (66) per category, we obtain 112 types. Table 20 shows the percentages of the represented categories with respect to this figure, whereas Table 21 refers to the tokens. The picture is quite similar: the most represented categories are adjectives and action nouns.

Table 20 **Lexical categories represented in VeV lexemes (types)**

Categories		**(n) %**	
adjectives			(44) 39.3%
nouns	action	(68) 60,7%	(43) 38.4%
	concrete		(15) 13.4%
	game		(10) 8.9%

Table 21 **Lexical categories represented in VeV lexemes (tokens in *laR*)**

Categories		**(n) %**	
adjectives			(584) 32.8%
nouns	action	(1197) 67.2%	(997) 56.0%
	concrete		(54) 3.0%
	game		(146) 8.2%

2.3.4.3. Semantic relationship between the members

In this section we explore the semantic relationship between V1 and V2. From our analysis, it turns out that V1 and V2 entertain two types of semantic relationships, namely:

- **lexico-semantic** relationships *strictu sensu* (cf. Cruse 1986, Ježek 2005), i.e., V1 and V2 may be opposites (**OPP**) (of the kind called "reversive" by Cruse 1986: 226) (e.g., (22a)), synonyms (**SYN**) (e.g., (22b)), co-hyponyms (**CO-HYPO**) (e.g., (22c)), or, finally, have an intrinsic cause-effect relation (**CONSEQ**) (e.g., (22d));

[21] In the only context in which it occurs ([…] *in epoca di* consuma-e-getta *e di film in tv ridotti a panini imbottiti di pubblicità* […] 'in an age of consume-and-throw away and of TV movies turned into sandwiches filled with commercials'), this lexeme denotes a concrete noun referring to commercial products.

(22) a. *apri e chiudi* lit. open and close
b. *scappa e fuggi* lit. run_away and run_away
c. *mangia e bevi* lit. eat and drink
d. *cerca e trova (lavoro)* lit. look_for and find (job)

- **frame-like** relationships: in this case the two verbs share the same frame (in the Fillmorean sense), and often define it. For instance, in example (23a), *grattare* 'scratch' and *sostare* 'park' are not linked by a particular lexico-semantic relationship, but are linked at an encyclopedic level since they form a well-defined complex event which is part of the shared knowledge of a community, i.e., the act of buying and scratching a special ticket so as to be authorized to park a car for a particular length of time. The same holds for (23b), where unwrapping and swallowing are two sequential events that belong to the frame of EATING or FOOD: some foods are packaged, so in order to eat them one needs to unwrap the packaging first.

(23) a. *gratta e sosta* lit. scratch and park 'scratch and park ticket'
b. *scarta e inghiotti* lit. unwrap and swallow 'easy to unpack (snack)'

In this last group we have a number of subtypes:

- sequential (**SEQ**): this group contains a V1 and a V2 that simply refer to events that occur one after the other, like (23) above.
- aspectual (**ASP**): this group displays two specific verbal bases as V2, namely *getta* and *butta* (both meaning 'throw (away)'); rather than identifying a proper second phase of the complex event denoted by the whole lexeme, these verbal bases have more the effect of modifying the V1 event; more precisely, they convey an aspectual meaning of "semelfactivity" of the V1 event, which therefore cannot be reiterated (for instance *leggi e getta* – in (24) – denotes an editorial product which can be read – or is worth reading – only once); these lexemes refer to different kinds of disposable (hence, often, low-quality, low-price) commercial products;

(24) *leggi e getta* lit. read and throw away 'low-quality, low-price editorial product'

- manner (**MAN**): quite similarly, this group also presents two specific verbal bases as V2 – *fuggi* and *scappa* (both meaning 'run away') – that have the function of modifying the manner in which the V1 action is accomplished: in this case, V2 conveys an adverb-like meaning of "in a hurry" and "superficially" (see for example *mordi e fuggi* in (25), which, used as an adjective, often refers to a kind of frantic and superficial tourism);

(25) *mordi e fuggi* lit. bite and run_away '(too) quick and hasty'

Table 22 shows the number of types and tokens per semantic relationship.

Table 22 **Semantic relationships between V1 and V2**

semantic relationships		**abbreviations**	**number of types**	**number of tokens (*laR*)**
lexico-semantic	opposites	OPP	18	755
	cause-effect	CONSEQ	8	201
	cohyponyms	CO-HYPO	1	1
	synonyms	SYN	1	7

frame-like	sequential	SEQ	18	55
	manner	MAN	11	282
	aspectual	ASP	9	480

As can be easily seen, there are five main categories (opposites, cause-effect, sequential, manner and aspectual), whereas synonyms and cohyponyms are isolated exceptions, both in terms of types and tokens.

If we compare the semantic groups just outlined and the type of verbal classes (of various kinds) involved in each group, we note that there is a tendency towards dishomogeneity. However, there are some regularities, since the verbs in V2 position of the ASP and MAN classes obviously display the very same features: in the ASP class, V2 all belong to the semantic class "Throwing" (labelled as 17), to the syntactic class "transitive" and to the Aktionsart class "Achievement"; in the MAN class, V2 all belong to the semantic class "Motion" (labelled as 51), to the syntactic class "intransitive" (and more precisely "unaccusative") and to the Aktionsart class "Achievement" (cf. also section 2.2.3.3).

As for lexical categories, as shown in Table 23, almost all semantic classes are in line with the figures presented above for the whole corpus (cf. Table 20), i.e., in almost all classes there is a majority of adjectives and action nouns.

Table 23 **Semantic relationships and lexical categories**

semantic classes		**lexical categories (types per category, total 112)**			
		Adj	**Naction**	**Nconcrete**	**Ngame**
opposites	(OPP)	8	17	2	2
cause-effect	(CONSEQ)	4	3	2	5
cohyponyms	(CO-HYPO)	1	1	1	-
synonyms	(SYN)	1	1	-	-
sequential	(SEQ)	13	10	3	1
manner	(MAN)	9	9	1	2
aspectual	(ASP)	8	2	6	-

No exclusive relationship is found between semantic classes and lexical categories. All semantic classes display more than one lexical category and most display all of them. However, there are some interesting correlations that are worth mentioning, namely:

- most names of games are found in the CONSEQ group; the latter also displays another peculiar feature: it is the only class in which the semantic roles of the subjects of V1 and V2 do not coincide. As mentioned in section 2.2.3.2, there is a strong tendency towards the Agent role, but in V2 position one may also find verbs whose subjects have the role of Patient. These verbs appear to gather in the CONSEQ class, which contains in V2 position verbs such as *vinci* 'win' or *perdi* 'lose';
- the ASP group contains a rather high number of concrete nouns; this is due to a nominalization process that has the adjectives of the same group as a base: for instance, *usa e getta* used as an adjective (as in *un rasoio usa e getta* 'a disposable razor') is then nominalized by dropping the nominal head: *un usa e getta* 'a disposable [razor]' (cf. Thornton 2004 for an analysis of this nominalization process in Italian).

3. Summing up: input and output

In the preceding sections we have analyzed the phonological, morphological, syntactic and semantic properties of VeV lexemes and of their input verbs.

The results of our investigation are summarized in the following tables. As can be seen, VeV lexemes display certain tendencies and a number of rather strong restrictions. Besides, they can be classified in a number of well identified semantic classes. In the following section we will analyze in more detail how these classes have arisen by means of a corpus-based investigation.

Table 24 **Input: constraints and tendencies in VeV lexeme formation**

level	**constraints**
phonology	the input verbs should be **disyllabic** (cf. section 2.2.1, Tables 1 and 2)
morphology	the distribution of **inflectional classes** of the input verbs corresponds to that found in token frequency in speech and writing, not to that found in type frequency in the lexicon (cf. section 2.2.2.1, Tables 4 and 5)
	verbs taking the ***-isc-*** suffix are **banned** (cf. section 2.2.2.1)
	the input verbs may be **prefixed** but not suffixed (cf. section 2.2.2.2)
syntax and semantics	zero-argument and pronominal verbs are banned from VeV lexemes; there is a high tendency to have intransitive/transitive or transitive verbs in V1 position (cf. section 2.2.3.1, Table 7)
	the input verbs tend to have subjects with the semantic role of **Agent** (cf. section 2.2.3.2, Tables 8 and 9)
	stative verbs are **banned** (cf. section 2.2.3.3, Table 11)
	the input verbs mostly denote **change, causation and motion** (analysis based on Levin's 1993 classes) (cf. section 2.2.3.4)

Table 25 **Output: constraints and tendencies in VeV lexeme formation**

level	**constraints**
phonology	V1 and V2 tend to have maximally **different stressed vowels** (cf. section 2.3.1.1, Table 12, Figures 1 and 2)
	V1 and V2 tend to have **equal stressed syllables of equal weight** (cf. section 2.3.1.2, Table 13)
	V1 and V2 tend to display **syllable length harmony**; if not, they tend to comply with Behagel's *Gesetz der wachsenden Glieder* ('law of increasing members') (cf. section 2.3.1.2, Table 14)
	V1 and V2 form **distinct phonological words** (cf. section 2.3.1.4)
morphology	VeV lexemes are **invariable** (cf. section 2.3.3.1)
	the verbal bases are **homophonous to the imperative** (cf. section 2.3.3.2, Table 17)
syntax and semantics	V1 and V2 **share the subject** (cf. section 2.3.4.1, Table 19)
	the output lexical categories are **noun** and/or **adjective** (cf. section 2.3.4.2, Tables 20 and 21)
	V1 and V2 are linked by a **lexico-semantic relationship** or share the same **frame** (cf. Tables 22 and 23)

4. VeV lexemes in use

In order to reach a better understanding of the role of the VeV semantic classes identified above, and of how the different semantic classes have arisen, we carried out a corpus-based quantitative and qualitative analysis of all VeV lexemes in our corpus.

As for the quantitative analysis, the results based on the *la Repubblica* corpus give us insight into the matter. The distribution of the tokens per each type is indeed quite telling. As Table 26 shows, for almost all proper semantic classes (i.e., opposites, cause-effect, manner and aspectual), there is a "leader" in terms of tokens, the type that emerges due to its remarkably higher number of tokens with respect to the other items in the same group. The only exception to this generalization is the sequential class.

Table 26 **Types, tokens, leaders and hapaxes for each semantic class**

semantic classes		**number of types**	**number of tokens (*laR*)**	**leader of the group**	**% of leaders (tokens)**	**number of hapaxes (*laR*)**
opposites	OPP	18	755	*tira e molla* (589 tokens) *va e vieni* (97 tokens)	78.00% 12.80% (90.8%)	7

cause-effect	CONSEQ	8	201	*gratta e vinci* (194 tokens)	96.50%	1
cohyponyms	CO-HYPO	1 (*mangia e bevi*)	1	-		1
synonyms	SYN	1 (*scappa e fuggi*)	7	-		-
sequentials	SEQ	18	55	-		6
manner	MAN	11	282	*mordi e fuggi* (261 tokens)	92.60%	7
aspectual	ASP	9	480	*usa e getta* (462 tokens)	96.25%	3

In what follows, we set forth the results of our qualitative analysis of the occurrences of VeV lexemes in the *la Repubblica* corpus. These results show that each of the leader lexemes is used in one or a few specific frames, and that new lexemes are formed that fit into the same frame and differ only slightly from the leader.

4.1. Manner (MAN) class: *mordi e fuggi*

The Manner class is led by the VeV *mordi e fuggi*:

(26) *mordi e fuggi* lit. bite and run_away
Adj = '(too) quick, hasty'
Naction = 'quick operation' [in some specific domain]

Of a total of 261 tokens of *mordi e fuggi* in the *laR* corpus, over 30% are adjectives. We have checked what kinds of nouns *mordi e fuggi* as an adjective modifies. The overwhelming majority of these nouns belong to the frame of TOURISM. The highest token-frequency is found with *turismo* 'tourism' (13 tokens):

(27) *turismo* 'tourism' modified by *mordi e fuggi* (13 tokens)
a. *nella stragrande maggioranza dei casi, si tratta di un turismo mordi e fuggi che riempie le spiagge e lascia deserte le altre strutture.*
'in the great majority of cases, it is a bite-and-run-away tourism, which fills the beaches but leaves other places empty'
b. *Agrigento, le cui aspettative sono concentrate su un turismo meno "mordi e fuggi"*
'Agrigento, whose hopes focus on a less "bite and run away" tourism'
c. *il turismo d'affari e quello mordi e fuggi*
'business tourism and the "bite and run away" one'
d. *le nostre battaglie sono […] per un turismo dolce stanziale e non "mordi e fuggi"*
'we fight for a nice stationary tourism, not a "bite and run away" one'

Besides *turismo*, as many as 14 other nouns referring to travelling for pleasure occur modified by *mordi e fuggi*, for a total of 31 tokens; they are listed in (28):

(28) Action nouns belonging to the frame TOURISM modified by *mordi e fuggi* (14 types, 31 tokens)
arrembaggi turistici 'touristic boarding', *assalto* __[22] *ai centri storici* '__ assault on historic centres', *crociere* 'cruises', *ferie* 'vacation', *gita* 'day trip', *incursioni* 'incursions', *rapidi blitz* 'quick blitzes', *ritiri (spirituali)* (2)[23] '(spiritual) retreats', *soggiorno in fattoria* 'stay in a farm', *trasferta* 'trip', *rendez-vouz [sic]* 'rendez-vous', *vacanza* (2) 'vacation', *viaggi* (2) 'trips', *visita* (2) 'visit'

There is also a minority of nouns referring not to the activity, but to the agents who undertake quick pleasure trips (29):

[22] The underscore marks the position of the VeV lexeme in the string.
[23] The figure between round brackets indicates the number of tokens.

(29) Agent nouns belonging to the frame TOURISM modified by *mordi e fuggi* (4 types, 9 tokens)
turista (2) / *turisti* 3 'tourist(s)', *visitatore* / *visitatori* 'visitor(s)', *vacanzieri* 'vacationers', *tedeschi* 'Germans'[24]

Altogether, the TOURISM frame hosts about 50% of the adjectival tokens of *mordi e fuggi*.

No other frame is as richly represented as the TOURISM frame among nouns modified by *mordi e fuggi.* Other small clusters, sometimes attaining frame-like coherence, hardly ever surpass a 10% of the tokens. They are listed in (30):

(30) Frames within which *mordi e fuggi* occurs as an Adjective
a. ECONOMY (10 tokens)
venditore 'seller', *acquisto* 'buying', *investimento* 'investment', *operazione di Borsa* (2) 'stock exchange operation', *spesa* 'shopping', *spedizioni industriali* (2) 'industrial shipping', *fabbriche* 'factories', *imprese* 'enterprises'
b. GENERIC ACTIONS AND STATES (8 tokens)
iniziative 'initiative', *operazione* (2) 'operation', *attività* 'activity', *atteggiamento* 'attitude', *stile* (3) 'style'
c. MISCELLANEOUS ACTIVITIES (7 tokens)
sesso (2) 'sex', *ciclismo* 'bicycle racing', *neodiplomazia* 'neo-diplomacy', *passioni kennediane* 'kennedian passions', *agricoltura* 'agriculture', *ascolto* 'listening'
d. MEDIA, COMMUNICATION, JOURNALISM (8 tokens)
opinionista 'columnist', *minirecensioni* 'mini-reviews', *interviste* 'interviews', *dichiarazione* 'statement', *battute* 'utterances', *quotidiano __ Usa today* 'daily newspaper *USA Today*', *giornalismo* 'journalism', *televisione* 'television'
e. KIDNAPPINGS, CRIME (5 tokens)
rapimento 'kidnapping', *rapimenti* 'kidnappings', *sequestri* 'kidnappings', *crimine* 'crime', *attacchi* 'attacks'
f. OTHER AGENTS (4 tokens)
politicante 'politician', *assessori* 'councilmen', *voi* 'you (PL)', *raider* 'raider'

Some of the same frames host *mordi e fuggi* used as an Action noun, as can be seen in (31). To make the data more easily comparable, we have analyzed as many tokens of *mordi e fuggi* used as a noun as of *mordi e fuggi* used as an adjective (covering over 50% of the noun tokens).

(31) Frames within which *mordi e fuggi* occurs as an Action noun
(based on the analysis of 50% of the noun tokens)

ECONOMY	38 tokens
GUERRILLA	9 tokens
SPORTS	8 tokens
ENTERTAINMENT	8 tokens
FAST FOOD	8 tokens
TOURISM	7 tokens

Only a handful of tokens refer to generic actions or belong to other frames (for example, only one token belongs to the KIDNAPPINGS frame).

In these frames, *mordi e fuggi* as an action noun refers to some action which is characterized by quickness and is defined within the frame: rapid buying and selling in ECONOMY, quick

[24] For non-Italians it is maybe necessary to explain that Germans are the prototypical tourists in Italian culture, or at least they have been for a long time. A search in the *laR* corpus for the phrases corresponding to 'German tourist' (in all possible gender and number combinations) yields 381 tokens, whereas 'American tourist' occurs only 293 times, and 'Japanese tourist' a mere 167 times (not to speak of Greek tourists, mentioned only 12 times in 15 years of *la Repubblica*!).

actions in SPORTS and GUERRILLA, quick production of movies or concerts in ENTERTAINMENT, quick meals in FAST FOOD, quick visits in TOURISM. In some of these frames, quickness is not viewed as a negative feature, but rather as a positive one.

The TOURISM frame has generated other VeV lexemes used as adjectives: in the *laR* corpus we find a number of hapaxes connected to this frame, shown in (32).

(32) Other adjectival VeV lexemes (hapaxes) that modify nouns belonging to the TOURISM frame
a. *turista compra e scappa* 'buy and run_away tourist'
b. *visitatori prendi e scappa* 'take and run_away visitors'
c. *turista scatta (nel senso di foto) e torna a casa* 'take a snapshot and go_back home tourist'
d. *paesaggio tocca e scappa* 'touch and run_away landscape'
e. *turismo vedi e fuggi* (3 tokens)'see and run_away tourism'

In most of these lexemes, the actions referred to by V1 are not metaphorical, as in *mordi e fuggi* lit. bite and run_away; rather, V1s are descriptive, sometimes almost excessively so (as in the case of (32c) *scatta (nel senso di foto) e torna a casa*); clearly, their (nonce-)formation is guided by the TOURISM frame.

Another frame that has given rise to a few additional formations is that of MEDIA, COMMUNICATION, JOURNALISM, in which we find a few lexemes referring to situations in which there is a component of quickness (33):

(33) The MEDIA, COMMUNICATION, JOURNALISM frame
a. *libri kiss and tell*, *alla lettera bacia e racconta* '*kiss and tell*, literally kiss and tell, books'
b. *libri istantanei*, *prendi-e-getta* 'instant, take-and-throw_away, books'
c. *cartoni guarda-e-compra* 'look-and-buy cartoons'

While example (33a) shows that some of the VeV lexemes have been formed as calques of pre-existing English expressions, (33b) is particularly interesting because it shows that sometimes a VeV lexeme is formed even when it is somehow redundant: in this context there was already another adjective (*istantaneo* 'instant') that expressed the idea of quickness and was used to calque the English phrase *instant book*; *prendi-e-getta* adds a nuance. This lexeme brings us to another semantic class of VeV lexemes, the Aspectual class, that we will now describe.

4.2. Aspectual (ASP) class: *usa e getta*

The leader of the Aspectual class is *usa e getta*:

(34) *usa e getta* lit. use and throw_away
Adj = 'disposable'
Naction = 'a way of life characterized by wasting environmental resources'

There are 462 tokens of *usa e getta* (403 tokens spelled *usa e getta* and 59 tokens spelled *usa-e-getta)* in *laR*. Of these, more than 80% are adjectives. The analysis of the contexts in which *usa e getta* appears has been conducted on 100 tokens (all the 59 hyphenated tokens, and the first 41 non-hyphenated tokens). The distribution of nouns and adjectives in this sample matches the overall distribution of lexical categories in the full set of data.

As an adjective, *usa e getta* is mostly used to modify nouns that refer to disposable products, as the data in (35) show:

(35) Nouns denoting products modified by *usa e getta* (from a 100-token sample)
a. Generic, superordinate nouns (12 tokens)

articoli / articolo 'item(s)', *confezioni* 'package', *materiali* (2) 'supplies', *oggetti in plastica* 'plastic objects', *prodotti* (2) 'products', *strumenti* 'tools', *qualcosa* 'something', *pacchetto* [di lenzuola e federe] (2) 'package [containing bed linen]'

b. Specific, basic level nouns (29 tokens)
abiti di carta 'paper clothes', *accendini* (2) 'lighters', *aghi* (2) 'needles', *assemblages e collages __ di ritagli* '__ assemblages and collages of paper clippings', *bazooka del tipo __* 'bazooka of the __ kind', *biancheria* 'linen', *bicchiere di plastica* 'plastic cup', *bicchieri* 'cups', *bicchieri di carta* 'paper cups', *bottiglia / bottiglie* 'bottle(s)', *bottigliette in vetro* 'glass bottles:DIM', *cronometro* 'chronometer', *federe di carta* 'paper pillow-cases', *kodak* 'Kodak camera', *lamette* 'blades:DIM', *lenti a contatto* 'contact lenses', *giocattoli* 'toys', *missili* 'missiles', *pannolini* 'diapers', *penne a sfera e rasoi* 'ball-point pens and razors', *penna biro* 'ball-point pen', *portatile* 'notebook (computer)', *rasoi* 'razors', *stadio* 'stadium', *vestiti di piume* 'feather clothes', *volumi* 'volumes'

The second strongest cluster of nouns modified by *usa e getta* can be analyzed as belonging to a frame that we have called MEDIA AND ENTERTAINMENT (36):

(36) Nouns modified by *usa e getta* belonging to the MEDIA AND ENTERTAINMENT frame (13 tokens)

a. *successi __ per teenager* 'hits for teenagers', *canzonette* 'songs:DIM', *musichetta* (2) 'music:DIM', *scenografie* 'stage sets', *tv* 'tv', *raccontini rosa* 'chick-lit:DIM', *film* 'movie', *pellicolette* 'films:DIM', *teatro* 'theatre', *programmazione spesso __* 'often __ tv programs'

b. *la trasmissione del sabato sera, definita [da Intini] "un non programma usa e getta"*
'the Saturday night show, defined [by Intini] "a use and throw_away non-show"'

c. *tutto quello che lo stesso pubblico […] aveva considerato come "usa e getta"*
'everything that the same audience […] had considered as "use and throw_away"'

A few more nouns are connected to communication (particularly verbal communication), and could easily belong to the MEDIA, COMMUNICATION, JOURNALISM frame that was quite strong with *mordi e fuggi* (37):

(37) MEDIA, COMMUNICATION, JOURNALISM
paroletta 'word:DIM', *slogan* 'slogan', *battuta* 'utterance', *giornalismo* 'journalism'

In the same frames, *usa e getta* also occurs twice as a Concrete noun, as in (38):

(38) a. *non so ancora se un Vocabolario sia un effimero usa-e-getta come un giornale*
'I do not know yet whether a dictionary is an ephemeral use-and-throw_away like a newspaper'

b. *il teatro […] non è un usa e getta*
'theatre […] is not a use and throw_away'

A few nouns refer to people (39a), or groups of people (39b), and another small group of nouns modified by *usa e getta* is constituted of action nouns (39c):

(39) a. *amante* (2) 'lover', *uomini* 'men', *sindaco debole, __* 'weak, __ mayor', *leader* 'leader', *bambino* 'kid'

b. *sindacati kleenex, __* 'kleenex-like, __ unions', *pendolariato stile __* 'commuters in the __ style'

c. *tradimento __ che dura un giorno solo* '__ betrayal, that lasts only one day', *candidatura* 'candidature', *risata* 'laugh', *arrampicata* 'climbing', *arrampicate* 'climbings', *utilizzazione __ degli shoppers* '__usage of shopping bags''

An example of convergence between the Aspectual and the Manner classes is found in the quotation in (40), from our corpus:

(40) *Lui è il turista fast food*, *il turista usa e getta*, *il turista compra e scappa*, *il turista scatta (nel senso di foto) e torna a casa*.
'He is the fast food tourist, the use-and-throw_away tourist, the buy-and-run_away tourist, the take-a-snapshot-and-come_back-home tourist.'

In fact, the two classes have a common semantic flavour: both are used to qualify objects and actions that, for one reason or another, are not prototypical instances of their kind. Activities performed in a *mordi e fuggi* way, or agents that act in a *mordi e fuggi* way, are considered, by the speakers who use this adjective to qualify them, too quick: the activities are not performed in the correct, standard, typical way. Similarly, *usa e getta* objects are not the typical, permanent, long-lasting objects; they are not worth keeping. There is often a connotation of lesser value and lesser quality connected to the referents of the nouns that are modified by *usa e getta*; this is particularly clear with the nouns that belong to the MEDIA AND ENTERTAINMENT frame, that often appear in a diminutive form[25] (cf. (36a)).

This negative connotation stays with *usa e getta* even in those relatively few cases in which this lexeme is used as an action noun, listed in (41). As an action noun, *usa e getta* refers to a philosophy, an approach to life, and occurs mostly in contexts where the author criticizes it (41a). *Usa e getta* refers mostly to the production of disposable products, criticized because it causes damage to the environment, but is sometimes used metaphorically also to refer to an approach to human relations and life in general, as in (41b).

(41) *Usa e getta* as action noun
   a. *civiltà dell'usa e getta* 'use and throw_away civilization', *società dell'usa e getta* 'use and throw_away society', *mentalità dell'usa e getta* 'use and throw_away mentality', *ideologia dell'usa e getta* 'use and throw_away ideology', *all'insegna dell'usa e getta* 'under the banner of the use and throw_away', *società fondata sull'usa e getta* 'society based on use and throw_away', *il fascino effimero dell'usa e getta* 'the ephemeral charm of the use and throw_away', *L'usa e getta* 'the use and throw_away', *l'usa e getta frenetico* 'the frantic use and throw_away'
   b. *l'usa e getta delle idee e delle ideologie* 'the using-and-throwing away of ideas and ideologies'

Starting from *usa e getta*, a few more lexemes have developed, containing the verbs *getta* or *butta* 'throw_away' as V2. They are listed in (42) with the nouns they modify:

(42) a. *libri / romanzoni /lavoro culturale leggi e getta* 'books / novels:AUG / cultural work read-and-throw_away'
   b. *lenzuola metti e butta* 'put (to use) and throw_away sheets'
   c. *telefonino parla e getta* 'speak-and-throw_away cell phone'
   d. *i libri istantanei prendi e getta* 'instant take-and-throw_away books'
   e. *rasoio / rasoi radi e getta* 'shave and throw_away razor / razors'
   f. *[film] vedi e getta* 'see and throw_away [movies]'

These new formations appear to be driven by the "collocation" the modified noun entertains with a particular verb: books are read, movies are seen, cell phones are used to speak, razors to shave, and so on.

Starting from the element of quickness contained in *mordi e fuggi*, and probably helped by the negative connotation of haste of the *usa e getta* type, some lexemes have been created that sometimes contain neither *fuggi / scappa* nor *getta / butta* as V2, but share the general negative

[25] Lesser importance or value is one of the possible connotations of diminutives (cf. Dressler & Merlini Barbaresi 1994: 121).

connotation and refer to undesirable or objectionable quickness (in the mind of the speaker); most of these adjectives occur in context having to do with quick sex, as shown in (43a-d):

(43) a. *camiciole apri-e-gusta* 'open-and-enjoy shirts:DIM'
b. *una signora sbatti-e-butta* 'a fuck-and-throw_away lady'
c. *scopate prendi e fuggi* 'take-and-run_away fucks'
d. *l'idea del sesso è prendi e getta* 'the idea of sex is take-and-throw_away'
e. *una componente tocca e fuggi* 'a touch-and-run_away component'
f. *una chiesa quasi prega e fuggi* 'an almost pray-and-run_away church'

The lexemes in (33a-b) above (*libri kiss and tell*, *alla lettera bacia e racconta*, and *libri istantanei*, *prendi-e-getta*) also belong to this group.

In a few cases (44), quickness has a positive, or at least non-negative, connotation (that could be linked to the practicality also connected with disposable objects, qualified by *usa e getta*):

(44) a. *poliestere lava e indossa* 'wash and wear polyester'
b. *merendine scarta e inghiotti* 'unwrap and swallow snacks'

This group of lexemes which carry a positive connotation is preceded by a use of *apri e gusta* that is not found in our *laR* corpus, but is well known to most middle-aged Italians: *apri e gusta* was a slogan first used in 1971[26] (perhaps as a sequence of two real imperatives) in tv commercials for Manzotin canned beef (45):

(45) *Manzotin: apri e gusta* 'Manzotin[27]: open and enjoy'

In our corpus, *apri e gusta* refers to sexy shirts, easy to unbutton, and therefore has entered the QUICK SEX frame. *Scarta e inghiotti*, instead, is a true descendant of *apri e gusta*, developed in the frame of ready-to-eat food. *Lava e indossa* is a calque of English *wash and wear*: it is noticeable that no attempt to replicate the alliterative make-up of the English model has been made in the calque.

4.3. Cause-effect (CONSEQ) class: *gratta e vinci*

The leader of the cause-effect class is *gratta e vinci*:

(46) *gratta e vinci* lit. scratch and win
Ngame = 'instant scratch lottery'
Nconcrete = 'instant scratch lottery ticket'
Adj = 'scratch-and-win-like' or 'to be scratched'

There are 190 tokens of *gratta e vinci* in *laR* corpus. Over 70% of them are used as the name of a specific lottery existing in Italy, or of any kind of instant lottery.

As a concrete noun, *gratta e vinci* refers to instant lottery tickets; only once, in a satirical article, does it refer to a slip of paper of a different kind (a garlic-flavoured slip of paper replacing the hamburger in a McDonald's sandwich, to save money).

As an adjective, again *gratta e vinci* occurs mostly to modify nouns that refer to an instant lottery or its tickets (47a), although in a few cases (47b) it spreads to other referents:

(47) Nouns modified by *gratta e vinci* used as an adjective

[26] The date of first usage of the *apri e gusta* slogan was obtained from the *Manzotin Consumers Service* (e-mail to Anna M. Thornton of October 26, 2007).
[27] The etymology of Manzotin as given on the company's official website (www.manzotin.com) is the following: *manzo* 'beef' + *tin* 'tin can'.

a. *lotteria* (9) 'lottery', *lotterie* (6) 'lotteries', *biglietti* (3) 'tickets', *lotterie e concorsi* 'lotteries and competitions with prizes'
b. *schede magnetiche tipo* __ 'magnetic cards of the __ type', *pellicola* 'film', *concorsone* 'state_exam:AUG', *la nostra società* 'our society''

Clearly, the main frame for this class is the INSTANT LOTTERY frame. Some other lexemes are formed on the model of *gratta e vinci* by exploiting syntagmatic or paradigmatic relations with one or both of its constituents, and always refer to instant lotteries or some other game (48):

(48)			
	gratta e vinci	'scratch and win'	
	gratta e **perdi**	'scratch and lose'	(antonym of *vinci*)
	gratta e **stravinci**	'scratch and over-win'	(derived from *vinci* with an intensifying prefix)
	gioca e vinci	'play and win'	(superordinate term in the frame of games)
	tira e vinci	'kick and win'	(co-hyponym of *gratta*, specifying the action to be performed to play)

However, a few adjectival tokens show that it is spreading to other frames, having to do more with the organization of society. In one instance *gratta e vinci* modifies the noun *concorsone* 'state_exam:AUG (a public selection to obtain a job)', and a *concorsone gratta e vinci* is a selection in which the outcome is considered, in the view of the speaker, to depend on sheer luck, and not on the candidates' qualifications. In one instance, *gratta e vinci* modifies 'our society', to express the view that luck counts more than qualification in Italian society overall.

Three other lexemes belonging to this class do not seem to have developed on the model of *gratta e vinci*. They are more closely connected to proverbs, and in general to gnomic sayings (49).

(49)				
	a.	*cerca e trova* 'look for and find'	→	Proverb: *Chi cerca trova* 'Seek and you will find'
	b.	*guarda e impara* 'look and learn'	→	Proverb: *Sbagliando s'impara* 'You learn from your mistakes'
	c.	*prevedi e previeni* 'foresee and forestall'		

In his paper on the stylistic usages of imperatives, Spitzer has shown that one possible usage, which he calls "impératif gérondial", is in sayings that are closely related to proverbs, in that they depict a situation with a repetition of the same imperative twice, and then conjoin this sequence of imperatives, by means of the conjunction corresponding to *and*, to a verb form that depicts the "necessary consequence" of the first action. Spitzer maintains that the conjunction is used to express "la conséquence nécessaire" of the actions named in the first half of the construction. Spitzer observes that this "*et* de la conséquence nécessaire" is found in proverbs and in gerundial imperatives. He illustrates this with the Catalan expression *cerca que cerca, y a la fí el trobarem* 'look and look, *and* at the end we'll find it' (Spitzer 1951: 457, n. 1). It is possible that some of our lexemes belonging to the CONSEQ class have been formed by association with "impératifs gérondials" rather than with descriptive imperatives, as is the case in most of our data.

4.4. Sequentials (SEQ) class

A few lexemes with *gratta* as V1 occur also in the SEQ semantic class: the best attested is *gratta e sosta*, which refers to a system of paying parking fees by buying and scratching pre-paid vouchers, and is also used as a concrete noun to refer to the vouchers themselves. The SEQ class does not seem to have a leader lexeme; most of its members have already been explained in connection with frames developing from the leader lexemes of the ASP and MAN classes; a

small cluster of nouns in the SEQ class refers to games or specific kinds of actions to be performed in certain games and sports (cf. (50)); the few remaining lexemes do not fit in any specific frame.

(50) GAMES, ACTIONS IN GAMES

corri e tira	lit. run and shoot
mangia e passa	lit. eat and pass
tocca e corri	lit. touch and run

It must also be observed that a relatively high number of lexemes in the SEQ class is represented by calques from English, as shown in (51):

(51) Lexemes in the SEQ class that are calques from English

a.	*copia e incolla*	<	*copy and paste*
b.	*taglia e incolla*	<	*cut and paste*
c.	*lava e indossa*	<	*wash and wear*
d.	*bacia e racconta*	<	*kiss and tell*
e.	*corri e tira*	<	*run and shoot*
f.	*tassa e spendi*	<	*tax and spend*
g.	*gratta e sniffa*	<	*scratch and sniff*

All in all, it seems that so far the SEQ class has not established itself firmly. Its existence may be an illusion on our part: this supposed class collects an almost random set of lexemes that are better understood as calques or as modelled on leaders of other, better established and semantically coherent classes.

4.5. Opposites (OPP) class: *tira e molla*, *va e vieni*

The leaders of the Opposites class are *tira e molla* and *va e vieni*:

(52) a. *tira e molla* lit. pull and let_go ‘see-saw, playing fast and loose’
(attested since at least 1890)

b. *va e vieni* lit. go and come ‘coming and going, to and fro’
(attested since at least 1870)

This class contains some of the oldest lexemes of the VeV type: *va e vieni* is attested since at least 1870, *tira e molla* is attested since at least 1890, and *sali e scendi* in the variant without conjunction *saliscendi* is attested as an action noun since 1836, and as a concrete noun as early as 1400.

Almost all the lexemes in this class are primarily action nouns; some of them can also be used as adjectives or concrete nouns (but interestingly, the leader nouns seem to be impervious to becoming adjectives: *tira e molla* can be interpreted as an adjective in only 3 of its 572 tokens in *laR* corpus).

Semantically speaking, this class is close to other kinds of verb-based lexemes, such as those VV compounds that are action nouns and VV reduplicative compounds, which are also action nouns. The only formal difference between the VeV lexemes in this class and other VV action nouns resides in the presence *vs.* absence of the conjunction *e* ‘and’: but Zamboni (1986) has observed that several lexeme-types are attested in variants with and without the conjunction. Thus, within the OPP class the presence of the conjunction is not relevant to the semantics of the lexemes.

The VeV action nouns refer to complex events, consisting of two opposite actions that take place alternatively, performed by the same subject(s) or by different subjects.

Many of these nouns refer to a kind of stop-and-go in human relations, particularly in political negotiations, but also in love partnerships.

The nouns often occur in syntactic contexts that refer to the continual, lengthy, time-consuming alternation of opposite actions, as illustrated in (53):

(53) a. *un accendi e spegni continuo*
'a continual switch_on and switch_off'
b. *l'interminabile cuci e scuci di maggioranze interne alla DC*
'the unending sew and unsew of majority groups within Christian Democracy'
c. *una notte di tira-e-molla*
'a whole night of pull and let_go'
d. *il tira-e-molla era durato per quasi tutto l'autunno*
'the pull and let_go had lasted almost all autumn'
e. *un lungo tira-e-molla* (2)
'a long pull and let_go'
f. *dopo otto ore* / *dopo un mese* (2)/ *dopo anni* (2) *di tira-e-molla*
'after eight hours / one month / years of pull and let_go'

Other adjectives that occur often with *tira e molla* are *estenuante* (19) 'exhausting' and *faticoso* (4) 'heavy', that refer to how unbearable the sequence of actions is.

No specific frame is evoked, even if the high token frequency of *tira e molla* and the high frequency of contexts referring to political negotiations within it result in a high frequency of contexts having to do with politics.

5. Discussion

5.1. The emergence of VeV semantic classes on an exemplar basis

In this final section we would like to discuss briefly some theoretical implications of our findings, starting from the semantic classification of VeV lexemes.

The detailed corpus-based analysis in section 4 shows that each VeV semantic class has emerged thanks to the high frequency of a single type. For instance, the Manner class has clearly emerged due to the very frequent occurrence of the leader *mordi e fuggi* lit. bite and run_away '(too) quick, quick operation' (both in the nominal and the adjectival use) within certains frames, above all the TOURISM frame. The general meaning of "quickness" and "superficiality" conveyed by *mordi e fuggi* also extends to other frames, such as for instance ECONOMY, ENTERTAINMENT, SPORTS, etcetera. At the same time, within both the TOURISM frame and other frames, a number of other VeV lexemes emerge that are semantically similar to *mordi e fuggi*: see for instance *tocca e scappa* (lit. touch and run_away) or *vedi e fuggi* (lit. see and run_away). These new formations retain the general semantic meaning of *mordi e fuggi*, but feature other input verbs (especially in V1 position) that are more specific to the situation depicted. A clear example can be found in the Aspectual class (whose leader is *usa e getta* lit. use and throw_away 'disposable'): in the expression *libri leggi e getta* (lit. books read and throw_away), the verb 'read' is clearly used by collocation with 'books'.

Whereas some of the classes appear to be bound to one or more specific frames and new formations seem to be the result of analogical processes strictly speaking (this is the case of the above-mentioned Manner and Aspectual classes), other classes, such as the Opposites class, have a wider scope. For instance, the hapax *accendi e spegni* (lit. switch on and switch off) obviously shares with the leaders (*tira e molla* lit. pull and let_go 'playing fast and loose'; *va e vieni* lit. go and come 'coming and going') the abstract meaning of 'doing X and doing the opposite of X repeatedly', but its specific semantics is independent of the specific semantics of the leaders.

The phenomena just described can be straightforwardly accounted for in a usage-based constructionist model of language such as that outlined, among others, by Bybee (2006) (cf. also Bybee & Eddington 2006, Goldberg 2006, and Tomasello 2003 for an acquisitional perspective).

Bybee (2006: 711) claims that "grammar is the cognitive organization of one's experience with language" and that "certain facets of linguistic experience, such as the frequency of use of particular instances of constructions, have an impact on representation". More important for our current purposes, Bybee (2006: 716) insists on the importance of token frequency and

"exemplar" representations[28]: "new tokens of experience are not decoded and then discarded, but rather they impact memory representations. In particular, a token of linguistic experience that is identical to an existing exemplar is mapped onto that exemplar, strengthening it. Tokens that are similar but not identical […] to existing exemplars are represented as exemplars themselves and are stored near similar exemplars to constitute clusters or categories. […] Constructions emerge when phrases that bear some formal similarity as well as some semantic coherence are stored close to one another". Therefore, the clustering of exemplars plus a high level of repetition "can lead to the establishment of a new construction with its own categories" (2006: 719). Almost the same process is named "entrenchment" by Langacker (2000).

It is now clear that the emergence of VeV lexemes in Italian can be explained in terms of an "exemplar" effect: some expressions (i.e., the leaders of the semantic classes, which are themselves lexically specified constructions), due to their high frequency and to exemplar clustering, become entrenched and new more abstract or schematic constructions then arise (identified by the various semantic classes). The existence of these more abstract constructions accounts for the formation of new expressions that instantiate it.

5.2. Towards a family of constructions

Another issue we would like to discuss is whether there is ground to cluster together some of the semantic classes identified above, in order to draw some higher-level generalizations. In order to reach this goal, we have selected the following three criteria[29]:

- **± fixed order** of the events denoted by V1 and V2;
- **± iterativity** of the two events;
- **± obligatorily shared subject** for the two events.

By applying these criteria to our data, we are able to identify two main macro-classes of VeV lexemes, which are shown in Table 27. The two macro-classes cover 97% of the types in the corpus and leave out two types only, namely: *mangia e bevi* lit. eat and drink 'a sort of sundae' (the only item in the CO-HYPO class) and *scappa e fuggi* lit. run_away and run_away 'quick' (the only item in the SYN class). Both macro-types seem to be productive to some extent, since there are some hapaxes in each.

Table 27 **The macro-classes of VeV lexemes**

macro-class	fixed order	iterativity	shared subject	semantic types involved	n. of types	n. of tokens	n. of hapaxes
1	-	+	±	OPP	18	755	7
2	+	-	+	CONSEQ, MAN, ASP, SEQ	46	1025	17

According to the data in Table 27, there is a first class (**macro-class 1**) that has the following characteristics: first, it denotes a complex event described by the sequence of two opposite events that may occur more than once, with no priority of one event over the other (so that we can have both sequences V1-V2 or V2-V1, e.g., *prendi e lascia* lit. catch and release, where the events of catching and releasing are not strictly ordered); second, the events referred to by the two verbs may be performed either by one and the same subject or by different subjects, i.e., there may be a disjunct subject interpretation (cf. example (15) in section 2.3.4.1). In this second case, the different subjects can be interpreted as one collective and generic subject. Of course, this is closely linked to the positive value of the iterativity feature (and, secondarily, to the negative value of the fixed order feature).

[28] As Bybee (2006: 716) states, exemplar theory was originally developed in psychological studies of categorization and was then applied, in linguistics, to phonetics.

[29] Note that all the criteria make reference to the events denoted by the input verbs, not to the linguistic realization of the verbs themselves. Therefore, "fixed order" does not refer to the order of the constituents, but to the conceptual representation of the event denoted by the whole lexeme.

The situation is different in **macro-class 2** lexemes, which denote single complex events formed by two fixed phases – V1 and V2 – that are not iterated and have to be performed by one single subject. In general, V1 and V2 are not linked by a lexico-semantic relationship (apart from the CONSEQ class), but share the same frame. Some of the macro-class 2 lexemes, such as those belonging to the Aspectual and Manner classes, have developed a more abstract meaning (cf. section 2.3.4.3).

From what we have seen so far, the key feature here seems to be played by the "verbal number" category as described in Corbett (2000: 243-264): event/participant plurality is the feature that distinguishes the two macro-classes, and is also the feature that links the two macro-classes to VV reduplicative compounds. As mentioned at the beginning of the paper, according to Thornton (2007b) VV reduplicative compounds are a fairly productive construction that forms action nouns with a (compulsory) plural agent. Therefore, we can depict a *continuum* involving these three constructions like the one represented in Figure 6.

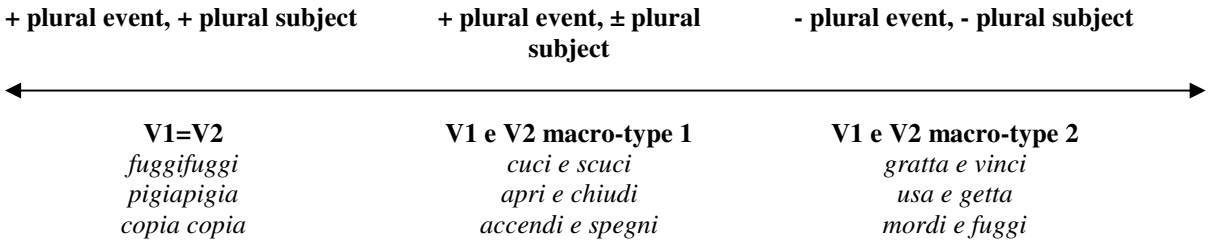


Figure 6 **Productive Italian V-(e)-V lexical constructions**

In conclusion, it seems that VV lexemes in Italian form a family of constructions that resemble one another in both form (all of them are made up of two verbs) and meaning (all of them have a complex event semantics) and can be distinguished according to a set of parameters, such as the ones listed above[30].

6. Conclusions

In this paper, we have offered a fine-grained analysis of Italian VeV lexical constructions, i.e., phrasal lexemes formed by two conjoined verbs that yield a noun or an adjective as output. These constructions are peculiar in that they cannot be regarded as prototypical words, but nonetheless have the status of lexemes.

In our analysis, we have used two approaches.

On the one hand, we have applied to VeV lexemes the traditional methodology used in Lexical Morphology in order to capture the generalizations underlying the formations of these items. Therefore, we have looked for constraints on the input and the output of VeV lexemes at the main levels of linguistic analysis, namely phonology, morphology, syntax and semantics.

On the other hand, we have made use of usage-based constructionist models in order to account for the rise of various semantic classes of VeV lexemes and new formations. Data form corpora show that the emergence of VeV lexemes is to be ascribed to the entrenchment of a few exemplars, which led to the creation of more abstract constructions that in turn license the formation of new lexemes.

Finally, VeV lexemes are found to be closely related to other VV lexemes in Italian that are usually ascribed to the realm of compounding: all these constructions are dedicated strategies to encode a complex event semantics into different linguistic forms, namely nouns and adjectives (but not verbs). This obvioulsy raises interesting typological questions regarding the nature and

[30] Some syntactico-semantic properties of VeV lexemes (such as their sharing arguments, and the kinds of semantic classes to which they belong) suggest a connection of VeV lexemes with the constructions known as "serial verbs" (cf. Aikhenvald 2006). As a path for further research, we suggest that at least some Italian VeV lexemes could be analyzed as nominalizations of virtual serial verbs.

role of lexicalization strategies involving two verbs in the languages of the world, which we leave for future research.

Appendix: the corpus[31]

N	VeV lexemes	Literal translation	Semantic category	Lexical category	Tokens per variant (*laR*)	Tokens per lexeme (*laR*)
1	***accendi e spegni***	switch on and switch off	OPP	Naction	1	1
2	***appila e spila***	plug and unplug	OPP	Naction	1	1
3a	***apri e chiudi***	open and close	OPP	Adj	6	9
3b	*apri-chiudi*	open-close	OPP	Adj	1	
3c	*apri e chiudi*	open and close	OPP	Naction	2	
4	***apri-e-gusta***	open-and-enjoy	SEQ	Adj	1	1
5	***attacca e stacca***	attach and detach	OPP	Adj	0	0
6a	***bacia e racconta***	kiss and tell	SEQ	Adj	2	3
6b	*bacia e racconta*	kiss and tell	SEQ	Naction	1	
7a	***cerca e trova (lavoro)***	look for and find (job)	CONSEQ	Naction	1	2
7b	*cerca-trova*	look for-find	CONSEQ	Nconcrete	1	
8	***compra e fuggi***	buy and run away	MAN	Naction	1	1
9	***compra e scappa***	buy and run away	MAN	Adj	1	1
10	***compra-e-vendi***	buy-and-sell	OPP	Naction	1	1
11	***consuma-e-getta***	consume-and-throw away	ASP	Nconcrete	1	1
12a	***copia e incolla***	copy and paste	SEQ	Adj	0	0
12b	*copia e incolla*	copy and paste	SEQ	Naction	0	
13a	***copri e scopri***	cover and uncover	OPP	Adj	0	1
13b	*copri e scopri*	cover and uncover	OPP	Nconcrete	0	
13c	*copri e scopri*	cover and uncover	OPP	Naction	1	
14a	***corri e tira***	run and shoot	SEQ	Adj	2	11
14b	*corri e tira*	run and shoot	SEQ	Naction	9	
15a	***cuci e scuci***	sew and unsew	OPP	Naction	3	4
15b	*cuci-scuci*	sew-unsew	OPP	Naction	1	
16	***dai e prendi***	give and take	OPP	Naction	1	1
17	***fotti e chiagni***	cheat and cry	SEQ	Naction	2	2
18	***gioca & stravinci***	play & over-win	CONSEQ	Ngame	0	0
19a	***gratta e perdi***	play and lose	CONSEQ	Adj	0	2
19b	*gratta e perdi*	play and lose	CONSEQ	Ngame	1	
19c	*gratta-e-perdi*	play-and-lose	CONSEQ	Ngame	1	
20a	***gratta e sniffa***	scratch and sniff	SEQ	Adj	0	0
20b	*gratta e sniffa*	scratch and sniff	SEQ	Nconcrete	0	
21a	***gratta e sosta***	scratch and park	SEQ	Adj	0	8

[31] Each type is marked in bold type and numbered progressively. The spelling variants (based mainly on the corpus *la Repubblica*; when the item is not attested in this corpus, the spelling found in the original source is used) and the different lexical categories (based on both the corpus *la Repubblica* and Google) are listed under each type in Roman type (with progressive letters, e.g., 13a, 13b, and so on). All token figures are based on the corpus *la Repubblica*. However, most types are attested on the internet.
After completing our analysis, we found five more VeV lexemes: ***agita e gusta*** (lit. shake and taste, a proper name indicating a food product by Bonduelle); ***punta & clicca*** (lit. point and click, used as an adjective in combination with the noun *avventura* 'adventure' to indicate a computer adventure); ***strappa e vinci*** (lit. strip and win, a sort of game similar to *gratta e vinci*); ***arrusti e mangia*** (Regional Italian spoken in Catania, Sicily; lit. roast and eat, a takeaway food shop where horse meat is roasted and served in a sandwich, and also the sandwich itself); ***gratta e truffa*** (lit. scratch and cheat, used as an adjective with the noun *francobollo* 'stamp').

21b	*gratta e sosta*	scratch and park	SEQ	Nconcrete	3	
21c	*gratta e sosta*	scratch and park	SEQ	Naction	5	
22	***gratta e spara***	scratch and gun down	SEQ	Adj	1	1
23	***gratta e stravinci***	scratch and over-win	CONSEQ	Ngame	0	0
24a	***gratta e vinci***	scratch and win	CONSEQ	Adj	25	194
24b	*gratta-e-vinci*	scratch-and-win	CONSEQ	Adj	2	
24c	*gratta e vinci*	scratch and win	CONSEQ	Naction	2	
24d	*gratta e vinci*	scratch and win	CONSEQ	Nconcrete	23	
24e	*gratta-e-vinci*	scratch-and-win	CONSEQ	Nconcrete	1	
24f	*gratta e vinci*	scratch and win	CONSEQ	Ngame	140	
24g	*gratta-e-vinci*	scratch-and-win	CONSEQ	Ngame	1	
25	***guarda e compra***	look and buy	SEQ	Adj	0	0
26a	***guarda e impara***	look and learn	CONSEQ	Adj	0	0
26b	*guarda e impara*	look and learn	CONSEQ	Naction	0	
27	***lascia-e-piglia***	leave-and-take	OPP	Naction	1	1
28	***lava e indossa***	wash and wear	SEQ	Adj	1	1
29a	***leggi e getta***	read-and-throw away	ASP	Adj	3	6
29b	*leggi-e-getta*	read-and-throw away	ASP	Adj	1	
29c	*leggi-e-getta*	read and throw away	ASP	Nconcrete	2	
30a	***mangia e bevi***	eat and drink	CO-HYPO	Adj	1	1
30b	*mangia e bevi*	eat and drink	CO-HYPO	Nconcrete	0	
30c	*mangia e bevi*	eat and drink	CO-HYPO	Naction	0	
31	***mangia e passa***	eat and pass	SEQ	Ngame	0	0
32	***metti e butta***	put on and throw away	ASP	Adj	1	1
33	***metti e togli***	put and remove	OPP	Naction	2	2
34a	***mordi e fuggi***	bite and run away	MAN	Adj	92	261
34b	*mordi-e-fuggi*	bite-and-run away	MAN	Naction	143	
34c	*mordi e fuggi*	bite and run away	MAN	Nconcrete	2	
34d	*mordi-e-fuggi*	bite-and-run away	MAN	Adj	11	
34d	*mordi-e-fuggi*	bite-and-run away	MAN	Naction	12	
34e	*mordi-fuggi*	bite-run away	MAN	Naction	1	
35	***parla e fuggi***	talk and run away	MAN	Naction	1	1
36a	***parla e getta***	talk and throw away	ASP	Adj	0	0
36b	*parla e getta*	talk and throw away	ASP	Nconcrete	0	
37	***piglia e lascia***	take and leave	OPP	Naction	0	0
38	***porta il riso in cascina e piangi miseria***	take the rice home and plead poverty	SEQ	Naction	1	1
39a	***prega e fuggi***	pray and run away	MAN	Adj	1	1
39b	*prega e fuggi*	pray and run away	MAN	Naction	0	
40	***prega e segui***	pray and follow	SEQ	Naction	0	0
41a	***prendi e fuggi***	take and run away	MAN	Adj	4	6
41b	*prendi e fuggi*	take and run away	MAN	Naction	2	
42a	***prendi e getta***	take and throw away	ASP	Adj	1	2
42b	*prendi-e-getta*	take-and-throw away	ASP	Adj	1	
42c	*prendi-e-getta*	take-and-throw away	ASP	Naction	0	
43a	***prendi e lascia***	catch and release	OPP	Naction	1	2
43b	*prendi-e-lascia*	catch-and-release	OPP	Naction	1	

43c	*prendi-e-lascia*	catch-and-release	OPP	Ngame	0	
43d	*prendi-e-lascia*	catch-and-release	OPP	Adj	0	
44a	***prendi e scappa***	take and run away	MAN	Adj	1	1
44b	*prendi e scappa*	take and run away	MAN	Naction	0	
44c	*prendi e scappa*	take and run away	MAN	Ngame	0	
45	***prevedi e previeni***	foresee and forestall	CONSEQ	Adj	1	1
46a	***radi e getta***	shave and throw away	ASP	Adj	2	4
46b	*radi-e-getta*	shave-and-throw away	ASP	Adj	2	
46c	*radi e getta*	shave and throw away	ASP	Nconcrete	0	
47a	***sali e scendi***	ascend and descend	OPP	Naction	7	14
47b	*sali-scendi*	ascend-descend	OPP	Naction	3	
47c	*sali e scendi*	ascend and descend	OPP	Nconcrete	2	
47d	*sali-scendi*	ascend-descend	OPP	Nconcrete	1	
47e	*sali e scendi*	ascend and descend	OPP	Adj	1	
48	***sbatti-e-butta***	fuck-and-throw away	ASP	Adj	1	1
49a	***scappa e fuggi***	run away and run away	SYN	Adj	5	7
49b	*scappa e fuggi*	run away and run away	SYN	Naction	2	
50	***scarta e inghiotti***	unwrap and swallow	SEQ	Adj	1	1
51	***scatta (nel senso di foto) e torna a casa***	snap (as with a camera) and come back home	MAN	Adj	1	1
52a	***stira e ammira®***	iron and admire	SEQ	Adj	2	3
52b	*stira-e-ammira®*	iron-and-admire	SEQ	Adj	1	
52c	*stira e ammira®*	iron and admire	SEQ	Nconcrete	0	
53a	***taglia e cuci***	cut and sew	OPP	Adj	2	28
53b	*taglia e cuci*	cut and sew	OPP	Naction	21	
53c	*taglia-e-cuci*	cut-and-sew	OPP	Naction	5	
54a	***taglia e incolla***	cut and paste	SEQ	Adj	0	5
54b	*taglia e incolla*	cut and paste	SEQ	Naction	5	
55a	***tassa e spendi***	tax and spend	SEQ	Adj	6	17
55b	*tassa-e-spendi*	tax-and-spend	SEQ	Adj	1	
55c	*tassa-spendi*	tax-spend	SEQ	Adj	1	
55d	*tassa e spendi*	tax and spend	SEQ	Naction	9	
56a	***tira e molla***	pull and let go	OPP	Adj	2	589
56b	*tira e molla*	pull and let go	OPP	Naction	532	
56c	*tira e molla*	pull and let go	OPP	Ngame	1	
56d	*tira-molla*	pull-let go	OPP	Naction	16	
56e	*tira-e-molla*	pull-and-let go	OPP	Naction	37	
56f	*tira-e-molla*	pull-and-let go	OPP	Adj	1	
57	***tira e vinci***	kick and win	CONSEQ	Ngame	2	2
58	***tocca e corri***	touch and run	SEQ	Naction	1	1
59a	***tocca e fuggi***	touch and run away	MAN	Adj	2	5
59b	*tocca e fuggi*	touch and run away	MAN	Naction	3	
60a	***tocca e scappa***	touch and run away	MAN	Adj	0	1
60b	*tocca e scappa*	touch and run away	MAN	Naction	1	
60c	*tocca e scappa*	touch and run away	MAN	Ngame	0	
61a	***togli e metti***	remove and put	OPP	Adj	1	3
61b	*togli e metti*	remove and put	OPP	Naction	2	
62	***tura e stura***	cork and uncork	OPP	Naction	1	1

63a	***usa e getta***	use and throw away	ASP	Adj	337	462
63b	*usa e getta*	use and throw away	ASP	Naction	52	
63c	*usa e getta*	use and throw away	ASP	Nconcreto	14	
63d	*usa-e-getta*	use-and-throw away	ASP	Adj	53	
63e	*usa-e-getta*	use-and-throw away	ASP	Naction	4	
63f	*usa-e-getta*	use-and-throw away	ASP	Nconcrete	2	
64a	***va e vieni***	go and come	OPP	Naction	84	97
64b	*va-e-vieni*	go-and-come	OPP	Naction	13	
65a	***vedi e fuggi***	look/watch and run away	MAN	Adj	1	3
65b	*vedi e fuggi*	look/watch and run away	MAN	Naction	2	
66a	***vedi e getta***	look/watch and throw away	ASP	Adj	1	3
66b	*vedi e getta*	look/watch and throw away	ASP	Nconcrete	2	

References

Aikhenvald, A. Y. (2006), *Serial Verb Constructions in Typological Perspective*, in Aikhenvald, A. & Dixon, R. M. W., *Serial Verb Constructions. A Cross-linguistic Typology*, Oxford, Oxford University Press, 1-68.

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G. & Mazzoleni, M. (2004), *Introducing the* "la Repubblica" *corpus: A large*, *annotated*, *TEI(XML)-compliant corpus of newspaper Italian*, in *Proceedings of the 4th International Conference on Language Resources and Evaluation* (LREC 2004, Lisbona, 26-28 maggio), Parigi, ELRA, 1771-1774.

Behaghel, O. (1909), *Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern*, in "Indogermanische Forschungen" XXV, 110-142.

Behaghel, O. (1928), *Deutsche Syntax. Eine geschichtliche Darstellung. Band III: Die Satzgebilde*, Heidelberg, Carl Winters Universitätsbuchhandlung.

Bertinetto, P. M. (1986), *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*, Florence, Accademia della Crusca.

Blake, B. (1994), *Case*, Cambridge, Cambridge University Press.

Booij, G. (2005a), *Compounding and derivation: evidence for Construction Morphology*, in Dressler, W. U., Kastovsky, D., Pfeiffer, O. E. & Rainer, F. (eds.), *Morphology and its Demarcations*, Amsterdam/Philadelphia, Benjamins, 109-132.

Booij, G. (2005b), *Construction-dependent morphology*, in "Lingue e Linguaggio" IV (2), 163-178.

Booij, G. (2005c), *Construction Morphology*, Vrije Universiteit Amsterdam, Ms.

Bybee, J. (2006), *From usage to grammar: the mind's response to repetition*, in "Language" 82 (4), 711-733.

Bybee, J. & Eddington, D. (2006), *A Usage-Based Approach to Spanish Verbs of 'Becoming'*, in "Language" 82 (2), 323-355.

Corbett, G. (2000), *Number*, Cambridge, Cambridge Univerity Press.

CORDE = REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español.* <http://www.rae.es>

Crocco Galèas, G. & Iacobini, C. (1993), *Parasintesi e doppio stadio derivativo nella formazione verbale del latino*, in "Archivio glottologico italiano" 78, 167-199.

Cruse, A. D. (1986), *Lexical semantics*, Cambridge, Cambridge University Press.

Dardano, M. (1978), *La formazione delle parole nell'italiano di oggi*, Rome, Bulzoni.

Dressler, W. U. & Thornton, A. M. (1991), *Doppie basi e binarismo nella morfologia italiana*, in "Rivista di linguistica" 3 (1), 3-22.

Dressler, W. U. & Merlini Barbaresi, L. (1994), *Morphopragmatics. Diminutives and Intensifiers in Italian*, *German*, *and Other Languages*, Berlin & New York, Mouton de Gruyter.

Dressler, W. U., Kilani-Schoch, M., Spina, R. & Thornton, A. M. (2003), *Le classi di coniugazione in italiano e francese*, in Giacomo-Marcellesi, M. & Rocchetti, A. (eds.), *Il verbo italiano: studi diacronici*, *sincronici*, *contrastivi e didattici*, Rome, Bulzoni, 397-416.

Folena, G. (1958), Review of Prati 1958, in "Lingua nostra" XIX, 3, 104.

Fillmore, C. J., Kay, P. & O'Connor, M. C. (1988), *Regularity and idiomaticity in grammatical constructions: the case of* let alone, in "Language" 64 (3), 501-538.

Giordano, R. & Voghera, M. (2002), *Verb system and verb usage in spoken and written Italian*, in Morin, A. & Sébillot, P. (eds.), Jadt 2002, *6èmes Journées Internationales d'Analyse Statistique des Données Textuelles*, IRISA, Rennes, 289-299.

Goldberg, A. (1995), *Constructions. A Construction Grammar Approach to Argument Structures*, Chicago, The University of Chicago Press.

Goldberg, A. (2006), *Constructions at Work*, Oxford, Oxford University Press.

GRADIT = *Grande dizionario italiano dell'uso*, edited by Tullio De Mauro, Turin, UTET, 1999.

Guevara, E. & Scalise, S. (forthcoming), *Searching for Universals in Compounding*, in Scalise, S. *et alii* (eds.), *Universals of Language*, Amsterdam, Springer.

Hopper, P. & Thompson, S. A. (1984), *The Discourse Basis for Lexical Categories in Universal Grammar*, in "Language" 60 (4): 703-752.

Jespersen, O. (1905), *Growth and Structure of the English language*, Leipzig, B. G. Teubner.

Ježek, E. (2003), *Classi di verbi tra semantica e sintassi*, Pisa, Edizioni ETS.

Ježek, E. (2005), *Lessico*, Bologna, Il Mulino.

Kahane, R. (1938), *Italienische Marinenwörter im Neugriechischen anlässlich*, in "Archivum Romanicum" XXII, 510-582.

Langacker, R. W. (2000), *A Dynamic Usage-Based Model*, in Barlow, M. & Kemmer, S. (eds.), *Usage Based Models of Grammar*, Stanford, CSLI Publications, 1-63.

Levin, B. (1993), *English Verb Classes and Alternations*, Chicago, Chicago University Press.

LIF = Bortolini, U., Tagliavini, C. & Zampolli, A., *Lessico di frequenza della lingua italiana contemporanea*, Milan, IBM Italia, 1971.

LIP = De Mauro, T., Mancini, F., Vedovelli, M. & Voghera, M., *Lessico di frequenza dell'italiano parlato*, Milan, Etas, 1993.

Lo Duca, M. G. (2000), *Proprietà valenziali e criteri di descrizione lessicografica*, in Simone, R. (ed.), *Classi di parole e conoscenza lessicale*, Special issue of "Studi Italiani di Linguistica Teorica e Applicata" XXIX (2), 219-242.

Masini, F. (2006), *Binomial constructions: inheritance*, *specification and subregularities*, in "Lingue e linguaggio" V (2), 207-232.

Masini, F. (2007), *Parole sintagmatiche in italiano* [Phrasal words in Italian], PhD dissertation, Roma Tre University.

Migliorini, B. (1957[2]), *I nomi maschili in* -a, in Migliorini, B., *Saggi linguistici*, Florence, Le Monnier, 53-108. [1st ed.: *I nomi maschili in* -a *(Appunti di morfologia italiana)*, in "Studj romanzi già editi da E. Monaci e ora da Vittorio Rossi" 25, 5-76, 1935.]

Morawski, J. (1927), *Les formules rimées de la langue espagnole*, in "Revista de Filología Española" 14, 113-133.

Nespor, M. (1993), *Fonologia*, Bologna, Il Mulino.

Peperkamp, S. (1997), *Prosodic Words*, Den Haag, Holland Academic Graphics.

Plénat, M. (1994), *L''extramétricité' des voyelles initiales*, in Lyche, C. (ed.), *French Generative Phonology: Retrospective and Perspectives*, Salford, ESRI.

Rainer, F. (2001), *Compositionality and paradigmatically determined allomorphy in Italian word-formation*, in Schaner-Wolles, C., Rennison, J. R. & Neubarth, F. (eds.), *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60th birthday*, Turin, Rosenberg & Sellier, 383-392.

Ricca, D. (2005), *Al limite tra sintassi e morfologia: i composti aggettivali V-N nell'italiano contemporaneo*, in Grossmann, M. & Thornton, A. M. (eds.), *La formazione delle parole*, Rome, Bulzoni, 465-486.

Ricca, D. (in prep.), *I composti verbo-nome in italiano* (Compo-Net PRIN).

Salvioni, C. (1902), *Rassegna bibliografica di Rob. Longley Taylor "Alliteration in Italian", New Haven, The Tuttle, Morehouse and Taylor Company, 1900*, in "Giornale storico della letteratura italiana" XXXIX, 366-391.

Scalise, S. (1983), *Morfologia lessicale*, Padova, Clesp.

Spitzer, L. (1918), *Über den Imperativ im Romanischen*, in Spitzer, L., *Aufsätze zur romanischen Syntax und Stilistik*, Halle, Niemeyer.

Spitzer, L. (1941), *L'impératif des marins*, in "Modern Language Quarterly" II, 531-550.

Spitzer, L. (1951-1952), *Sur quelques emplois métaphoriques de l'impératif. Un chapitre de syntaxe comparative*, in "Romania" LXXII, 433-478; LXXIII, 16-63.

Thornton, A. M. (2004), *Conversione A → N*, in Grossmann, M. & Rainer, F. (eds.), *La formazione delle parole in italiano*, Tübingen, Niemeyer, 508-515.

Thornton, A. M. (2005), *Morfologia*, Rome, Carocci.

Thornton, A. M. (2007a), *Phenomènes de réduction en italien*, in Delais-Roussarie, E. & Labrune, L. (eds.), *Des sons et des sens. Données et modèles en phonologie et en morphologie*, Paris, Hermès Science/Lavoisier, 241-268.

Thornton, A. M. (2007b), *Lessemi "Verbo+Verbo" in italiano e in altre lingue romanze: dalla parole alle parole*, talk presented at CILPR 2007 (Innsbruck, 3-8 September 2007).

Thornton, A. M., Iacobini, C. & Burani, C. (1997), *BDVDB. Una base di dati sul Vocabolario di base della lingua italiana*, Rome, Bulzoni.

Tomasello, M. (2003), *Constructing a language. A usage-based theory of language acquisition*, Cambridge (Mass.)/London(UK), Harvard University Press.

Vendler, Z. (1967), *Verbs and Times*, in Vendler, Z., *Linguistics in Philosophy*, Ithaca (N.Y.), Cornell University Press, 97-121.

Xu, Z., Aronoff, M. & Anshen, F. (2007), *Deponency in Latin*, in "Proceedings of the British Academy" 145, 127–143.

Zamboni, A. (1986), *Considerazioni sull'it. andirivieni e sul relativo tipo composizionale*, in Lichem, K, Mara, E. & Knaller, S. (eds.), *Parallela 2*, *Aspetti della sintassi dell'italiano contemporaneo*, Tübingen, Narr, 329-341.

The 生 *shēng/sheng* Complex Words in Chinese between Morphology and Semantics

Emanuele BANFI Giorgio Francesco ARCODIA
Università degli studi di Milano-Bicocca
emanuele.banfi@unimib.it giorgio.arcodia@unimib.it

Abstract

In the long history of the Chinese language we may find a number of complex words with the morph 生 *shēng*, meaning "to be / to be by nature", as one of their constituents. For some instances of 生 *shēng* (also in the neutral tone) in such words, this morph may be regarded as a suffix-like formative, derivating nouns (such as e.g. in 先生 *xiānsheng* "sir, mr.", lit. "born earlier", 學生 *xuésheng* "student", "study + scholar", 醫生 *yīshēng* "physician", "medicine + professional"), verbs (出生 *chūshēng* "to be born", "go out + be born") and adverbs (好生 *hǎoshēng* "quite, exceedingly", "quite + very"). There are also instances, albeit rare and non-productive, of 生 *shēng* being used as a prefix-like element, specifically as an intensifier preceding verbs of perception (as in e.g. 生疼 *shēng téng* "very painful").

Here we shall first analyze and describe, basing on historical data, the nature and the importance of complex words with 生 *shēng*; in a typological-comparative perspective, we shall then turn to the examination of "parallel" complex words of some Indo-European languages, where one of the consitutents is a grammatical / lexical form which may be traced back to one of the two I.E. roots *$h_1$es- / *s- and *b^hweh_2-, whose original meaning is "to be / to be by nature", just as the Chinese morph under examination here.

Excluding, for obvious reasons, genealogical kinship between Sino-Tibetan and Indo-European languages, the fact that they have parallel morphological phenomena hints to the possibility of the existence of "semantic universals" operating at a cognitive level.

0. Introduction

Modern Chinese, although it has often been regarded as a typical isolating language, indeed has a number of (proto-?)morphological phenomena, mainly (but not limited to) complex words made of lexical morphemes. In this paper we shall analyse, both from the diachronic and from the synchronic perspective, complex words where one constituent is the morph(eme) 生 *shēng* / *sheng* "to be (born), to be by nature", which has had several historically attested semantic values as a constituent of polymorphemic words (and, to a lesser extent, as an affix / affixoid).

In the second part of our paper we shall turn to the analysis of "parallel" word-forms attested in some Indo-European languages; those words have as one of their constituents a morph with a (nearly) analogous meaning as Ch. 生 *shēng*, which, although having different phonological (and, needless to say, ortographic) shapes for different languages, may be in all cases traced back to one of the I.E. roots *$h_1$es- / *s- // *b^hweh_2-. From the typological point of view, the interest in such comparison lies in the fact, that Chinese and I.E. languages in the field of word formation, seem to have a fundamental difference: whereas in the former language the morphematic structure of complex words is, normally, completely transparent (even for many older word-forms), mainly because of the logographic nature of the Chinese writing system, in the I.E. languages, as it is known (see e.g. Hopper 1990), it is quite common for morpheme boundaries to become opaque in the historical development of a word, thus making "originally" compound or derivate words look like simple words in the syncrhonic stage; we shall provide several examples for this in the following sections.

As we can definitely n exclude a genealogical (or contact) explanation for the existence of parallel structures in complex words from Chinese and I.E. languages, such similarity suggests the possibility of "meaning universals" or, rather, common cognitive strategies operating in the

speakers' mind when it comes to "recruiting" morphemes for complex word building (also, grammaticalization: see Heine & Kuteva 2002).

1. Theoretical background. What is a complex word?

As mentioned in the introduction, our research is centred on the analysis of *complex words*, and we shall now explain what we include under this label.

In the literature on word formation, a *complex word* is a word-form[1] containing either one lexical morpheme and a derivational affix, or more than one lexical morpheme / *subword* (Bauer 2006[2]), and, optionally, a derivational affix, including *derivatives* and *compound words*; this notion, as it is known, is a descriptive label which is synchronic in nature, i.e. applicable to a word-form caught at a certain stage of its development. Here we shall adopt a rather different approach: for the purposes of our analysis, we shall apply the notion of complex word to the diachrony of the language: what may be regarded as a simple word now might be the evolution of a complex word, made up of more than one lexical / derivational morpheme, as the following Latin adjectives:

(1) *prō-b-us* "honest" < *prō-bh-os "(one) who is, by nature, before, in a prominent position"
super-b-us "arrogant" < *super-bh-os "(one) who is by nature above".

The Latin words have the appearance of simple words, containing only one lexical morpheme. However, by enclosing the "b" between hyphens we wanted to highlight the presence of a derivative form of the I.E. root *$b^h weh_2$- "to be, to be by nature" (see **0.**), already "blended" in the root in the Latin adjectives in (1), the reconstructed "pre-Latin" forms of which are those on the right side; by comparing the Latin forms and the reconstructed ones, we may notice the structural opacity of the former ones. In the Chinese lexicon, instead, owing to the peculiar nature of the writing system and of the morphology of the language there is an almost absolute transparency in word formation, differently from what typically happens in the Indo-European languages (see Belardi 1990). Let us examine a Chinese example:

2) 渡 *dù* "to cross (a sheet of water)" + 船 *chuán* "boat" = 渡船 *dùchuán* "ferry"

Needless to say, the meaning of a Chinese complex word is not always predictable: when a word has a "deep" level of lexicalization[3] (such as e.g. 壓根 "completely", lit. "press+root"; example in Packard 2000:222), it may acquire a "non-transparent" meaning; nevertheless, the morpheme structure of the word will still be visible.

Having made clear what we mean by "complex words", we shall turn now a brief discussion of some issues concerning the application of the notion of "morphology" to the Chinese language.

2. Some remarks on multi-morphemic words in the history of Chinese

[1] On the notion of "word", see Ramat (1990 and 2005) and the recent work by Dixon & Aikhenvald (2004).

[2] The classical definition of the compound, as it is known, regards only words as possible constituents for it (see e.g. Fabb 1998); Haspelmath (2002) and prefers the label "lexeme" for the forms which take part in the creation of compound words. The definition of what consitutes a "lexeme" or a "lexical morpheme" as a complex word constituent may well depend on the language one considers (especially, depending on the morphological type, such as agglutinating *vs.* inflectional, etc.), as it was pointed out by Bauer (2006:719): "[t]he implication of this is that the forms in which the individual subwords appear may be differently defined in different languages; a citation form in one, a stem in another, a specific compounding form in yet a third, a word form in a fourth".

[3] See Packard (2000:216-225), where a categorization of five "lexicalization types" is proposed for Chinese.

As we already mentioned in the introductory paragraph, Modern Chinese *has* morphology, and the number of works inquiring into word formation phenomena in this language is, nowadays, vast[4].

What is, possibly, not yet widespread known is that Chinese has always had multi-morphemic words, even in the old stages of its development[5]; in (3) we provide an example of a sub-sillabic suffix present in Old Chinese, the*-*s* suffix, usually deriving nouns from adjectives or verbs (Baxter & Sagart 1998:54):

(3) 傳 *chuán* "transmit" < *driwen* < **drjon* →
傳 *zhuàn* "a record" < *drjwenH* < **drjon-s* (*"something transmitted")

This kind of sub-sillabic affixes, however, do not survive up to the modern language (and, in fact, they had already started deteriorating in Old Chinese). The model for morphological phenomena which "won the challenge of history", rather, is the agglutination of morphemes (mostly, monosyllabic), typically bearing lexical meaning, as we can see in this Old Chinese word (from Pulleyblank 2000:1736):

(4) 百 *bǎi* "100" + 姓 *xìng* "surname" = 百姓 *bǎixìng* "commmon people".

The spread of this typology of word-formation processes lead a linguist such as Lin (2001:62) to define Chinese as "a language of compounded words"[6].

In the huge number of complex words in Modern Chinese we may notice a few instances where one of the consituent morphemes (most frequently, the rightmost one) seems to have become grammaticalized in a derivational morpheme, showing no formal difference with the corresponding lexical morpheme (if present), but having apparently undergone some sort of semantic shift towards "bleaching" of meaning (Hopper & Traugott 2003): this seems to be the case, e.g., for 吧 *bā* "bar" (a loanword from English), the meaning of which has "faded", deriving words such as 棋吧 *qíbá* "chess club", where it means something like "meeting place". However, at present there is no agreement in the literature on the topic on whether such cases should be regarded as instances of grammaticalization[7]: here we shall not discuss the issue any further and we shall just stress the fact that the notions of "affix", "derivation" and "compounding" are controversial, and here we shall not always be using them advisedly.

3. A case study: complex word forms with 生*shēng/sheng*

Having made explicit the theoretical and empirical basis of our research, let us now turn to the analysis of our historical data on complex words containing the morpheme 生 *shēng / sheng*[8].

[4] To name but a few, Packard (1998 and 2000), Pan, Ye & Han (2004) and Dong (2002, 2004). On the possibility that a "typological drift" is under way in Chinese word formation, see Banfi (2005) and Banfi & Arcodia (forthcoming).

[5] "To speak of morphological processes in Old Chinese may surprise some readers, for there is a widespread belief that early Chinese had only impoverished morphology if it had any at all." (Baxter & Sagart 1998:35).

[6] According to the figures in Shi (2002:70-71), around 80% of the words in the Chinese lexicon are disyllabic, and the vast majority of those are, also, bimorphemic (on the genesis of disyllabic word-forms and on their supposed compound nature, see Feng 1998 and Arcodia 2007).

[7] For a summary of the issue of affixation in the Chinese linguistic bibliography, see Pan, Ye & Han 2004).

[8] Data on complex words containing the morpheme 生 *shēng/sheng* have been drawn mainly, for Chinese, from the dictionaries 漢語大詞典 *Hànyǔ Dàcídiǎn* and 現代漢語詞典 *Xiàndai Hanyǔ Cídiǎn*, from the reverse lexicon 現代漢語逆序詞典 *Xiàndai Hanyǔ Nìxù Cídiǎn* and from the corpora of the *Academia Sinica* (古漢語語料庫 *Gǔhànyǔ yǔliàokù* for Classical Chinese, 近代漢語標記語料庫 *Jìndài*

In the modern language, the character 生 *shēng/sheng* is a morph, lexical and bound (although it is worth remembering that, for Chinese, the bound-free status may depend on the micro-syntactic context, rather than being a stable feature of the morpheme itself; Zhou & Marslen-Wilson 1994); in Old Chinese, it was a free morph, a lexeme which could be used both as an independent word form and as a complex word constituent (*subword*; see **1.**). The 漢語大詞典 *Hànyǔ Dàcídiǎn* dictionary lists almost 50 semantic values for 生 *shēng/sheng*; below are the main semantic areas:

i) to give birth to, to grow, life, all one's life, living, to cause, to burn fuel;
ii) unripe, raw, rough, unknown / unfamiliar, stiff / unnatural, intensifying prefix for perception verbs / adjectives;
iii) student / pupil, intellectual, male role in Chinese opera, suffix for professionals;
iv) suffix for some adverbs.

In the following sentences (5 and 6), we may see two instances from the Classical language of a "shifted" meaning being assigned to 生 *shēng / sheng* (semantic area "i"), in its usage as a free word:

(5) "to have by nature, inborn" (天生、生來 *tiānshēng* (sky / nature + *sheng*), *shēnglái* (*sheng* + come); attested at least since the 3rd century BCE)

石生而堅, 芷生而芳
shí shēng ér jiān zhǐ shēng ér fāng
stone innate CONJ hard Dahurian angelica innate CONJ fragrant
"Stone is hard by nature and Dahurian angelica is fragrant by nature" (文字 上德 *Wénzi Shàngdé*)

(6) "occupation, living" (生業 *shēngyè* (*sheng* + occupation), attested at least since the first century BCE)

勃以織薄曲為業
Bó yǐ zhī bóqǔ wèi shēng
Bo PREP weave reeds as living
"Bo's living comes from bamboo trays weaving" (史記 絳侯周勃世家 *Shǐjì*, "Records of the Grand Historian", *Jiàng hóu Zhōu Bó shìjiā*)

As we mentioned in the introduction, in Modern Chinese *shēng / sheng* chiefly acts as a complex word constituent, a bound morph; let us see some examples of modern words containing it. The morpheme 生*shēng/sheng* may be used as a suffix(oid), bearing the meaning "student, intellectual" (7) , also forming nouns for occupations (8; meaning area "iii"):

(7)

大學生	初中生	儒生
dàxuéshēng	*chūzhōngshēng*	*rúshēng*
"university student"	"jun. sec. school student"	"Confucian scholar"

Hànyǔ biāojì yǔliàokù for early Mandarin Chinese and現代漢語平衡語料庫 *Xiàndài Hànyǔ pínghéng yǔliàokù* for Modern Chinese); Japanese data come mainly from the岩波広辞苑国語辞典 *Iwanami Kōjien Kokugo jiten* and from the reverse lexicon 逆引き広辞苑 *Gyakuhiki Kōjien* (see the references below).

(8)

醫生	陰陽生
yīshēng	*yīnyángshēng*
"doctor"	"astrologer"

We also have instances of 生*shēng/sheng* as a right-hand constitutent in complex verbs (meaning area i):

(9)

再生	派生	增生
zàishēng	*pàishēng*	*zēngshēng*
"be reborn, regenerate"	"derive"	"proliferate, multiply"

We have a few words where 生 *shēng/sheng* is used as a suffix-like formative for some adverbs (meaning area iv):

(10)

好生	怎生
hǎoshēng	*zěnshēng*
"quite, exceedingly"	"how, in which way"

The morph 生 *shēng/sheng* may also be the left-hand constituent in complex nouns, carrying the meaning "unknown" or "raw, rough" (area "ii"):

(11)

生詞	生魚	生藥
shēngcí	*shēngyú*	*shēngyào*
"new word"	"raw fish"	"unprocessed medicine"

Moreover, 生 *shēng/sheng* may be used as a (proto-derivational?) intensifying prefix(oid), only with a restricted set of perception verbs ans adjectives (area ii):

(12)

生怕	生疼
shēngpà	*shēngténg*
"(very) afraid"	"(very) painful"

Not all of these word-formation patterns, however, are equally worth investigating on, from the morphological point of view: those in 8, 10 and 12 seem to be no longer productive and, moreover, 生 *shēng/sheng* is used as an intensifier and as a constituent in adverbs (10 and 12) only in a handful of words, fully lexicalized. In what follows, we shall rather focus on complex words where 生 *shēng/sheng* seem to act as a suffix(oid) meaning "student" (see the examples in 7), as such pattern seem to be still productive nowadays; in the next section, we shall provide some historical data on this word-formation model.

3.1 Some diachronical considerations on complex words containg 生 shēng/sheng

In table 1 we summarized data on those complex words where 生 *shēng/sheng* conveys the meaning "student, intellectual" or "professional"[9], which, as we shall see in the following paragraph, fit well in the comparison with the Indo-European complex words under examination here.

[9] The main sources for the data were the dictionaries 漢語大詞典 *Hànyǔ Dàcídiǎn* and the 古漢語大詞典 *Gǔhànyǔ Dàcídiǎn*, as well as web corpora (see footnote 8).

Table I. Some complex words with 生 *shēng/sheng* and their attested meanings.

Period → Word ↓	**4th – 3rd cent. BCE**	**漢 *Hàn* dinasty (206 BCE– 220 CE)**	**南北朝 *Nánběi Cháo* - Northern and Southern Dynasties (420-589)**	**唐 *Táng* dynasty (618-907)**	**宋 *Sòng* dynasty (960-1279)**	**明 – 清 *Míng* and *Qīng* dynasties (1368-1644 and 1644-1911)**	**From 1912 to the present day**
先生 ***xiānsheng***	The elders / father and elder brothers (論語 *Lùnyǔ*); Senior erudite / scholar (孟子 *Mèngzǐ*)。	Teacher, master (禮記 *Lǐjì*); Literate / scholar (史記 *Shǐjì*); Title for physiognomists, diviners, singers, medical practitioners and *Feng Shui* experts (史記*Shǐjì*)。					Antiquate term for people doing secretariat and supervision jobs (三元裡人民抗英斗爭史料 *Sānyuánlǐ rénmín kàng Yīng dòuzhēng shǐliào*); Used for men of high position in some field or area (解放日報 *Jiěfàng Rìbào*).
學生 ***xuéshēng / xuésheng***			Student in a school (后漢書 *Hòu Hàn shū*).				
廩生 ***lǐnshēng***						Student supported by the state (醒世姻緣傳 *Xǐngshì yīnyuán zhuàn*)	
監生 ***jiànshēng***						Student in the Imperial Academy (金瓶梅 *Jīnpíngméi*)	
門生 ***ménshēng***			Second-generation disciple (后漢書*Hòu Hàn shū*); Servant, follower (后漢書*Hòu Hàn shū*).	Humble term for candidates for the Imperial exams (唐白居易 *Táng bái jū yì*).	Humble term for reccomended candidates for the Imperial exams (朝野類要 升轉 *Cháoyě lèiyào shēngzhuǎn*).	Assistant to a high officer / advisor of a powerful person (禮部尚書張公墓志銘 *Lǐbù shàngshū Zhānggōng mùzhì míng*)	
醫生 ***yīshēng***				An official's aide who is engaged in medical studies (唐典太醫署 *Táng liùdiǎn tàiyī shǔ*).	A person with a knowledge of pharmacy whose activity is healing (書事 *Shūshì*).		
新生 ***xīnshēng***	Newborn (莊子 *Zhuāngzǐ*)	Beginner Confucian scholar (風俗通 皇霸三王 *Fēngsútōng Huángbà sānwáng*).					Newly enrolled student (光明日報 *Guāngmíng rìbào*)
書生 ***shūshēng***			Confucian scholar, student, scholar (后漢書*Hòu Hàn shū*).				

The path which led to the acquisition of the "student, intellectual" meaning for the morpheme 生 *shēng/sheng* is quite peculiar, and appears to be very different from the one of the I.E. roots which shall be discussed here. According to Jiang (2005 [1989]:90), the meaning of 生 *shēng/sheng* developed from "to be born, to grow" (meaning area i, **3.**) to "student" through compound shortening / truncation (簡縮 *jiǎnsuō*): "By 'shortening' we mean the expression of a meaning, which originally was proper of a compound or phrase, by a single morpheme (a Chinese character)" (my translation). By the process outlined above, the morph 生 *shēng/sheng* takes on the "new" meaning out of having been a constituent in a complex word which, originally, posessed that semantic value. Of course, this happens also in I.E. languages, albeit maybe not so often:

(13) **Homo**sexual → **homo**phobia

In the example above, a morpheme *homo-*[10] "one, same", becomes a "substitute" for "homosexual": we could analyse those instances in several different ways (treating them as analogical formations, as clipping, etc.), but such a discussion is far beyond the aims of the present paper.

Let us illustrate the steps by which the value of 生 *shēng/sheng* evolved from "to be born" into "student" in Jiang's account (2005 [1989]:90):

(a) In the 秦 *Qín* (221 - 226 BCE) and 漢 *Hàn* (206 BCE – 220 CE) dynasties, 生*shēng* is used as the short form for 先生 *xiānshēng*, lit. "born earlier", a respectful address →

(b) in the *Mencius* (孟子 *Mèngzǐ*, 3rd cent. BCE), 先生 *xiānshēng* is used in the sense of "senior erudite / scholar" (cfr. table I) →

(c) the short form 生 *shēng* later (at least since the 史記 *Shǐjì*, 1st cent. BCE) becomes equivalent to "(Confucian) scholar, student", no matter what the age was→

(d) this meaning is preserved until today: 畢業生 (degree+*shēng*) "graduate student", 招生 *zhāoshēng* (recruit+*shēng*) "recruit students".

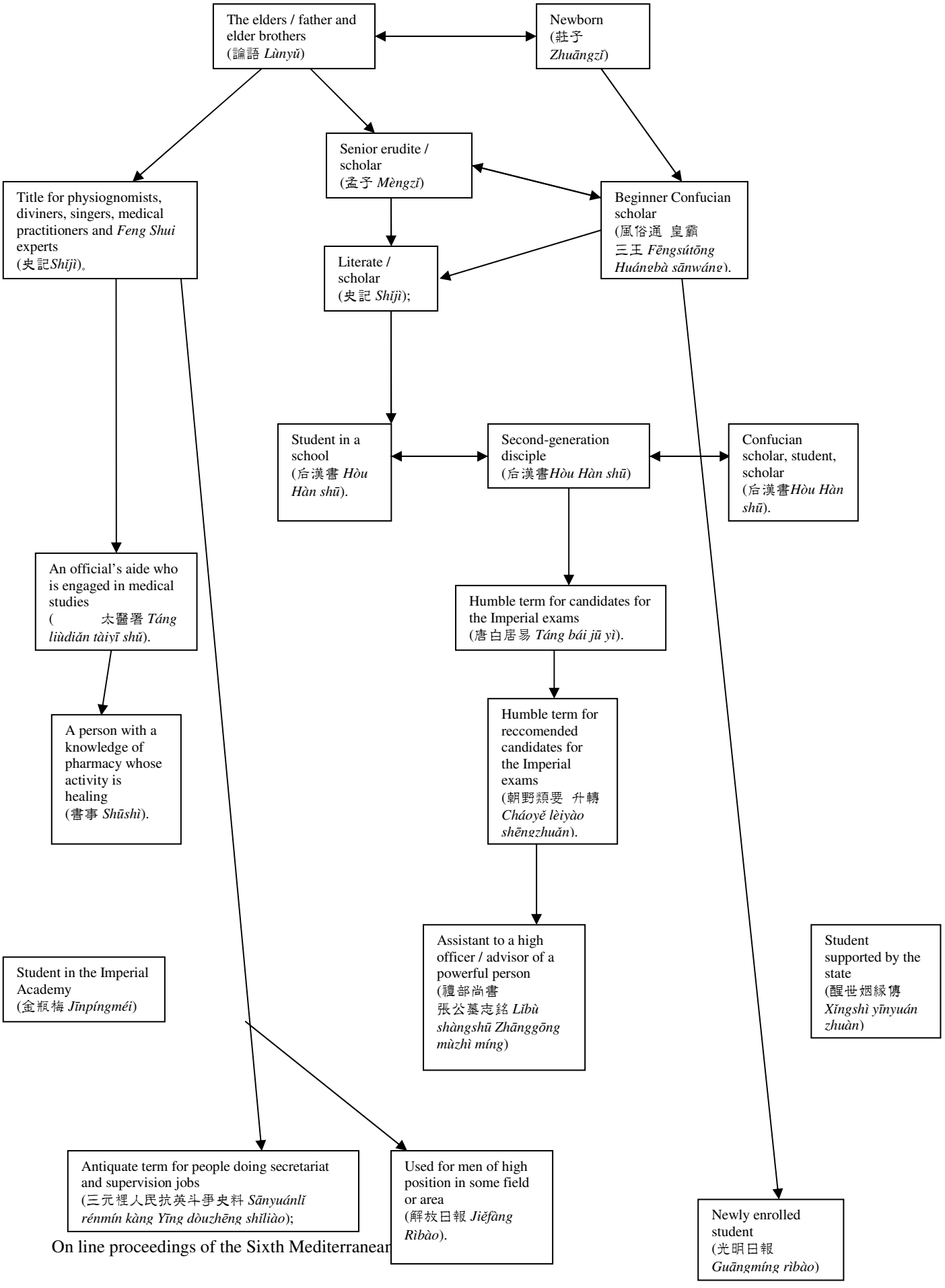
To sum up:

i) **To be born > to be born earlier / be older**
> ii) **person deserving respect**
> iii) **scholar, medicine man, diviner**
> iv) **student, doctor**

As the evolution of 生 *shēng*/*sheng* was driven by the complex words in which it appeared as a constituent, as we made clear above, it is interesting to analyse the history of those words. In table II, we provide a two-dimension representation of the semantic shifts which occurred in the 生 *shēng*/*sheng* complex words: along the vertical axis, historical relationships among different occurrences of a single word (or of "cognate" words) are shown, whereas the arrows (horizontal axis) make the semantic connections between different words at the synchronic level visible.

[10] From the Ancient Greek prefix ὁμο-, connected to the adjectival lexemes ὁμος / ὁμοῖος "same, similar" (see DELI 1999).

Table II. Some complex words with 生 *shēng/sheng* and their attested meanings.



3.2 Some remarks on the Sino-Japanese complex words containg 生 shēng/sheng (sei)

To conclude our discussion on the evolution of 生 *shēng/sheng*, we shall provide some data on the "destiny" of the corresponding Sino-Japanese morpheme (pronounced *sei*).

According to the 岩波広辞苑 *Iwanami Kōjien* Japanese dictionary, 醫生 *yīshēng* "doctor" (see table I) in Japanese stands for "medicine student" (pron. *isei*). In the Japanese lexicon, there is a number of complex words where 生*sei* has the meaning of "student"; below are some examples:

(14)	院生	越年生
	insei	*etsunensei*
	"postgraduate school student"	"second year student"
(15)	寄宿生	寮生
	kishukusei	*ryōsei*
	"boarding student"	"dormitory student, boarder"
(16)	給費生	自費生 / 私費生
	kyūhisei	*jihisei* / *shihisei*
	"fellowship student"	"self-supporting student"[11]

At a first glance, one might think that 生 *shēng* / *sei*, in Japanese, has undergone a grammaticalization process: this morph, in fact, appears in suffixal position in a series of complex words with a consistent meaning. Let us postpone the discussion on this to the concluding remarks; in the following paragraph, we shall introduce our Indo-European complex words data.

4. Parallel complex words in the Indo-European languages

In the preceding paragraph we illustrated the diachronical evolution of the Chinese morpheme 生 *shēng/sheng* "to be / to be by nature", including instances where, apparently, this morph has (partially?) grammaticalized into an affix(oid) (see examples 7, 8, 10, 12).

What has been observed in Chinese complex word formation, exemplified by the morpheme 生 *shēng/sheng*, lead us to make a comparison with similar developments in the I.E. languages, where there is a relevant number of word formation patterns, including a constituent semantically connected to the idea of "to be / to be by nature" (see par. **1.** for our definition of "complex word").

In I.E. languages, the most interesting case is that of some complex words, the structure of which may be traced back to a matrix "lexical root + V", which in some linguistic traditions go back to participial forms. In their semantic quality of "verbal nouns", participles share formal features of the nominal and verbal categories, thus representing a particularly interesting form of "transition" between these two points of the semantic *continuum*. An example for this could be the Latin word *studēns, studēntis* < *stud-ent-s, stud-ent-is: it is formed by a lexical root *stud- "to study" with a suffix *-ent, derivated from an allomorphic form of the I.E. root *es- / *$h_1$es- "to be, to exist". In other words, from a semantic perspective the structural matrix of Lat. *studēns, studēntis* is "the one who is in the dimension of studying". This structural matrix is

[11] Compare the words in 16 to Chin. 公費生 *gōngfēishēng* "fellowship student" (lit. "student at public expenses) and 自費生 *zìfēishēng* "self-supporting student".

actually analogous, in the synchronic perspective, to the Chinese "學生 *xuéshēng*-type" constructions.

Based on these remarks, we shall now examine a series of complex words drawn from I.E. languages that involve a morpheme in the semantic area of "to be / to be by nature".

As widely known, from the very beginning of research in Indo-European linguistics, comparison within members of the family made it possible to recognize two very productive roots, both conveying the meaning of "to be / to exist by nature": these roots are *$h_1$es- / *s- and *b^hweh_2-, diffused throughout I.E. languages and may be exemplified by the double forms (jointly expressed in a paradigm) such as Skr. (*ásti / bháviti*), Gr. (ἐστί / ἔφυ), Lat. (*est / fuīt*), in Germanic languages (Ger. *ist / bin*; Engl. *is / to be*), in the Slavic group (Russ. *jest' / byt'*), and, as a single form, in the Baltic languages (Latv. *būti*). In this section we shall analyse the occurrence of such roots (including their allomorphs) in word formation; in many of these cases, it is appropriate to talk about grammaticalization, albeit sometimes only partial.

First of all, it should be noticed that the two roots under examination here, as far as the formation of allomorphic paradigms of "to be" (such as Lat. *esse / sum / fuī*), the existential verb *par excellence*, are widely attested in I.E. languages through a number of word-forms which were originally participles and later evolved to the *status* of (somewhat) complex autonomous words, such as:

- Gr. τὰ ὄντα "the things which exist > truth", Lat. *entia* "existing things", derived from a participle form with apophonic variation (*ent- / *ont-);

- Hitt.. *ašant-* "existing > true, real", Skr. *sánt-* "existing, real", Av. *hant-* "existing > true, real", Lat. *sōns, sont-is* "the one who is (at the centre of a judgement) > guilty", O. Icel. *sannr* "existing > true", O.Engl. *sōð* "existing > true, actual". The Latin verb *sentire* (lit. "to identify oneself with sthg." > "to feel") is quite interesting since it involves a special case of morphological redundancy: it is derived from a participle form (with apophonic variation) *(h_1e)s-ent / *(h_1e)s-ont-, in which the base form of the verb "to be" is followed by a morpheme connected with the the present participle form of the very same verb (compare Lat. *essentia*; *4.1*).

In what follows, we shall analyse instances of (partial?) grammaticalization of the I.E. roots discussed above, providing examples of their usage (or, rather, of the usage of a form derived from them) as complex word constituents.

4.1 The I.E. roots *$h_1$es- / *s- and *b^hweh_2- in word formation

Participle forms (with apophonic variation) *(h_1e)s-ent / *(h_1e)s-ont- and *ent- / *ont- combined with prepositions yield nouns such as Lat. *ab-sent-ia* "assenza", *prae-sent-ia* "presenza" (related to the corresponding verb forms *abesse, praeesse*), Gr. παρουσία "presence" (< παρά + *ὀντ-ϳα, realated to the verb παρειμί). In combination with verbal roots of the *(e)d-ont- / *(e)d-ent- type, one finds words such as Hitt. *adant-*, Gr. ὀδούς, ὀδόντος, Church Sl. *ěd-ęt,* Skr. *da(n)t,* Avest. *dantan-,* Lat. *dēns, dēntis*, O. Ir. *dét,* Goth. *tunþus,* O. H. Germ. *zand,* Latv. *dantis* "tooth". The structural matrix behind these words consists of a lexical root "to eat" + *ent- / -ont- (a morpheme of the present participle of the verb "to exist (by nature)") > "that is eating". Another interesting case is this peculiar instances of morphological redundancy: Latin *essentia* (< *esse + *-ent-ja, derived from an unattested participle **essēns*, following the lexical pattern of *patientia* < *patiens, sapientia* < *sapiens*) and *sententia* (instead of the expected **sentientia,* derived from *sentiō* or, rather, from a participial base *sent- + *-ent-ja).

From the root *b^hweh_2-, Latin has developed an intervocalic morpheme –b- (regular derivation *-bh- > -b-), which may be found in a series of adjectives formed by:

a) preposition + *-b-, such as *prō-b-us* (< *prō-bh-os, lit. "(one) who is, by nature, before, in a prominent position" > "honest"; compare. Ved. *pra-bhúḥ* "eminent, prominent"), *super-b-us* (< *super-bh-os, lit. *super-bh-os "(one) who is by nature above" > "arrogant");

b) nominale/adjectival base + *-b-, such as *acer-b-us* (<*akro-bho-s < *ak- "to be sharp", lit. "which is sharpened (by nature) > sour".

This pattern is also found in a high number of Greek adjectives (where *-bh- > *-ph-). In such structures the first element may be on of the following:

a) a preposition, as in ἐμ-φυ-ής "innate, natural" and προσ-φυ-ής "naturally attached";
b) an adjective, as μεγαλο-φυ-ής "of noble nature";
c) an adjective acting as an adverb, such as in εὐρυ-φυ-ής "vast + to be (by nature) > developing";
d) an adverb, as in the adjective δι-φυ-ής "dual, double in nature";
e) a prefixoid, such as in the adjective αὐτο-φυ-ής "born by oneself, natural".

It is also worth noticing that the morph *-b- may be found also in the verbal paradigm of Latin and Italic languages, both as an aspect marker indicating past imperfective event (the type of Lat. *am-ā-b-a-m, mon-ē-b-a-m*, with its Oscan equivalent *fu-fa-ns* "they were") and as a marker of imminent / progressive event in Latin future verb-forms (the type of *am-ā-b-ō, mon-ē-b-ō*, the Faliscan equivalent of which is *pi-pa-fo* "I will drink").

The primary meaning of the morpheme *-ent- / *-ont- (and of its allomorph *-(V)nt-, zero degree of the root) must have been something like "who / what is (by nature)": later on, with a rather predictable semantic shift, it came to mean "related to, connected to" > "belonging to". This semantic variation may explain some features of a set of I.E. complex words: among these, we find collective / plural proper names in Anatolian (the Hittite type *utneyant-/udneyant-* "population of a region" < *utne/udne* "region"; *tuzziyant* "troops, army" < *tuzzi* "troop") with possible continuations, according to Solta (1958: 8-9) – following Kretschmer's and Oikonomides's thesis – in the plural / collective forms found in the Modern Greek dialects of Ponto (such as βρεχή, ἡ "rain" / βρεχάντας, τά "rains", χαρά, ἡ "joy" / χαράντας, τά "joy, wedding".

The most visible among the semantic values of one of the forms of the morpheme *-ent- / *-ont- / -*(V)nt- is that of participle marker (active, middle and passive voice), which can be considered to be located at the boundaries between the categories of adjectives proper and verbal adjectives. According to Adrados, Bernabé & Mendoza (1996: 321-323), there are serious problem in estabilishing clear boundaries betweeen the above mentioned categories: some ancient Latin denominative adjectives formed with the morpheme *-(a)nt-* (from the formal point of view, present participles of Classical Latin) used in poetry are particularly interesting in this respect, such as e.g. *animāns* "endowed with the spirit of life", *stellāns* "provided with stars", *comāns* "provided with hair", being in fact transitional between verbs and nouns / adjcetives (see also Hofmann, Szantyr & Leumann 1972: 589).

Yet, the most important semantic features of the morpheme *-ent- / *-ont- / -*(V)nt-, as a participle marker, is the clear insensitivity to diathesis. This typologically very significant feature occurs in many I.E. languages:

In Buddhist Sanskrit one finds several passive participles, meaning conveyed by the active suffix *-(a)-nt-(a)* < *-(V)nt-: it is as such words as *dahy-ant-as* (nom. sing. masvc.) "which is burned" semantically analogous to *dahyamānas*; the same happens in the Prakrit languages.

In Iranian, as in Sanskrit, the derivatives of the suffix *-(V)nt- normally have an active, transitive or intransitive meaning, depending on the verb to which they are agglutinated; yet, just as in Sanskrit, one finds visible traces of insensitivity to diathesis of this suffix in Avestic, as evidenced words like Av. *haošy-ant-a* "that ought ot be squeezed" (Solta 1958: 42).

In Greek, just as in Sanskrit and in Avestic, the continuations of the suffix *-(V)nt- > -(V)ντ- form participles with active diathesis (present, future, aorist, like in λύ-ο-ντ-, λύ-σο-ντ-, λύ-σα-ντ-), whereas for the mediopassive (present, future, passive future, middle aorist,

perfect) one finds the continuations of the suffix *-meno-: the types of λυ-ό-μενο-ς, λυ-σό-μενο-ς, λυ-θη-σό-μενο-ς, λυ-σά-μενο-ς, λε-λυ-μένο-ς. Such a division generally applies, and the suffix *-(V)*vτ*- normally has active, transitive or intransitive meaning, depending on the nature of the verb to which it is attached. However, an original insensitivity towards diathesis emerges from the fact that the suffix involved plays a role *also* in the formation of the aorist passive participle (see the model -(θ)ε-ντ-: λυ-θ-έ-ντ- > λυθεῖς) and in ancient noun forms in which the suffix *-(V)*vτ*- is clearly used with a middle or even passive meaning: see Chantraine (1990: 218) on Gr. *γέϱ-ο-ντ- > γέϱων, γέϱοντος "that has grown old, made to grow old > old" comparable to Skr. *járant-*, Ossetian *zärond*.

In Hittite, the continuations of the suffix *-(V)nt- (besides forming plural / collective nouns, which we already mentioned) yield the only participle of the originally diathesis indifferent verb (Benveniste 1984: 126).

In Latin, the morpheme -(V)*nt*- participates in the formation of present participle forms both of active (such as *dīcēns, dīcentis*) and deponent verbs (*sequēns, sequentis*), with an active-transitive or intransitive meaning. Latin has no present mediopassive participle deriving from the suffix *-m(e)no- for the expression of the aspectual notion "not completed", whereas the passive and deponent past participles are supposedly derived from the ancient verbal adjcetive *-(C/V)to- (the types represented by *fac-tu-m* and *mon-i-tu-m*). Ancient participle forms of the type *e-vid-ens* "that is seen", *in-fans* "that cannot be told" (translating Gr. ἄῤῥητον: compare. Acc. 189 *infāns facinus*; Lucil. 140 *facta nefantia*), *gignentia* "plant, growth" which renders Gr. τὰ φυόμενα. The usage of the participle *-(V)*nt*- with an intransitive-reflexive or, also, passive meaning (e.g. *amantissimus* vs. *amatissimus* and *desiderans* vs. *desideratus* found in funeral inscriptions already during the Republican era and present throughout Medieval Latin epigraphs) continues throughout the entire Latin era and it is also attested in later authors, as one may see from the many examples in Hofmann, Szantyr e Leumann (1972: 387-390).

In Celtic, the category of present participle is absent from the verbal system. However, the diachronic analysis of Celtic data confirms that the continuations of the participle suffix *-(V)nt- could not only combine with active bases, but could also be a part in middle or even passive construction, being insensitive to diathesis: See e.g. the O. Ir. adjective *té/tee* "hot" (< *tep-ent-s; compare. Skr. *tapant-*, Lat. *tepens*), the O. Ir. nouns *care/carae* (masc. sing.) "friend" (< *kar-ā-nt-s, lit. "the one who loves"; compare Welsh *Carantus, Carantius, Carantillus,* etc.), *bragae* "prisoner" (< *brag-a-nt- "the one who is being guarded"; compare Goth. *bairgan* "to guard", Russ. *beregu* "I am guarding"), where a participle *-(V)nt- had an evident passive meaning

In Germanic and Tocharian one also finds strong insensitivity to diathesis in participle forms that continue the suffix *-(V)nt-: in Old Nordic, participles in *-(V)nt- may be found only in popular prose, with intransitive verbs (Solta 1958: 26-27; Evangelisti 1965: 158). In Gothic, one finds a single case of present participle in *-(V)*nt-,* with mediopassive meaning, which is *aland* < *alan* "to feed, to nourish" (translating Gr. ἐντϱεφόμενος of Timoth. I, IV, 6); analogously, we have Goth. *þai daupjandans* (corresponding to Greek οἱ βαπτιζόμενοι of K 15, 29): the verb on which it depends – *daupjan* "to baptize" – is clearly transitive. In both Tocharian dialects the continuation of the suffix *-(V)nt- (Toch. A: *-nt,* Toch. B: *ñca*) form the active present participle, as opposed to the middle present participle in *-meno- (Toch. A: *māṃ*; Toch. B: *-mane*). The semantic values yield complete indifference to diathesis also in Tocharian participles. To be sure, Tocharian has collective nouns attesting the continuation of the suffix *-(V)nt-, marking plural (Toch. A *akär / akrunt* "tear / tears", *tärkär / tärkrunt* "cloud / clouds").

Insensitivity to diathesis in participle formations with the *-(V)nt- morpheme is a typologically relevant feature that hints to the presence of remnants of a grammaticalization process of units in the morphological component in an ancient phase of the Indo-European linguistic tradition, if considered side by side with the morphemes present in some complex words (in Vedic, Greek, Latin) which are the continuation of the root *$b^h weh_2$-: it is a very ancient phase, in which the I.E. system yielded agglutinating phenomena, with a high degree of semantic transparency in word formation.

4.2 Summary and (provisional) conclusions

The comparative analysis of complex words in Chinese containing the morpheme 生 *shēng* / *sheng* and of the continuations in I.E. languages of morphs deriving from the roots *$h_1$es- / *b^hweh$_2$- (excluding, of course, any genealogic realationship between the two linguistic areas) suggests, thus, three facts:

a) a process of progressive enrichment of word formation in Chinese, which is an isolating system that has, partially, acquired typical agglutinating features. In such a system, however, in complex word formation the elements involved maintain their shape unaltered and, consequently, allow for a high level of semantic transparency;

b) the parallel process of the increasing in morphological complexity in the I.E. system, with its apparently early grammaticalization features and where individual morphemes that acquired a high degree of formal opacity, following the loss of (formal) connection to their original form, have acquired a high degree of formal opacity. This loss of transparency was (is) at such a high level that only the historical linguistic reconstruction yields insights into its underlying mechanisms;

c) from the methodological point of view, the observation of "parallel" morphological strategies documented in typologically and diatopically distant linguistic systems ultimately allows for hypothesizing the presence of semantic universal features operating at the cognitive level, in the speakers' linguistic awareness, leading to other fascinating research topics.

5. Concluding remarks

5.1 Grammaticalizazion and "typological drift" in Chinese

A general assessment of the typological status of the morphology of Chinese is obviously beyond the scope of this talk. Nevertheless, we may take the data presented here as an example of the kind of development which characterizes Modern Chinese.

A "typological drift" toward the agglutinating type, i.e. the acquisition of agglutinating features, may be seen, above all, in the domain of word formation; in the modern language there is a tendency to the development of word formation patterns with "hybrid" properties, which possess both features of isolating morphology (such as the 1:1 ratio between morph and lexical morpheme) and characteristics of the agglutinating "word type" (words made of long, clearly segmentable, morpheme chains).

It is still questionable whether word formation elements such as 生 *shēng/sheng* should be regarded as lexemes or as affixes. In Chinese, the origin of the "suffix-like" 生 *shēng/sheng* is rather peculiar, being born out of the abbreviation of a compound (先生 *xiānsheng*, see above), similarly to Engl. *–gate* or it. *–poli* "scandal related to X", a somehow productive suffixoid which was created by analogy with the word *tangentopoli* "tangent town" (used in reference to Milan in the early nineties, later to indicate the system of corruption in Italian politics and the scandal which followed its unveiling).

The question as to whether 生 *shēng/sheng* has undergone grammaticalization requires a "split judgement": as a suffix-like formative used in nouns for occupations and in its usage with some adverbs (exx, 4 and 6), as well as in its prefix-like identity (with intensifying value) for V and Adj (ex. 8), we may talk about "aborted grammaticalization", as the number of words which are built according to these patterns is negligible, and they are no longer productive; we might speak of grammaticalization / morphologization when 生 *shēng/sheng* is used as a suffix-like formative meaning "student" (ex. 3), where we have both a specialization of meaning and (limited) productivity. The grammaticalization of 生 *shēng/sheng* in a suffix(oid) meaning "student" seems more evident in Japanese, where we have a consistent series of complex words with such morph in suffixal position and with a stable meaning.

5.2 A comparative outlook

In its diachronical evolution, Chinese morphology (specifically, word formation) has shifted from processes modifying the syllable to *word formation schemas* (Booij 2006) based on the agglutination of lexical morphemes; we may also regard some of these lexemes as derivational morphemes, i.e. grammatical morphemes;

The transparency of complex words in Modern Chinese is very high; the main "challenge" to this feature is the production of abbreviated forms which generate "new morphemes" (Packard 2000) →

(17)	勞動	保險	勞保
	láodòng	*bǎoxiǎn*	*lǎobǎo*
	"labour+act"	"protect+risk"	"labour+protect"
	"labour"	"insurance"	"labour insurance"

In the case of 生 *shēng/sheng* used as a suffix(oid), meaning transparency is ensured by the "paradigmatic" relationship with other complex words in the same "word family" (see Booij 2006).

The opposite occurred in the Indo-European system, where, as we have seen, agglutinating features in word formation are not "active" anymore, and early grammatical morphs have become fused as part of the lexical root, thus making the structure of the word opaque, whose original shape is of interest only for etymologcal research.

Having found two morphemes with a common original meaning and similarities in the present semantic value in two unrelated (and typologically distant) language families, one is tempted to hypothesize the action on "meaning universals"; such an explanation, which can hardly hold in the lexical domain, seems to be valid for the evolution of lexical morphemes into grammatical ones (see Heine & Kuteva 2002). In the latter case, one might well put forward the hypothesis that lexemes meaning "to be (born), to be by nature" are a typical source of grammatical morphs meaning "to be in the state / activity of". Of course, this is to be regarded as a "clue", and should be tested on a proper language sample; it is also worth remembering once more that the "student" meaning for 生 *shēng/sheng* has been acquired from the truncation of a compound, and thus it may not easily fit into the comparison with the Indo-European roots investigated here. Although…

Abbrevations used in the glosses

CONG.: conjunction
PREP.: preposition
PTC.: structural particle (possessive, relative)

Selected references

F. Adrados, A. Bernabé and J. Mendoza, Manual de Lingüística indoeuropea. II. Morfología nominal y verbal, Ediciones Clásicas, Madrid, 1996.
P. Baldi, The Foundations of Latin, Mouton de Gruyter, Berlin-New York, 1999.
W.H. Baxter and L. Sagart, Word formation in Old Chinese, in J.L. Packard (ed.), New approaches to Chinese Word Formation, Mouton de Gruyter, Berlin-New York, 1998, pp. 35-75.
W. Belardi, Linguistica generale, critica e filologia del testo, Bonacci, Roma, 1990.
É. Benveniste, Origines de la formation des noms en indo-européen, Maisonneuve, Paris, 1984.
G. Booij, Inflection and derivation, in K. Brown *et alii* (eds.), Encyclopedia of language and linguistics, Elsevier, Oxford, 2006, pp. 654-661 (V vol.).
P. Chantraine, La formation des noms en grec ancien, Klincksieck, Paris, 1979 [1933].
E. Evangelisti, Ricerche sul suffisso -NT- di participio, Acme 18, 1965, pp. 205-22 [Reprinted in E. Evangelisti, Scritti tocarici e altri studi, Brescia, 1990, pp. 139-161 (quotations from the reprint)].
T. V. Gamkrelidze and V. Ivanov Vjačeslav, Indo-European and the Indo-Europeans. A Reconstruction and Historical Analysis of a Proto-Language and a Proto-Culture, Mouton de Gruyter, Berlin-New

York, 1995.
M. Haspelmath, On Directionality in Language Change with particular Refence to Grammaticalization, in O. Fischer, M. Norde and H. Perridon (eds.), Up and down the Cline – The nature of Grammaticalization, Benjamins, Amsterdam-Philadelphia, 2004, pp. 17-44.
B. Heine and T. Kuteva, World Lexicon of Grammaticalization, Cambridge University Press, Cambridge, 2002.
P. J. Hopper Paul, Where do words come from?, in W. Croft, D. Keith and S. Kemmer (eds.), Studies in Typology and Diachrony. Papers presented to Joseph H. Greenberg on his 75th birthday, John Benjamins, Amsterdam-Philadelphia, 1990, pp. 151-160.
P. Hopper and E. Traugott, Grammaticalization, Cambridge University Press, Cambridge, 1993.
C. Lehmann, Thoughts on Grammaticalization, Lincom Europa, München, 1995 [1982].
H. Lin, A Grammar of Mandarin Chinese, Lincom Europa, München, 2001.
F. Mawet, Les participes indo-européens (*-nt-), in G. Bolognesi and V. Pisani (eds.), Linguistica e Filologia. Atti del VII Convegno internazionale di linguisti (Milano, 12-14 IX 1984), Paideia, Brescia, 1987, pp. 357-374.
G. Meiser, Historische Laut- und Formenlehere der lateinische Sprache, Wissenschaftliche Buchgesellschaft, Darmstadt, 1998.
A. Meillet and J. Vendryes, Traité de Grammaire comparée des langues classiques, Paris, 1979[5].
J. L. Packard, The Morphology of Chinese. A Linguistic and Cognitive Approach, Cambridge University Press, Cambridge, 2000.
W. Pan, B. Ye and Y. Han, 汉语的构词法研究 (*Hanyu de goucifa yanjiu* – Research on word formation in Chinese), Huadong Shifan Daxue Chubanshe, Shanghai, 2004.
E. G. Pulleyblank, From Archaic Chinese to Mandarin, in G. Booij, C. Lehmann and J. Mugdan (eds.), Morphologie-Morphology, Walter de Gruyter, Berlin-New York, 2000, pp. 1730-1740.
R. Solta, Gedanken über das nt-Suffix, Österreische Akademie der Wissenschaften, Phil.-Hist. Kl. 232/1, 1957, pp. 3-47.
C. Watkins, Il proto-indoeuropeo, in A. Giacalone Ramat and P. Ramat (eds.), Le lingue indoeuropee, Il Mulino, Bologna, 1993, pp. 43-93.

Dictionaries and corpora

CCD (2002), The Contemporary Chinese Dictionary - 现代汉语词典 (*Xiandai Hanyu cidian*), Foreign Language Teaching and Research Press, Beijing.

GHKJE (2002), 逆引き広辞苑 (*Gyakuhiki Kōjien –Kōjien* reverse order Japanese dictionary), Iwanami Shoten, Tokyo (electronic edition).

GHYDCD (2000), 古汉语大词典 (*Guhanyu Dacidian* – Great dictionary of Classical Chinese), Shanghai Cishu Chubanshe, Shanghai.

HYDCD (1991), 汉语大词典 (*Hanyu Dacidian* – Great dictionary of Chinese), Hanyu Dacidian Chubanshe, Shanghai.

INKGJT (1994), 岩波国語辞典 (*Iwanami Kokugo jiten* – *Iwanami* Japanese dictionary), Iwanami Shoten, Tokyo.

NXCD (2005), 现代汉语逆序词典 (*Xiandai Hanyu nixu cidian* – Modern Chinese reverse order dictionary), Yanbian Renmin Chubanshe, Yanji.

Academia Sinica corpus of Old Chinese (古漢語語料庫 *Gǔhànyǔ yǔliàokù*): http://www.sinica.edu.tw/~tibe/2-words/old-words/.

Academia Sinica tagged corpus of early Mandarin Chinese (近代漢語標記語料庫 *Jìndài Hànyǔ biāojì yǔliàokù*): http://www.sinica.edu.tw/Early_Mandarin/.

Academia Sinica Balanced Corpus of Modern Chinese (現代漢語平衡語料庫 *Xiàndài Hànyǔ pínghéng yǔliàokù*): http://www.sinica.edu.tw/ftms-bin/kiwi.sh?ukey=344276996&qtype=1.