

**FIRST  
INTERNATIONAL  
SYMPOSIUM OF  
MORPHOLOGY**

Premier Symposium  
International  
de Morphologie



# iSMo 2017

PREFIX

DERIVATION

MORPHOLOGY

COMPOUNDING

SUFFIX

TYPOLOGY

University  
of Lille  
France

**DECEMBER  
13-15<sup>TH</sup>  
2017**

LILLE

<https://colloque-ismo.univ-lille3.fr>



# Livret d'abstracts

# Booklet of abstracts

Copyright © 2017 International Symposium of Morphology (ISMo)

Permission is granted under a Creative Commons license to replicate and distribute this report freely, for noncommercial purposes. That attribution is provided as illustrated in the citation below. To view a copy of this licence, visit: <http://creativecommons.org/licenses/by-nd/4.0/>

**Citation:** International Symposium of Morphology (2017). Booklet of Abstracts, 13th – 15th December 2017, Lille, France.

A full-text PDF version of this report is available to download from <https://colloque-ismo.univ-lille3.fr/fr/static13/livret-colloque-ismo-2017>



## Table of contents

Presentation.....	1
Committees.....	2
Programme .....	3
DECEMBER 13 <sup>TH</sup> .....	3
DECEMBER 14 <sup>TH</sup> .....	3
DECEMBER 15 <sup>TH</sup> .....	4
Posters.....	5
List of Abstracts .....	7





## Presentation

The International Symposium of Morphology (ISMo) takes over from two previous scientific events: the Forum de Morphologie, organized every two years alternatively in Lille and in Toulouse between 1997 and 2002, and the also biennial conference Les Décembrettes, organized in Toulouse (with Lille for its last edition) from 2004 to 2016.

The creation of this new event, ISMo, has been decided at the last Décembrette's edition, in order to enlarge the number of French universities implied in its organization and to make visible the vitality of morphology in France. The biennial conference ISMo shares the aim of its predecessors: being a regular event of high scientific level where one can present results or discuss data and analyses.

\*\*\*

La présente manifestation scientifique « International Symposium of Morphology » (ISMo), dont c'est ici la première édition, prend la suite de deux manifestations précédentes : le Forum de morphologie, organisé en alternance à Lille et à Toulouse et qui a connu trois éditions de 1997 à 2002, et les Décembrettes, organisées à Toulouse selon une périodicité bisannuelle de 2004 à 2016.

Le principe de la création d'ISMo a été décidé lors de la dernière édition des Décembrettes en décembre 2016, afin d'élargir le nombre d'universités partenaires et de rendre visible à l'international la vitalité de la morphologie sur l'ensemble du territoire français. ISMo poursuit le même objectif que ses prédecesseurs : constituer un événement bisannuel de haut niveau scientifique où faire se rencontrer morphologues français et de l'international autour de résultats, de données ou d'analyses.

## **Local Organizing Committee**

Dany AMIOT, Georgette DAL, Delphine TRIBOUT (UMR Savoirs, Textes, Langage, CNRS et Université de Lille, France)

## **Permanent Committee**

Dany AMIOT (STL / Université de Lille, France), Olivier BONAMI (LLF / Université Paris Diderot, France), Gilles BOYE (CLLE-ERSS / Université de Bordeaux, France), Georgette DAL (STL / Université de Lille, France), Bernard FRADIN (LLF / Université Paris Diderot, France), Hélène GIRAUDO (CLLE-ERSS / Université Jean Jaurès, France), Nabil HATHOUT (CLLE-ERSS / Université Jean Jaurès, France), Stéphanie LIGNON (ATILF / Université de Lorraine, France), Fabio MONTERMINI (CLLE-ERSS / Université Jean Jaurès, France), Fiammetta NAMER (ATILF / Université de Lorraine, France), Delphine TRIBOUT (STL / Université de Lille, France), Florence VILLOING (MoDYCO / Université Paris Ouest Nanterre La Défense, France)

## **Scientific Committee**

Paolo ACQUAVIVA (University College, Dublin, Ireland), Anna ANASTASIADIS-SYMEONIDIS (University of Thessaloniki, Greece), Giorgio Francesco ARCODIA (Università di Milano Bicocca, Italia), Mark ARONOFF (Stony Brook University, USA), Jenny AUDRING (Universiteit Leiden, Nederlands), Matthew BAERMAN (University of Surrey, United Kingdom), Laurie BAUER (Victoria University of Wellington, New Zealand), Geert Booij (Universiteit Leiden, Nederlands), Dunstan BROWN (University of York, United Kingdom), Basilio CALDERONE (CNRS CLLE-ERSS / Université Jean Jaurès, Toulouse, France), Bruno CARTONI (Google Research, Switzerland), Séverine CASALIS (SCALab / Université de Lille, France), Berthold CRYSMANN (CNRS LLF / Université Paris-Diderot, France), Serena DAL MASO (Università di Verona, Italia), Walter DE MULDER (Universiteit Antwerpen, België), Laura DOWNING (Göteborgs Universitet Sverige), Wolfgang U. DRESSLER (Universität Wien, Austria), Louise ESHER (CNRS CLLE-ERSS / Université Jean Jaurès, France), Antonio FABREGAS (Universitetet i Tromsø, Norge), Zoe GAVRILIDOU (Democritus University of Thrace, Greece), Nicola GRANDI (Università di Bologna, Italia), Stephan HARTMAN (Universität Hamburg, Deutschland), Claudio IACOBINI (Università di Salerno, Italia), Marianne KILANI SCHOCK (University of Lausanne, Switzerland), Maria Rosa LLORET (Universitat de Barcelona, España), Francesca MASINI (Università di Bologna, Italia), Chiara MELLONI (Università di Verona, Italia), Muriel NORDE (Nordeuropa-Institut, Humboldt-Universität zu Berlin, Deutschland), Enrique PALANCAR (CNRS SEDYL-CELIA & Surrey Morphology Group, France & United Kingdom), Vito PIRRELLI (Istituto di Linguistica Computazionale «A. Zampolli», Italia), Marc PLENAT (Université Jean Jaurès, France), Franz RAINER (Institut für Romanische Sprachen, Wien, Austria), Angela RALLI (University of Patras, Greece), Davide RICCA (Università di Torino, Italia), Andrew SPENCER (University of Essex, United Kingdom), Pavel ŠTICHAUER (Charles University, Prague, Czech Republic), Dejan STOSIC (CLLE-ERSS / Université Jean Jaurès, France), Gregory STUMP (University of Kentucky, USA), Anna Maria THORNTON (Università de l'Aquila, Italia), Kristel VAN GOETHEM (Université de Louvain-La-Neuve, Belgique), Madeleine VOGA (LLACS / Université Paul Valéry, France)

## Programme

### December 13<sup>th</sup>

12h00-14h15	<b>Registration</b>	
14h15-14h30	<b>Opening</b>	
14h30-15h30	<b>Invited speaker</b> I. Plag	<i>How morphological structure affects phonetic encoding: Modeling the duration of morphemic and non-morphemic S</i>
15h30-16h00	S. Beniamine, O. Bonami, J. McDonough	<i>When segmentation helps. Implicative structure and morph boundaries in the Navajo verb.</i>
16h00-16h30	L. Esher	<i>Morphemic structure(s) and the Occitan synthetic conditional.</i>
16h30-17h00	<b>Coffee break</b>	
17h00-17h30	M. Voga, A. Anastasiadis	<i>Connecting lexica in monolingual and bilingual cross-script morphological processing: base and series effects.</i>
17h30-18h00	M. Schlechtweg	<i>The naming potential and acquisition of novel compounds and phrases.</i>
18h00-18h30	M. Shimada, A. Nagano	<i>A relational nominal structure in nominal predicates.</i>

### December 14<sup>th</sup>

9h30-10h00	<b>L. Laks</b>	<i>(Dis)favoring polycategoriality in word formation? Evidence from Palestinian Arabic and Hebrew.</i>
10h00-10h30	<b>G. Schalchli</b>	<i>Les doubles suffixes revisités: pour un changement radical de l'analyse dérivationnelle.</i>
10h30-11h00	N. Hathout, S. Lignon, F. Namer	<i>Morphophonology and paradigms in the Demonette derivational database.</i>
11h – 11h30	<b>Coffee break</b>	
11h30-12h00	E. Dugas J. Michaud	<i>Word formation in general and special languages: the case of super- and s- prefixations.</i>
12h00-12h30	<b>C. Szecel</b>	<i>Construction Morphology applied to the analysis of medical neologisms in the Middle Ages: benefits and challenges.</i>
12h30 - 14h	<b>Lunch</b>	

## ISMo 2017 – Lille

14h00-14h30	<b>E. Fornasiero</b>	<i>Morphology through the hands and face: evaluative morphology in Italian Sign Language.</i>
14h30-15h00	<b>Y. Sennikova, B. Garcia</b>	<i>Structuration du lexique de la Langue des Signes Française (LSF). Focus sur la nature et les fonctions des composants infra-unité lexicale.</i>
15h00-15h30	<b>E. Göbbel</b>	<i>English ie-hypocoristics: A DM analysis.</i>
15h30-15h45	<b>Coffee break</b>	
15h45-16h30	<b>Poster Session</b>	
16h30-17h00	<b>N. Faust</b>	<i>[lo kaze aj-aj-aj]: Haplology in Modern Hebrew plural marking.</i>
17h00-17h30	<b>G. Boyé</b>	<i>Inflectional Morphology without Inflectional Classes: a French Conjugation Reflection.</i>

## December 15<sup>th</sup>

9h30-10h30	<b>Invited speaker L. McNally</b>	<i>Roots and functional morphology: A view from combined formal and distributional semantics</i>
10h30-11h00	<b>B. Fradin</b>	<i>Deverbal nominalizations denoting places.</i>
11h00-11h30	<b>Coffee break</b>	
11h30-12h00	<b>A. Fabregas, R. Marin</b>	<i>Lexical categories and semantic primitives: the case of Spanish –ncia.</i>
12h00-12h30	<b>M.L. Knittel</b>	<i>French nominals in -ant: semantic properties.</i>
12h30-14h00	<b>Lunch</b>	
14h00-14h30	<b>F. Picard</b>	<i>Stem alternation at the morphology-phonology interface: examples from Skolt Saami.</i>
14h30-15h00	<b>B. Crysmann</b>	<i>Patterns of Allomorphy in Benabena: the case for multiple inheritance.</i>
15h00-15h30	<b>B. Herce</b>	<i>Form-form relations in the paradigm.</i>
15h30-16h00	<b>Coffee break</b>	
16h00-16h30	<b>J. Radimsky</b>	<i>Does French have verbal-nexus Noun+Noun compounds? A corpus-based study.</i>
16h30-17h00	<b>F. Villoing</b>	<i>Verbes statifs et mots composés Verbe-Nom du français : une préférence discrète.</i>
17h00-17h15	<b>Closing</b>	

## Posters

R. Blin	<i>Représentation sémantique des noms propres dérivés par suffixation, en japonais. Le cas du suffixe go « langue (linguistique) ».</i>
A. Millet, N. Auclair-Ouellet, M Fossard	<i>Compéences morphologiques compositionnelles en production chez des sujets sains francophones dans une tâche de dénomination d'images.</i>
I. Pujol	<i>From Manner to Result Verbs: the Evolution of -izar Complex Verbs in Old Spanish.</i>
Y. Schauwecker	<i>Typological flexibility: Anglo-French morphology &amp; syntax and its sociolinguistic implications.</i>
F. Nemo, F. Badin	<i>Permutations in the Lexicon: an Automated Approach.</i>
S. Moradi, M. Aronoff, L. Repetti	<i>The Continuum from Epenthesis to Linking Elements.</i>



## Invited Speakers' Abstracts

<b>McNALLY Louise</b> , <i>Roots and functional morphology: A view from combined formal and distributional semantics .....</i>	9
<b>PLAG Ingo</b> , <i>How morphological structure affects phonetic encoding: Modeling the duration of morphemic and non-morphemic S.....</i>	10

## List of Abstracts

<b>BENIAMINE Sacha, BONAMI Olivier, McDONOUGH Joyce</b> , <i>When segmentation helps. Implicative structure and morph boundaries in the Navajo verb .....</i>	11
<b>BLIN Raoul</b> , <i>Représentation sémantique des noms propres dérivés par suffixation, en japonais - le cas du suffixe go « langue (linguistique) » .....</i>	16
<b>BOYÉ Gilles</b> , <i>Inflectional Morphology without Inflectional Classes: a French Conjugation Reflection.....</i>	21
<b>CRYSMANN Berthold</b> , <i>Patterns of Allomorphy in Benabena: the case for multiple inheritance ..</i>	25
<b>DUGAS Edwige, MICHAUD Jérôme</b> , <i>Word formation in general and special languages: the case of super- and s- prefixations .....</i>	31
<b>ESHER Louise</b> , <i>Morphemic structure(s) and the Occitan synthetic conditional .....</i>	36
<b>FÁBREGAS Antonio, MARÍN Rafael</b> , <i>Lexical categories and semantic primitives: the case of Spanish -ncia .....</i>	40
<b>FAUST Noam</b> , <i>[lo kaze aj-aj-aj] : Haplology in Modern Hebrew plural marking .....</i>	45
<b>FORNASIERO Elena</b> , <i>Morphology through the hands and face : evaluative morphology in Italian Sign Language.....</i>	48
<b>FRADIN Bernard</b> , <i>Deverbal nominalizations denoting places .....</i>	53
<b>GÖBBEL EDWARD</b> , <i>English ie-hypocoristics: A DM analysis .....</i>	55
<b>HATHOUT Nabil, LIGNON Stéphanie, NAMER Fiammetta</b> , <i>Morphophonology and paradigms in the Demonette derivational database .....</i>	57
<b>HERCE Borja</b> , <i>Form-form relations in the paradigm .....</i>	62
<b>KIHM Alain</b> , <i>Predicate highlighting in Yiddish and the question of “pseudo-infinitives” .....</i>	65
<b>KNITTEL Marie Laurence</b> , <i>French nominals in -ant: Semantic properties .....</i>	70

<b>LAKS Lior, (Dis)favoring polycategoriality in word formation? Evidence from Palestinian Arabic and Hebrew .....</b>	<b>76</b>
<b>MILLET Alice, AUCLAIR-OUELLET Noémie, FOSSARD Marion, Compétences morphologiques compositionnelles en production chez des sujets sains francophones, dans une tâche de dénomination d'images.....</b>	<b>79</b>
<b>MORADI Sedigheh, ARONOFF Mark, REPETTI Lori, The Continuum from Epenthesis to Linking Elements .....</b>	<b>84</b>
<b>NEMO François, BADIN Flora, Permutations in the Lexicon: an Automated Approach .....</b>	<b>89</b>
<b>PICARD Flore, Stem alternation at the morphology-phonology interface: examples from Skolt Saami.....</b>	<b>95</b>
<b>PUJOL PAYET Isabel, From Manner to Result Verbs: the Evolution of –izar Complex Verbs in Old Spanish.....</b>	<b>99</b>
<b>RADIMSKÝ Jan, Does French have verbal-nexus Noun+Noun compounds? A corpus-based study .....</b>	<b>102</b>
<b>SENNIKOVA Yana, GARCIA Brigitte, Structuration du lexique de la Langue des Signes Française (LSF). Focus sur la nature et les fonctions des composants infra-unité lexicale .....</b>	<b>105</b>
<b>SCHALCHLI Gauvain, Les doubles suffixes revisités : pour un changement radical de l'analyse dérivationnelle .....</b>	<b>111</b>
<b>SCHAUWECKER Yela, Typological flexibility: Anglo-French morphology &amp; syntax and its sociolinguistic implications.....</b>	<b>116</b>
<b>SCHLECHTWEG Marcel, The naming potential and acquisition of novel compounds and phrases .....</b>	<b>120</b>
<b>SHIMADA Masaharu, NAGANO Akiko, A relational nominal structure in nominal predicates.....</b>	<b>123</b>
<b>SZECEL Céline, Construction Morphology applied to the analysis of medical neologisms in the Middle Ages: benefits and challenges .....</b>	<b>126</b>
<b>VILLOING Florence, Verbes statifs et mots composés Verbe-Nom du français : une préférence discrète .....</b>	<b>129</b>
<b>VOGA Madeleine, ANASTASSIADIS-SYMEONIDIS Anna, Connecting lexica in monolingual and bilingual cross-script morphological processing: base and series effects .....</b>	<b>134</b>

## **Roots and functional morphology: A view from combined formal and distributional semantics.**

**Louise McNALLY**

### **Description**

In recent work (e.g. McNally 2017, McNally & Boleda 2017) I have been exploring a specific approach to combining formal and so-called distributional semantic modeling (see e.g. Baroni, Bernardi & Zamparelli 2014, Boleda & Herbelot 2016 for discussions of the latter in relation to formal approaches to meaning). One of the interesting aspects of this mixed approach is that it leads naturally to positing a distinction between "root" and "functional" morphology in the spirit of that proposed by Borer in various works and also found in the Distributed Morphology literature. In this (very programmatic) talk, I offer what I hope will be a fresh and useful semantic modeling perspective on some of the debates and general issues that I have encountered in exploring this literature.

## How morphological structure affects phonetic encoding: Modeling the duration of morphemic and non-morphemic S

Ingo PLAG

Recent research on the acoustic realization of affixes has revealed differences between phonologically homophonous affixes, for example the different kinds of final S in English (Plag et al. 2015, Zimmermann 2016). Such results are unexpected and unaccounted for in widely-accepted models which separate lexical and post-lexical phonology, and in models which interpret phonetic effects as consequences of different prosodic structure. This paper demonstrates that the differences in duration between different kinds of final S in English (non-morphemic, plural, third person singular, genitive, genitive plural, cliticized *has*, and cliticized *is*) result from linguistic experience.

We use a Naive Discriminative Learning Model (Rescorla 1988), which is a two-layer model that computes association weights between cues and outcomes. These weights can be conceptualized as the extent to which a particular form can be expected to be associated with a particular meaning in the mind of the average speaker.

With data taken from the Buckeye Speech Corpus we show that the association weights between phonological and semantic information in the context of a given word form as cues, and the suffixal meaning as outcomes, are highly predictive of acoustic duration.

The results have important implications for theories of phonology-morphology interaction and for models of language production.

### References

- PLAG, I., HOMANN, J., and KUNTER, G. (2015). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, pages 1-36.
- RESCORLA, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43(3):151-.
- ZIMMERMANN, J. (2016). Morphological status and acoustic realization: Findings from New Zealand English. In Carignan, C. and Tyler, M. D. (eds.), *Proceedings of the 16th Australasian International Conference on Speech Science and Technology, 6-9 December 2016, Parramatta, Australia*. University of Western Sydney, Sydney.

## When segmentation helps Implicative structure and morph boundaries in the Navajo verb

**Sacha BENIAMINE,**

*Université Paris Diderot*

[sbeniamine@linguist.univ-paris-diderot.fr](mailto:sbeniamine@linguist.univ-paris-diderot.fr)

**Olivier BONAMI,**

*Université Paris Diderot*

[olivier.bonami@linguist.univ-paris-diderot.fr](mailto:olivier.bonami@linguist.univ-paris-diderot.fr)

**Joyce McDONOUGH,**

*University of Rochester*

[joyce.mcdonough@rochester.edu](mailto:joyce.mcdonough@rochester.edu)

Recent work in Word and Paradigm morphology argues that the implicative structure of paradigms is expressed in terms of relations between surface words, and that studying the structure of paradigms in terms of sub-word units is misleading if not outright impossible (Ackerman et al, 2009; Blevins, 2006, 2016; Bonami & Beniamine, 2016). The argument typically rests on the observation that a word can only be segmented in the context of its paradigmatic alternatives, and that different aspects of the paradigm lead to different segmentations for the same word.

This line of argumentation amounts to a claim about the empirical properties of some inflection systems. It is thus entirely possible that systems differ in this respect. In this presentation we show that there are systems where a uniform segmentation is possible and helpful to addressing implicative structure. Interestingly though, the segments that are identified lack the properties of classical morphemes.

### 1. The Navajo verbal system

The Athabaskan languages represent a classic example of polysynthetic morphology. In polysynthetic languages word internal structure lacks the transparency of more agglutinative types; that is, the proposed morphemes are not easily separable or identifiable. In a commonly adopted model of the highly complex and synthetic Navajo (Athabaskan) verb, the verb consists of a series of prefixes attached to a rightmost and prominent stem. An extensive ‘position class’ template provides a prosthesis for ordering among these prefixes. Lists of the morphemes for each posited position are deconstructed from fully inflected word forms, and are understood to be sound-meaning pairs. Extensive rewrite rules are needed to recompose forms from the elements of this model.

There are many problems with this position class template, apart from its dependence on reverse engineering to derive the morphemes. Importantly, the template focuses exclusively on the morphemes and their relative order. In focusing thus, it misses many important generalizations about the structure of the verbal complex and, particularly, the relationship of words to each other, including the existence of conjugation patterns, inflectional paradigms, and the internal inflections of the stem itself. One alternate approach is

to determine the actual forms in the complex that speakers may identify and use in word formation and in the organization of their lexicons. Based on the work of Young and Morgan (1980, 1987, 1992) (principle reference grammars of Navajo), McDonough and colleagues (1990, 1999, 2000, 2003, 2012, 2015; McDonough and Wood, 2008) have conducted a study of phonetic, phonotactic and phonological patterns in the verb across the Athabaskan languages and provided a consilience of arguments for the existence of two separable and independent but interdependent elements in the verb itself, identifying a minimal or ‘core verb’ obligatorily comprised of these elements, which carry the minimal morphosyntactic specification of a well formed verb. These are a *Mode* (*M*) element inflected for person and number, expressing the principle conjugational patterns of the verb, in penult position, and a monosyllabic *Stem* element (*S*), consisting of the ‘classifier’ (valence) plus ‘stem shape’ as the final and most prominent form in the verbal complex.

The final syllable in the verbal complex, the *Stem*, is phonetically prominent. The inflected *Mode* element on the other hand represents what Young and Morgan refer to as the *Base Paradigms*, a set of 4+ basic conjugations that all verbs are inflected in. This inflected *Mode* element may take a set of prefixal morphemes that serve to build up a very rich set of morpho- syntactic and -semantic meanings. These two elements, M(ode) and S(tem), represent independent yet inter-dependent dimensions of paradigmatic variation in the word.



## Prefix Mode Stem

In Young and Morgan's dictionary, each fully inflected verb form is given in 5 principal parts carrying distinct TMA values, and exemplified below for the verb BIDISHNE'. Somewhat confusingly, these five TMA values are also called 'modes'. Each verb involves a characteristic paradigm of mode elements and a characteristic paradigm of stem shapes. the Imperfective 1SG serves as the citation form.

TMA value	Mode	Stem	Surface form
Imperfective	bidish	(ł)ne'	bidishne'
Repetitive	bińdışh	(ł)niih	bińdışniih
Perfective	bidíí	łne'	bidííłne'
Future	bidideesh	(ł)niił	bidideeshniił
Optative	bidósh	(ł)ne'	bidóshne'

**Table 1** — Principal parts for the verb BIDISHNE'

The forms select each other to produce a rich set of aspectual meaning; these forms do not exhaust the possible combinations but act as principle components. Thus, although demonstrably independent, the inter-dependence of the M and S elements create word meaning; meaning is not compositional, but resides in the patterns of combinations of the elements in the verb. In this paper in particular we focus on the interdependence of the shapes of Stem and the Mode forms.

## 2. Data

Young & Morgan (1987) document the paradigms of Navaho verb in remarkable detail. Their dictionary presents information on inflectional paradigms in two guises. First, verb entries

provide five principal parts for each lexeme, corresponding to five main TMA combinations. Table 2 exemplifies some entries. These are in the 1SG except for impersonal verbs which are given in the 3SG. The dictionary contains 5073 such entries.

Imperfective	Repetitive	Perfective	Future	Optative	Translation
niishkaał	nániishkaał	niishkaal	dínéeshkaał	nooshkaał	support by pushing on
'áltániishkaał	'áltánániishkaał	'áltániilkaal	'áltádínéeshkaał	'áltánooshkaał	chop or split lengthwise
niish'eeł	nániish'oł	nii'éél	dínéesh'oł	noosh'eeł	dissolve
niishdóoh	nániishdoh	niildoi	dínéeshdoh	nooshdóoh	heat
yıldóoh	náldoh	yıldoi	doołdoh	wóldóoh	dry up during summer
yílhéésh	nálhésh	yílhézh	doołhésh	wólhéésh	move through the air (mushy matter)

**Table 2** — Sample principal part series from Young & Morgan

Second, the dictionary contains person-number-TMA paradigms for all prefixes+mode combinations. For instance, paradigms for *niish-*, *'altániish-*, and *yıl-*, the three combinations exemplified above, are provided. A few hundred such paradigms are provided. Full paradigms for all lexemes can easily be deduced from verb entries and prefix+mode paradigms, since no sandhi phenomena or other unpredictable alternations occur at the mode-stem boundary (unlike what happens at the prefix-mode boundary). Unfortunately though, the prefix+mode paradigms are not available in digital form yet. Hence we leave the examination of full paradigms for a future study.

For the present study we constructed a dataset on the basis of a digital version of Young & Morgan, which allowed easy semi-automatic tabulation of the five principal parts from each verb entry. After normalisation, error corrections, and exclusion of defective lexemes, we were left with a set of 1418 five cell (sub)paradigms. IPA transcriptions were deduced automatically for the orthographic forms, and mode-stem boundaries were introduced semi-automatically, by identifying and segmenting all consonant clusters occurring between the penultimate and final vowels. Table 3 exemplifies the results of these processes on the sample in Table 2.

Lexeme	Imperfective	Repetitive	Perfective	Future	Optative
NIISHKAAL	ni:j+kxa:l	náni:j+kxał	ni:j+kxa:l	tínéeʃ+kxał	no:j+kxa:l
'ALTANIISHKAAL	?áltxáni:j+kxa:l	?áltxánáni:j+kxał	?áltxáni:+łkxa:l	?áltxátinééʃ+kxał	?áltxáno:j+kxa:l
NIISH'EEL	ni:j+?e:l	náni:j+?oł	ni:+?é:l	tínéeʃ+?oł	no:j+?e:l
NIISHDOOH	ni:j+tó:h	náni:j+toh	ni:+łtoi:	tínéeʃ+toh	no:j+tó:h
YILDOOH	ji+łtó:h	ná+łtoh	ji+łtoi:	to:+łtoh	wó+łtó:h
YILHÉÉSH	ji+lhé:j	ná+lhęʃ	ni+lhę:ʒ	to:+lhęʃ	wó+lhę:j

**Table 3** — IPA version of the sample from Table 1, with mode-stem boundaries

### 3. Analysis

Building on previous work by Ackerman et al. (2009) and Ackerman and Malouf (2013), Bonami and Beniamine (2016) define implicative entropy as a way of assessing the

predictability of one paradigm cell from any other collection of paradigm cells. Unlike previous attempts at using entropy to address predictability, Bonami and Beniamine's algorithm does not presuppose a preexisting inflectional classification; hence it is readily applicable to new languages, as long as a large number of raw paradigms of surface forms is available.

We hence set out to apply Bonami and Beniamine's algorithm to the dataset derived from Young and Morgan (1987). When doing so, however, an immediate concern arose. At the heart of the algorithm is a generic method for inferring patterns of alternation from raw data. For the entropy calculations to be meaningful, it is essential that the patterns be as accurate as possible. Following Beniamine (2017), we assess the accuracy of a set of patterns by examining how well patterns extracted from a training set containing a random 90% of the data are used to predict the inflectional behavior of a test set consisting of the remaining 10% (using 10-fold cross-validation). Perfect accuracy is not expected (since the test set may contain inflectional behaviors not found in the rest of the system), but Beniamine's test cases lead us to expect accuracies in the 0.6-0.95 range. As it happens, the results were terribly bad: the accuracy of the patterns inferred from the raw dataset is only 0.28.

There is a rather direct explanation for this situation. As discussed earlier, the Mode and the Stem constitute independent dimensions of variation in the Navajo verb. This results in a combinatory explosion, where patterns of alternations between full forms involve a combination of a pattern of alternation for the mode and a pattern of alternation for the stem. Hence the test set is bound to contain many patterns not exemplified in the train set.

This combinatory explosion is a consequence of not taking advantage of the fact that every surface form of a verb can readily be segmented at the Mode-Stem boundary, as the stem coincides with the word's last syllable, a fact that speakers of the language are certainly attuned to. To assess the usefulness of this segmentation, we evaluated the very same algorithm on two datasets consisting of only the pre-stem material and only the stem. As indicated in Table 4, the results are then much more satisfactory, and stand in the range observed for other languages.

Dataset	Average accuracy	Average # of patterns
full words	0.33	537
pre-stem material	0.79	87
stems	0.75	102

**Table 4** — Accuracy of pattern inference (10-fold cross-validation)

This result indicates that segmentability is a useful feature of the Navajo conjugation system. The problem faced by the pattern inference algorithm discussed above certainly is a problem also faced by speakers learning the language: if they did not rely on segmentation, they would be bound to make a spectacular number of errors.

Once patterns of alternation for stem and pre-stem material have been computed, we are now in a position to assess the implicative entropy of the system. We computed independently the implicative entropy of the system of stems, the system of pre-stem sequences, and the combination of both (taking the pattern relating two full words as the combination of a pre-stem pattern and a stem pattern). The results confirm that, despite its strong intricacies, the level of predictability of the Navajo conjugation system is manageable. Moreover, full words are more predictive than their parts: despite being partly independent, pre-stem sequences are partly predictive of stems, and vice-versa. This indicates that, while the identification of inflectional patterns in the Navajo verb require a segmentation, full words still are the best unit of prediction in such a system.

Dataset	Entropy
full words	0.33
pre-stem material	0.53
stems	0.62

**Table 5** — Average unary implicative entropy**References**

- ACKERMAN, F., BLEVINS, J. P. & MALOUF, R. (2009). Parts and wholes: implicative patterns in inflectional paradigms. In James P. Blevins and Juliette Blevins (eds.), *Analogy in Grammar*, Oxford: Oxford University Press. 54–82.
- ACKERMAN, F. & R.MALOUF, (2013). Morphological organization: the low conditional entropy conjecture. *Language* 89: 429-464.
- BENIAMINE, S. (2017). Une approche universelle pour l'abstraction automatique d'alternances morphophonologiques. *Proceedings of TALN 2017*.
- BLEVINS, J. P. (2006). Word-based morphology. *Journal of Linguistics* 42: 531–573.
- BLEVINS, J. P. (2016). *Word and Paradigm Morphology*. Oxford: Oxford University Press.
- BONAMI, O. & Beniamine, S. 2016. Joint Predictiveness in inflectional paradigms. *Word Structure* 9.2: 156-182.
- ISKAROUS, K., McDONOUGH J. M. & D. H. WHALEN. (2012). A gestural account of velar contrast: the back fricatives in Navajo. *Laboratory Phonology* 3.1: 195-210.
- McDONOUGH, J. M. (2003). *The Navajo Sound System*. Kluwer Academic Press.
- McDONOUGH, J. M. & V. WOOD. (2008). The stop contrasts of the Athabaskan languages. *Journal of Phonetics*. 36.3: 427-449.
- YOUNG, R. W. & W.MORGAN. (1987). *The Navajo Language: A Grammar and Colloquial Dictionary*

## Représentation sémantique des noms propres dérivés par suffixation, en japonais - le cas du suffixe *go* « langue (linguistique) »

**Raoul BLIN,**  
*CNRS-CRLAO*  
 blin@ehess.fr

La dérivation par suffixation des noms japonais est très largement documentée pour ce qui touche la morphologie (voir par exemple récemment (Sugioka et Ito 2016)) ou la phonologie. Mais les analyses sémantiques restent intuitives et font l'impasse sur la distinction entre sens encodé et connaissances générales. Nous proposons une représentation sémantique formalisée pour un cas productif de dérivation par suffixation : nom propre + suffixe sino-japonais. Nous illustrerons notre propos en décrivant le fonctionnement du suffixe *go* lorsque celui-ci désigne la « langue (linguistique) » : *nihon-go* « Japon-langue ; langue du Japon ; japonais ». Nous confrontons plusieurs analyses et montrons que l'analyse compositionnelle est la plus élégante, à condition de ne pas adopter l'approche du lexique génératif. Nous implémentons notre analyse de la dérivation (morphologie et sémantique incluses) à l'aide d'une grammaire de contraintes intégrées avec base de connaissances (Renaud 2005). Dans les lignes qui suivent, par défaut, les « connaissances générales » sont celles d'un adulte standard. Les règles de contraintes et représentations sémantiques sont exprimées en théorie des types, dans une variante syntaxique du système *Q<sub>o</sub>* (Andrews 1986) augmentée d'un langage de contraintes sur les termes de traits.

### 1. Description

Conformément à la tradition, nous décomposons les noms dérivés comme *nihon-go* (« Japon-langue ; japonais ») en un nom propre (np) *nihon* et un suffixe *go*. Par prudence, nous distinguons ce suffixe d'au moins trois homophones et homographes qui font référence soit au « mot », soit au « parler » (ex : *wakamono-go* « jeune-parler ; le parler jeune »), soit à la « version » d'une langue à un moment donné (ex : *gendai-go* « présent-langue ; langue contemporaine »). *go* suffit trois types sémantiques de noms propres : les toponymes, les noms d'ethnies (hyponymes de *minzoku*) et les noms de langues. Le suffixe étudié ici est un morphe dépendant. Voici une description succincte des trois cas. Pour des questions de place, nous passons sur de nombreux détails, dont l'analyse ne remet pas en cause notre propos. De même, nous n'évoquons pas les quelques exceptions existantes (voir néanmoins discussion plus loin).

*go* suffit les np d'ethnies (hyponymes de *minzoku* « ethnie »). Le dérivé se glose alors par « langue associée à une ethnie » : *ainu-go* « langue(s) associée(s) à [l'ethnie] Aïnu ; aïnu ». Lorsque *go* suffit un toponyme correspondant à une entité territoriale (pays, région, continent...), le dérivé se glose « langue(s) parlée(s) associée à un toponyme » : *ahurika-go* « Afrique-langue ; langue(s) associée(s) au territoire Africain ; africain ». La nature de l'association relève des connaissances de chacun : langue(s) parlée(s) officiellement ou majoritairement par les habitants ou les natifs du lieu désignés etc. On voit même que le

nombre n'est pas déterminé et dépendra des connaissances de chacun. Le sens exact, nombre compris, ne sont donc pas « lexicalisés ». Nous pouvons seulement dire a minima que dans tous les cas, le dérivé désigne « une langue », ce qui figurera dans la représentation lexicale. Par contre, la définition exacte de l'association sera donnée dans la base de connaissances. Notons simplement enfin que, comme le montrent les gloses, l'entité à laquelle réfère le dérivé est calculée par rapport au radical. Nous parlerons d'interprétation fonctionnelle.

Le radical peut être aussi un np de langue. Le dérivé (*esperanto-go* « espéranto-langue ») est alors synonyme du np radical (*esperanto-go* = *esperanto*) et ne s'en distingue éventuellement que par le registre de langue. Nous ne sommes plus ici dans une interprétation fonctionnelle.

## 2. Règle de dérivation

Nous nous concentrerons sur l'interprétation fonctionnelle. La relation de synonymie pose moins de problème et sera exprimée sans difficulté dans la représentation sémantique (section 3). Trois implémentations de la relation fonctionnelle sont envisageables. Elles incarnent chacune une vision différente de la compositionnalité des dérivés en *go*.

1) La première implémentation correspond à une analyse compositionnelle. La règle de suffixation est unique. Les différents sens du mot dérivé sont « prévus » dans la représentation sémantique de l'entrée lexicale du suffixe. Informellement, la représentation sémantique de l'entrée lexicale du suffixe est la suivante :

- Le dérivé désigne une langue (c'est à dire est hyponyme de *gengo* « langue »)
- Si le radical est un toponyme : cette langue est associée au lieu désigné,
- Si c'est un nom de communauté : cette langue est associée à la communauté.
- Sinon : il n'y a pas de langue possible  
(rejet de la dérivation).

*go* étant un suffixe qui fondamentalement dépend d'un autre morphème, il paraît naturel d'intégrer dans sa représentation sémantique les sens qu'il prendra en fonction de son argument. La règle de suffixation est alors unique et très simple : elle se contente de passer la représentation sémantique du radical en argument de celle du suffixe. C'est la stratégie que nous avons adoptée. Nous l'implémentons comme ci-dessous. Pour des raisons de place, la contrainte de traits est ici très simplifiée. Nous nous contentons de citer les éléments qui permettent de gérer les exceptions. Le principe est de s'assurer que l'identifiant du radical (LEM1) n'appartient pas à la liste RI2 des identifiants de radicaux pour lesquels la règle ne peut être appliquée. La liste est fournie par le suffixe. Les exclusions (exceptions) sont donc déclenchées sur la base de critères lexicaux (identifiant du lemme) et non sémantiques.

```
rn_golang
nc_morph <- np suffixDeNom
U0 :: (...) ∧ U1 :: lemme : LEM1 ∧ U2 :: radinterdits : RI2 ∧ ¬(LEM1 :: RI2)
RSEM2.RSEM1
```

Pour rappel, un nom propre comme *nihon* « Japon » en contexte se présente comme ci-dessous. *japon<sub>i</sub>* est la constante propre à l'entité Japon (les traits sont simplifiés) :

*nihon* , np , lemme : npjapon ∩ registre : commun... ,  $\lambda P_{o(ou)} \lambda Q_{ou} P.\{japon\}$

2) La seconde approche est non compositionnelle. *go* a un sens unique. Il revient à la règle de suffixation de préciser le sens que prendra le dérivé en fonction du suffixe et du radical. Elle doit contenir l'interprétation de toutes les combinaisons (nom-suffixes) possibles. La règle vaut :

- si le suffixe est *go*
  - si le nom est un toponyme alors...
  - si c'est un nom de communauté, alors ...
  - sinon la règle ne s'applique pas
- si le suffixe est ...
  - ...
  - si ...

Nous ne sommes pas hostile par principe à l'idée de non compositionalité, mais pas appliquée de manière systématique. Dans le cas de la suffixation en *go*, ce principe pose plusieurs problèmes. De forme tout d'abord. La règle sémantique est monstrueuse par sa taille puisque dans la représentation elle doit décrire toutes les relations pour tous les suffixes et tous les radicaux. Sur le fond, la règle n'est pas productive puisqu'elle se limite strictement aux suffixes qui sont explicitement mentionnés. A chaque apparition d'un nouveau suffixe dans le vocabulaire de la langue, il faudra modifier la règle. La complexité et la non productivité vont à l'encontre de l'intuition que la dérivation par suffixation telles que mise en oeuvre pour *go* est un processus simple et très productif.

3) La troisième implémentation se situerait peut-être entre les deux précédentes. Le suffixe et la règle renvoient chacun une valeur unique. On considère que le suffixe n'accepte qu'une unique catégorie sémantique de radical. Par exemple, *go* attend un nom de communauté seulement, pas de lieu. En fait, la polysémie réside alors au niveau du radical. Conformément au principe du lexique génératif (Pustejovsky 1991), plusieurs sens sont dérivés du sens du np. L'un d'entre eux correspond au sens attendu par le suffixe. Voici une suggestion d'analyse, qui servira de support à la discussion :

*go* : langue parlée par la communauté désignée par le radical.

*nihon*<sup>1</sup> : toponyme

règle de dérivation sémantique 'rd' : tout toponyme peut désigner la communauté humaine associée au lieu qu'il désigne

*nihon*<sup>2</sup> (dérivé de *nihon*<sup>1</sup> par 'rd') : communauté humaine associée au Japon

L'attente du suffixe est aussi satisfaite par *nihon*<sup>2</sup>, un des sens dérivés de *nihon*<sup>1</sup>. *nihon-go* désigne alors « la langue parlée par la < communauté qui est liée au Japon> ». En apparence, l'analyse semble compositionnelle puisque la polysémie est inscrite dans le composant (le nom) et non dans la règle de dérivation elle-même. Dans la réalité, c'est à dire lorsqu'on tente de l'implémenter, elle ne l'est pas. En effet, aucune autre distribution que celle-ci n'exploite le sens « communauté» des toponymes. Si la dérivation sémantique 'rd' était appliquée librement, n'importe quel toponyme dans n'importe quelle distribution pourrait être interprété «communauté». Ce qui ne convient pas bien sûr. Pour éviter une telle surgénération, il faut n'autoriser cette la dérivation que lorsque le nom est suffixé de *go*. Or seule la règle de dérivation peut « héberger » l'autorisation d'appliquer 'rd' au toponyme. Plus généralement, la règle hébergera toutes les autorisations (ou interdictions) nécessaires au traitement de tous les suffixes qu'elle est supposée gérer. De sorte que, comme le montre l'illustration ci-dessous, cette règle sera semblable à la règle de l'implémentation 2, avec les mêmes défauts (lourdeur et non productivité).

si le suffixe est *go*  
 si le nom est un toponyme,  
     alors appliquer 'rd' au nom radical  
     puis construire la représentation du dérivé  
 si le nom est ...  
 ...  
 si le suffixe est ...  
 si le nom est ...  
 ...  
 ...

### 3. Représentation sémantique de *go*

Pour fonctionner avec la règle de dérivation énoncée précédemment (stratégie 1), après différentes ébauches que nous ne présentons pas ici faute de place, nous aboutissons à la représentation ci-dessous. Le principe est d'énoncer dans la représentation du suffixe les propriétés sémantiques des np qu'il peut suffixer, et les traitements à appliquer :

- Si le radical est un toponyme : le tout désigne la langue parlée dans le lieu désigné,
- Si c'est un nom de communauté : il désigne la langue parlée par cette communauté.
- Si c'est une langue : il désigne cette langue
- Sinon : la règle ne s'applique pas

La représentation complète de l'entrée est la suivante. Le trait *radicauxexclus* énumère les identifiants des lemmes que *go* ne peut suffixer par application de la règle *rn\_golangue*. Comme évoqué en section 1, la relation entre le radical et le dérivé varie d'un individu à l'autre. Nous proposons donc de désigner cette relation par un prédictat *ad hoc*. Sa valeur exacte sera définie dans la base de connaissances.

*go, SUFFIXEDENOM , lemme : golangue ∩ registre : commun ∩ ... ,*  
 $\lambda RSN \lambda P \lambda Q$   
*RSN . (λGN P.(λL ∃E lmlang.E.L*  
 $\wedge GN.X$   
 $\wedge$  si  $GN \subset entiteAdministrativeTerritoriale_{ou}$   
     alors  $\exists E lgterritoire.E.X.L \wedge Q.E$   
     sinon  
         si  $GN \subseteq lmminzoku_{ou}$   
             alors  $\exists E lgethnie.E.X.L \wedge Q.E$   
             sinon  
                 si  $GN \subset lmlangue_{ou}$   
                     alors  $GN.L$   
                     sinon faux ) )

 $.(\lambda E vrai)$ 

Après analyse, la représentation de *nihon-go* par exemple se présente :  $\lambda P \lambda Q P.(\lambda L \exists E lgterritoire.E.japon.L)$ . La représentation de *esperanto-go* vaut

$\lambda P \lambda Q \ P.\{esperanto\}$ , cette dernière est égale à la représentation du np *esperanto* seul.

Un travail exploratoire nous laisse penser que cette analyse est valide au moins pour les suffixes suivants :

*go<sup>parler</sup>* (« parler ») : *gakusei-go* (« parler Etudiant ; jargon Etudiant »)

*jin* (« personne ») : *nihon-jin* (« personne du japon ; Japonais »)

*gaku* (« étude ») : *nihon-gaku* (« étude du Japon ; japonologie »)

*sen* (« élection ») : *koronbi'a-sen* (« élection en Colombie »)

## Références

- ANDREWS, P. B. (1986). *An introduction to mathematical logic and type theory: to truth through proof*. Computer science and applied mathematics. Orlando: Academic Press.
- BLIN, R. (2017). « Désambiguisation des noms propres à l'aide des connaissances sur autrui ». *Linguisticae Investigationes* 42 (2).
- PUSTEJOVSKY, J. (1991). « The Generative Lexicon ». *Computational Linguistics* 17 (4).
- RENAUD, F. (2005). *Temps, durativité, télicité*. Peeters. Bibliothèque de l'Information Grammaticale 60. Louvain, Paris.
- SUGIOKA, Y. & T. ITO. (2016). « Derivational affixation in the lexicon and syntax ». In *Handbook of Japanese Lexicon and Word Formation*, Kageyama Taro, Kishimoto Hideki, 347-86. Handbooks of Japanese Language and Linguistics. Mouton de Gruyter.

## Inflectional Morphology without Inflectional Classes: a French Conjugation Reflection

**Gilles BOYE,**

*Université Bordeaux-Montaigne & CLLE-ERSSàB*

gilles.boye@u-bordeaux3.fr

Inflectional morphology descriptions usually adopt a top-down perspective using, for example, a partition of the lexicon into more or less fine-grained inflectional classes and describing the different classes (e.g. Network Morphology: Corbett and Fraser (1993), Brown and Hippisley (2012) or Natural Morphology: Kilani-Schoch and Dressler (2005)), or a set of stems for lexemes and rules of realizations for feature bundles (e.g. A-Morphous Morphology: Anderson (1992) or Paradigm Function Morphology: Stump (2001)). With Blevins (2006) and Ackerman *et al.* (2009), a different type of description with an *abstractive* approach has appeared built around Information Theory (Shannon, 1948), word-based and revolving around the Paradigm Cell Filling Problem (PCFP) in (1).

- (1) Given exposure to an inflected wordform of a novel lexeme, what licenses reliable inferences about the other wordforms in its inflectional family?

This has lead to a new line of word-based descriptions hingeing on implicative relations between forms, for example the *dynamic principal parts* of Stump and Finkel (2013), or the *joint predictiveness* of Bonami and Beniamine (2015)). But a new constructive paradox appears with these predictions. Finkel and Stump (2007), for example, describe a system of dynamic principal parts and an abstractive process to create an optimal system of principal parts. The analysis is word-based and the approach is abstractive in the sense of Blevins (2006) but as the authors comment themselves (p.65):

- (2) “unencountered members of a lexeme’s paradigm can only be inferred with full reliability if one assumes that one has already encountered all of a language’s contrasting inflection classes”.

In other words, (2) says the PCFP solution proposed relies on knowing exemplary paradigms for the contrasting inflection classes. In this talk, we argue that (2) is not plausible for French conjugation in view of the sparse data available to speakers and propose an alternative solution to the PCFP that can work in this context.

### 1. French Conjugation and Data Sparsity

To satisfy (2) for French conjugation and its more than 20 inflectional classes by all accounts,<sup>1</sup> speakers would have to know at least 20 exemplary paradigms, one belonging each class of the most parsimonious description.

---

<sup>1</sup> See (Sagot and Walther, 2011, pp.38–41) or an overview of inflectional class systems for French.

To estimate the exemplary paradigms a French speaker might have encountered, we used two different datasets:

- FrWaC, a corpus of web pages containing 1,328,628,428 annotated words, 12.6% of verb forms (Baroni *et al.*, 2009)
- BDLexique, a lexicon assembled from the phonological descriptions of BDLex (de Calmès and Pérennou, 1998) and the frequencies of Lexique3 (New *et al.*, 2001) with 328103 inflected verb forms

In FrWaC, to avoid the noise generated by tagging errors we followed the lead of Bonami and Beniamine (2015) and focused on the 6847 verbs documented in the *Lefff* lexicon (Sagot, 2010).<sup>2</sup> Taking into account the conservative estimate proposed by (Ramscar *et al.*, 2010, p.945) of child speakers being exposed to around 11 million spoken words a year, we counted the lexemes with complete paradigms at different samples size. At 11,000,000 verb forms, a sample size corresponding to age 8, only two verbs max out their paradigm, ÊTRE ('be') and AVOIR ('have'), and even with the full corpus (worth several lifetimes of tokens) only 9 verbs<sup>3</sup> appeared in all possible inflectional forms.

To further study the appearance of exemplary paradigms, we sampled BDLexique by repeatedly selecting forms at random following their frequency distribution and obtained similar results. With 11,000,000 tokens, only about 22% of the inflected forms had been sampled and two verbs had reached their full paradigm.

## 2. Data Sparsity and Morphological Analysis

In a framework based on Information Theory, the available data plays a crucial role in the analysis. We have to separate what is known from what we want to predict. In their analysis of French, both Stump and Finkel (2013) and Bonami and Beniamine (2015) use an oversized dataset containing exemplary paradigms for every contrasting inflectional classes to answer the PCFP. Seen from this perspective, it is not clear what the Low Entropy Conjecture (Malouf and Ackerman, 2013) says about morphology:

- (3) If one knows everything there is to know about an inflectional system, the remaining entropy should be low?

The solution we propose is based on entropy calculations similar to those used by Bonami and Boyé (2014) and Bonami and Beniamine (2015) with two differences.

- we consider only data from a sample not the complete inflectional lexicon
- we rely on **all** the abstracted knowledge for the predictions

The first step is the same. For every pair of cells in the sample, we calculate the analogies between the two cells with their phonological context minimally generalized, we constitute classes of analogies and compute the distribution of known forms in the classes.

The second step differs. As Bonami and Beniamine have shown *joint predictiveness* works better than *single predictiveness* and is definitely possible for many lexemes. It uses generalisations from several cell pairs to predict a single output. But if the lexeme under scrutiny, say DÉCOUPER ('cut out'), has four inflected co-forms in the sample, they will use

---

<sup>2</sup> This excluded about 2% of the 167,358,968 originally tagged as verbs.

<sup>3</sup> ACHETER ('buy'), AVOIR ('have'), CHOISIR ('choose'), ÊTRE ('be'), FINIR ('finish'), METTRE ('put'), PRENDRE ('take'), TROUVER ('find'), VENDRE ('sell')

the knowledge about the analogies between the four cells containing the co-forms and the PCFP output cell to predict the outcome. In this case, the prediction benefits only from the abstracted relations between the five cells, leaving aside the other abstracted relations between all the other pairs.

We propose to exploit all the abstracted relations by using, for example, the content of the four known cells to populate not only the PCFP output cell but the entire paradigm using just single predictiveness.

This produces a massively overabundant paradigm that can then be filtered by looking for analogically related co-forms. In other words, the target forms will all be related by analogies abstracted from the sample. We represent relations between form with a graph and extract maximal cliques. For all purposes, a clique of analogical relations is whatmorphologists traditionally call an inflectional class.

It is not necessary to have exemplary paradigms to obtain this result. To abstract enough analogies, we only need to have enough pairs of forms for every pair of cells to be able to connect the clique.

### 3. Conclusion

In our view, French verb inflectional classes emerge from the analysis of the sample. As such they are not part of the analytic tools but rather a result of the abstraction. At any rate, if we want to follow an abstractive perspective to its logical end, the status of inflectional classes and even paradigm shapes should be reevaluated. They belong to the linguist toolbox and are instrumental in explaining morphology in textbooks but they do not seem to be available to speakers in general.

Instead we proposed to use the available data from the sample (inflections) to abstract predictive analogies (reflections) and use both to predict missing forms in the paradigms of the lexemes in the sample and also paradigms of novel lexemes to come. The reliability of the predicted forms varies with the size of the sample as expected. The details of the solution outlined here are not as important as the fact that inflectional morphology should not be limited to textbook descriptions of inflectional systems.

Frequency distribution seems to play an important role that is generally not taken into account in morphology outside of psycholinguistic studies. Our argument about the unavailability of exemplary paradigm rests on the zipfian distribution of verbal inflected forms in French (see Blevins *et al.*, 2016, for a similar observation about German). Of course, the distribution bias might be neutralised by other factors. In extreme cases, such as the verbal inflection of Mauritian with only two cells (Bonami *et al.*, 2011), exemplary paradigms will, of course, be available to speakers early on. One might even wonder, what could happen with languages with large inflectional paradigms but very few candidate lexemes.

### References

- ACKERMAN, F., BLEVINS, J. P. & MALOUF, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins and J. Blevins, editors, *Analogy in grammar: Form and acquisition*, Oxford Scholarship Online. pages 54–82.
- ANDERSON, S. R. (1992). *A-Morphous Morphology*. Number 62 in Cambridge Studies in Linguistics. Cambridge University Press.
- BARONI, M., BERNARDINI, S., FERRARESI, A. & ZANCHETTA, E. (2009). The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3), 209–226.
- BLEVINS, J. P. (2006). Word-based morphology. *Journal of Linguistics*, 42, 531–573.

- BLEVINS, J. P., MILIN, P. & RAMSCAR, M. (2016). The zipfian paradigm cell filling problem. *Morphological Paradigms and Functions*. Brill.
- BONAMI, O. & BENIAMINE, S. (2015). Implicative structure and joint predictiveness. In V. Pirrelli, C. Marzi, and M. Ferro, editors, *Word Structure and Word Usage: Proceedings of the NetWordS Final Conference*.
- BONAMI, O. & BOYE, G. (2014). De formes en thèmes. *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux*, pages 17–45.
- BONAMI, O., BOYE, G. & HENRI, F. (2011). Measuring inflectional complexity: French and Mauritian. Paper presented at the Quantitative Measures in Morphology and Morphological Development Workshop.
- BROWN, D. & HIPPISLEY, A. (2012). *Network morphology: A defaults-based theory of word structure*, volume 133 of Cambridge Studies in Linguistics. Cambridge University Press.
- CORBETT, G. G. & FRASER, N. M. (1993). Network morphology: a datr account of Russian nominal inflection. *Journal of Linguistics*, 29(01), 113–142.
- DE CALMES, M. & PERENNOM, G. (1998). Bdlex : a lexicon for spoken and written french. In *1<sup>st</sup> International Conference on Language Resources and Evaluation*, pages 1129–1136. ELRA, Grenade.
- FINKEL, R. & STUMP, G. T. (2007). Principal parts and morphological typology. *Morphology*, 17(1), 39–75.
- KILANI-SCHOCH, M. & DRESSLER, W. U. (2005). *Morphologie naturelle et flexion du verbe français*. Gunter Narr Verlag, Tübingen.
- MALOUF, R. & ACKERMAN, F. (2013). The low entropy conjecture: The challenges of modern irish nominal declension. *Language*, 89(3).
- NEW, B., PALLIER, C., FERRAND, L. & MATOS, R. (2001). Une base de données lexicales du français contemporain sur internet: Lexique. *L'Année Psychologique*, 101, 447–462.
- RAMSCAR, M., YARLETT, D., DYE, M., DENNY, K. & THORPE, K. (2010). The effects of featurelabel-order and their implications for symbolic learning. *Cognitive Science*, 34(6), 909–957.
- SAGOT, B. (2010). The *Lefff*, a freely available and large-coverage morphological and syntactic lexicon for French. In *Proceedings of LREC 2010*.
- SAGOT, B. & WALTHER, G. (2011). Non-Canonical Inflection: Data, Formalisation and Complexity Measures. In C. Mahlow and M. Piotrowski, editors, *SFCM 2011 - The Second Workshop on Systems and Frameworks for Computational Morphology*, volume 100, pages 23–45, Zürich, Switzerland. Springer.
- SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423 & 623–656.
- STUMP, G. & FINKEL, R. A. (2013). *Morphological typology: From word to paradigm*, volume 138. Cambridge University Press.
- STUMP, G. T. (2001). *Inflectional Morphology*. Cambridge University Press.

# Patterns of Allomorphy in Benabena: the case for multiple inheritance

Berthold CRYSMANN,

CNRS, Laboratoire de linguistique formelle

In this talk, I shall investigate patterns of allomorphic variation in Benabena and conclude, following Young (1964), that the language's verbal morphology is heavily characterised by three systematic allomorphic processes, which contrast with the otherwise agglutinative nature of the system. I shall argue in particular that most current realisational models of morphology fail to capture the systematicity of the alternations and show how a novel approach based on underspecified descriptions of rules in multiple inheritance hierarchies (Crysman and Bonami, 2016) provides the necessary means to formally capture the relevant generalisations.

## 1. Data

### 1.1. Basic facts about Benabena verb inflection

Primary verbs in Benabena (Young, 1964) take both prefixal and suffixal markers. Any primary verb minimally inflects for subject agreement (person and number), tense (present/past/future) and mood (indicative/volitional). A casual glance at Table 1 will reveal that it is quite straightforward in the general case to arrive at a consensual segmentation of exponents and that it is equally straightforward to determine their primary function. E.g. it is quite easy to determine a constant future (*l*) or past marker (*?Vh*), or else the exponent for dual (*?i*), and even subject agreement marking, at least when focusing on non-present tense. However, despite this neat segmentability, it becomes equally apparent that we are confronted with massive allomorphy that essentially permeates the system.

Before we actually zoom in to the system of allomorphy, a few remarks on morphotactics are in order: according to Young (1964), the minimal primary verb observes the following template for suffixes: STEM-TENSE-SUBJECT-(EMPHATIC)-(QUESTION/REASON)-MOOD.

### 1.2. Patterns of allomorphy

According to Young (1964), Benabena verb morphology witnesses three systematic allomorphic patterns: an opposition between first and non-first person, an opposition between first person or singular (monofocal) and non-first/non-singular (polyfocal), and finally an opposition that singles out 2SG/1PL in contrast to a default form. All of these patterns affect more than one class of exponents, i.e. they show a certain degree of systematicity.

#### 1.2.1. Monofocal (1/SG) vs. polyfocal

The first split we are going to consider singles out non-first non-singular as the marked case. The pattern can most easily be observed with past stem alternation in Class A and B: we find, in Class B, an alternation between marked *bi* and unmarked *bu*, or in Class A, the equivalent alternation between marked *he* vs. unmarked *ho*.

This split is by no means restricted to stem allomorphy: it is equally attested for prefixal operative marker *no/ne* (roughly a continuative/habitual) and for stem selection in the imperative where singular and first person choose the past stem (e.g. *ho* for Class A), whereas non-singular, non-first person selects the future stem (e.g. *ha* for Class A).

### 1.2.2. Ego (1) vs. non-ego

The second split that will concern us targets primarily first person as opposed to all others. This pattern is clearly witnessed by the subject agreement markers (*-u/-a*) and the vowel alternation in the past marker (*-?oh/-?eh*). In the non-singular, we observe the basic binary opposition relating to first person, whereas in the singular there is an additional split that additionally singles out the third person for subject agreement (*-i*) or the second person for past tense (*-?ah*). The basic ego/non-ego split can be observed with present stem alternation: in Class A, first person shows preservation of the stem-final vowel and suppression of the person/number marker, while second and third person undergo clipping. In Class C, the pattern is reversed, witnessing stem clipping for first person, and stem preservation as well as suppression of person/number marking in the other cells.

Comparing the two patterns, it is clear that first person plays a pivotal role, so it makes sense to collapse the two, regarding monofocal/polyfocal alternation as a more restricted case of ego/non-ego alternation.

### 1.2.3. Morphemic splits (2sg/1pl vs. default)

The third alternation described by Young (1964) involves, *inter alia*, the indicative mood marker. In contrast to the first two alternations, this split is more morphemic in nature, opposing the unnatural class of second singular and first plural, marked by *-ne*, to all other cells, marked by *-be*.

Nevertheless, even this split generalises to other markers, namely the emphatic marker *-na/-ta* and the question marker *-fi/-pi*. These two markers are found in the vicinity of the mood marker. Alongside alternating markers, we also find invariant ones in this domain, including the volitive mood marker *-he*, which contrasts with the indicative, and the “reason” marker *-gi*, which contrasts with the question marker.

## 1.3. Discussion

The fact that allomorphic patterns in Benabena are recurrent and can be found across different syntagmatic classes poses non-trivial problems for classical approaches to realisational morphology, such as PFM (Stump, 2001) or AM (Anderson, 1992). While these approaches are of course capable to derive the correct surface forms, the fact that rules are organised into rule blocks makes it difficult to capture the generalisation that e.g. monofocal/polyfocal allomorphy affects exponents in different syntagmatic classes. In order to do full justice to systematic patterns of allomorphy, as witnessed in Benabena, it appears necessary to group different rules of exponence according to their allomorphic properties. In the next section, I shall argue that multiple inheritance hierarchies of realisational rules, as postulated in Information-based Morphology (Crysmann and Bonami, 2016), provide the necessary prerequisite to generalise simultaneously over properties of exponence, position and allomorphic pattern.

## 2. Analysis

The analysis of Benabena allomorphic patterns will be formalised within the framework of Information-based Morphology (Crysmann and Bonami, 2016). IbM is an inferential-realisation model of morphology that systematically builds on the logic of typed feature structures (Carpenter, 1992) to extract generalisations over realisation rules, i.e. realisation rules are organised in a multiple inheritance type hierarchy. In addition to (monotonic) inheritance, IbM uses Online Type Construction (Koenig and Jurafsky, 1994) to capture systematic alternations. Rule descriptions themselves are pairings of the morphosyntactic properties to express (MUD) with a set of exponent morphs (MPH), indexed for position. Rules can further be allomorphically conditioned on additional properties of the morphosyntactic property set (MS).

### 2.1. Motivated splits

The central intuition about systematic allomorphy is that rules should not only be grouped together in terms of the properties they express or their syntagmatic class, but also in terms of the allomorphic conditions they undergo. I shall therefore setup a cross-classifying type hierarchy (Figure 1) that simultaneously presents two views of the realisation rules at the leaves.

The two rule types on the top left generalise common properties shared by all variants of the realisation of a morphosyntactic property. Thus, they pair a MUD value (=morphology under discussion, i.e. the property expressed by the rule) with constraints on position and shape: e.g. all realisations of *sbj* are constrained to surface in position class (PC) 2, and the shape (PH) is constrained to be a vowel. Similarly, the most general type for the realisation of *pst* equally has a position class constraint, but the shared restrictions on shape are more specific, fixing the quality of the consonants, but leaving the vowel underspecified. All these properties are inherited by the specific subtypes.

Complementary to the types for realisation, I have set up a subhierarchy on the right that defines conditions on allomorphy. The highest type here constrains all its subtype to be conditioned on non-first person subject agreement. Its immediate subtypes now are just intersections of the general constraints on feature expression discussed above and the allomorphic constraint. The specific restrictions on phonological shape are then defined as respective subtypes of these intersections. Note that in the present case, we find again a generic realisation, as well as constraints for the shape of exceptional cells. Since rule application in IbM implements Panini's principle, only exceptional and marked variants display constraint on allomorphic conditioning.

For reasons of space, I have contented myself with presenting the ego/non-ego split. I shall therefore only briefly comment how the basic type hierarchy can be monotonically extended to incorporate monofocal/polyfocal alternations. All it takes is to define a more specific allomorphic condition as a subtype of our non-first constraint, imposing *non-sg* as a further allomorphic constraint. The specific constraints regarding the phonology of the allomorphically conditioned variant will just be embedded one level further down in the type hierarchy.

Similarly, rules of exponence where no allomorphic alternation can be observed are even more straightforwardly integrated into the system. All it takes is to insert the rule type high enough to be outside the scope of any allomorphy constraint, e.g. directly under the top. Absence of alternation then follows from the simple fact that there are just no alternate rules. As an example, I have included the rule for future tense.

## 2.2. Unmotivated splits

Now that we have seen how motivated splits can quite easily be captured using multiple inheritance hierarchies that are abstractions over rules, let us briefly discuss how unmotivated splits can be incorporated. The special twist with unmotivated splits is that they do not involve a natural class, so we cannot easily use underspecification of individual values to pick out all the non-default forms at once. In the case at hand, we have first plural and second singular pattern together, forming a distributed disjunction. However, the fact that the exact same split occurs with different markers makes it quite unsatisfactory to stipulate the non-default variant separately for each of the two exceptional cells.

What we need is a way to systematically distribute one constraint over two types. To do this, we can use Online Type Construction (Koenig and Jurafsky, 1994): setting up the constraints on expression and allomorphy in two separate dimension, Online Type Construction will systematically intersect, under unification, every leaf type from one dimension with every leaf type from the other. As a result, constraints on expression will be distributed over the types defining the allomorphic conditions that identify exceptional cells. The default variants are excluded from this since their types are already pre-linked to a type in the ALLOMORPHY dimension. As always, Panini's principle will ensure that the more general rules can only apply to the complement of the exceptional cells.

## 3. Conclusion

I have discussed recurrent allomorphic splits in Benabena verb morphology, and argued, following Young (1964), that both motivated and unmotivated splits permeate the system in highly systematic ways. I have concluded that this phenomenon necessitates generalisations about allomorphy across heterogeneous rules of exponence. I have shown that multiple inheritance type hierarchies over rules, as proposed in Crysmann and Bonami (2016), readily provide the expressive means to capture the generalisations, offering two simultaneous views on the relatedness of rules, i.e. in terms of allomorphic conditions, or in terms of expression. Furthermore, the current approach is easily extended to capture the similarity of the two motivated patterns identified by Young (1964): all that is necessary is to provide a non-first non-singular subtype under which polyfocal rules can be embedded. Finally, the difference between motivated and unmotivated, yet recurrent allomorphic patterns was formally represented as the difference between static and dynamic multiple inheritance.

## References

- ANDERSON, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- CARPENTER, R. (1992). *The Logic of Types Feature Structures*. Cambridge: Cambridge University Press.
- CRYSMANN, B. & BONAMI, O. (2016). 'Variable morphotactics in information-based morphology'. *Journal of Linguistics*, 52:311–374.

KOENIG, J.-P. & JURAFSKY, D. (1994). ‘Type underspecification and on-line type construction’. In *Proceedings of WCCFL XIII*. 270–285.

STUMP, G. T. (2001). *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

YOUNG, R. (1964). ‘The primary verb in Benabena’. In B. F. Elson (ed.), *Verb Studies in Five New Guinea Languages*. Norman: Summer Institute of Linguistics, 45–83.

	Class A: <i>ho/he/ha</i>			Class B: <i>bu/bi</i>			Class C: <i>fi</i>		
	PRS	PST	FUT	PRS	PST	FUT	PRS	PST	FUT
SG									
1	ho-be	ho-?oh-u-be	ha-l-u-be	b-u-be	bu-?oh-u-be	bi-l-u-be	f-u-be	fi-?oh-u-be	fi-l-u-be
2	h-a-ne	ho-?ah-a-ne	ha-l-a-ne	b-a-ne	bu-?ah-a-ne	bi-l-a-ne	fi-ne	fi-?ah-a-ne	fi-l-a-ne
3	h-a-be	ho-?eh-i-be	ha-l-i-be	b-i-be	bu-?eh-i-be	bi-l-i-be	fi-be	fi-?eh-i-be	fi-l-i-be
DU									
1	ho-?i-be	ho-?oh-u-?i-be	ha-l-u-?i-be	b-u-?i-be	bu-?oh-u-?i-be	bi-l-u-?i-be	f-u-?i-be	fi-?oh-u-?i-be	fi-l-u-?i-be
2	h-a-?i-be	he-?eh-a-?i-be	ha-l-a-?i-be	b-a-?i-be	bi-?eh-a-?i-be	bi-l-a-?i-be	fi-?i-be	fi-?eh-a-?i-be	fi-l-a-?i-be
3	h-a-?i-be	he-?eh-a-?i-be	ha-l-a-?i-be	b-a-?i-be	bi-?eh-a-?i-be	bi-l-a-?i-be	fi-?i-be	fi-?eh-a-?i-be	fi-l-a-?i-be
PL									
1	ho-ne	ho-?oh-u-ne	ha-l-u-ne	b-u-ne	bu-?oh-u-ne	bi-l-u-ne	f-u-ne	fi-?oh-u-ne	fi-l-u-ne
2	h-a-be	he-?eh-a-be	ha-l-a-be	b-a-be	bi-?eh-a-be	bi-l-a-be	fi-be	fi-?eh-a-be	fi-l-a-be
3	h-a-be	he-?eh-a-be	ha-l-a-be	b-a-be	bi-?eh-a-be	bi-l-a-be	fi-be	fi-?eh-a-be	fi-l-a-be

Table 1: Benabena verb inflection (indicative mood *be/ne*)

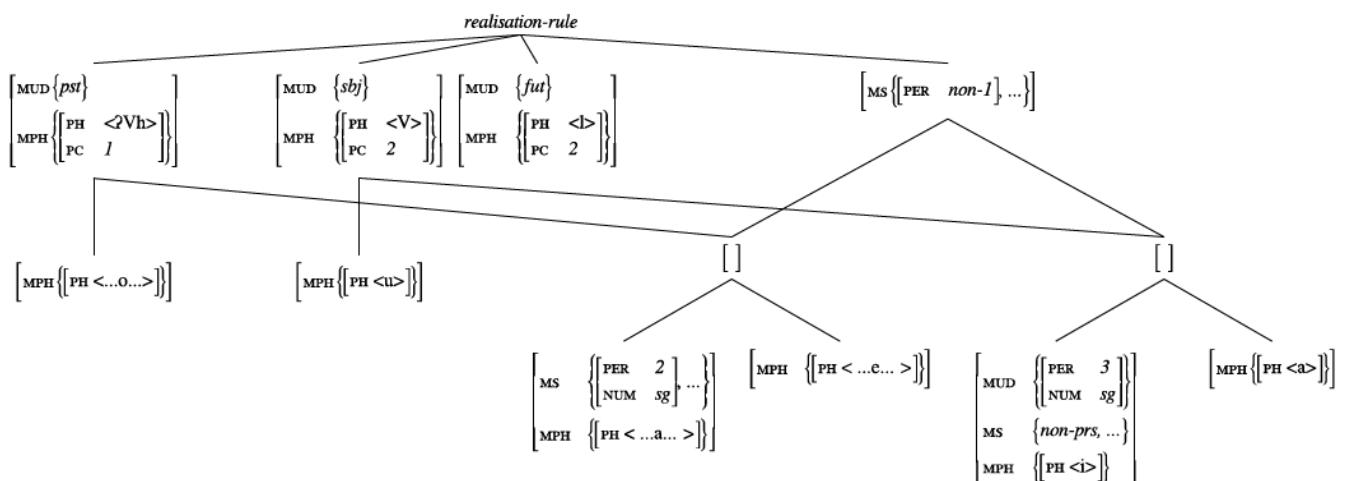


Figure 1: Type hierarchy of Benabena ego/non-ego split

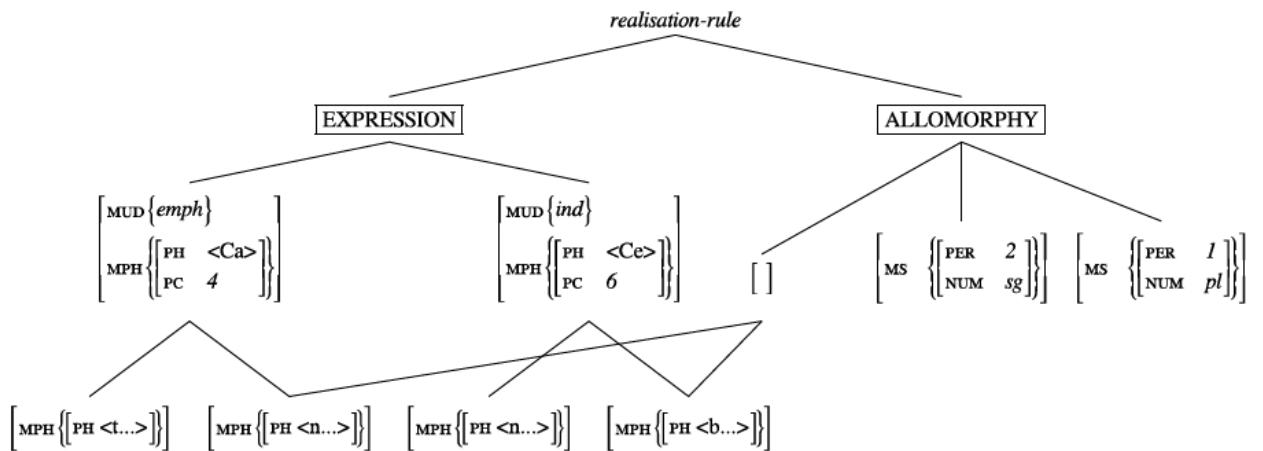


Figure 2: Two-dimensional type hierarchy for Benabena unmotivated allomorphy

## Word formation in general and special languages: the case of super- and s- prefixations

**Edwige DUGAS,**

*UMR CNRS 8163, Université de Lille*

[edwige.dugas@gmail.com](mailto:edwige.dugas@gmail.com)

**Jérôme MICHAUD,**

*Department of Physics and Astronomy, Uppsala Universitet*

[jerome.michaud@soc.uu.se](mailto:jerome.michaud@soc.uu.se)

This talk addresses the issue of the nature of special languages (henceforth SL) and how different they are from general language (henceforth GL) when it comes to morphological description and morphological theory, in the framework of Construction Grammar (henceforth CxG). As a case study, we consider the [*super-X*] word-formation pattern in the English SL of particle physics and, more specifically, of supersymmetry (abbreviated as SUSY) theory (supergeometry, supermultiplet) and compare it with the [*super-X*] word-formation pattern in English GL (SUPERHERO, SUPERTHIN). We make the hypothesis that the two constructions [*super-X*]<sup>GL</sup> and [*super-X*]<sup>SL</sup> share common properties<sup>1</sup> and that they are linked in the constructionon (Jurafsky 1991), the network of constructions of the language; our main objective is to propose a constructional network in which all [*super-X*] constructions are connected. In this study, we also examine the [*s-X*] word-formation pattern, which is specific of the SL of SUSY theory (SELECTRON, SQUARK), as the data shows that this pattern is in many respects similar to the [*super-X*]<sup>SL</sup> pattern.

For the purpose of our study, we have gathered and annotated a corpus of [*super-X*]<sup>SL</sup> and [*s-X*] lexemes, the characteristics of which we compared with the [*super-X*]<sup>GL</sup> constructions already described in the literature (Bauer 1983, Plag 2003, Lieber 2005, Bauer et al., 2013). We show that it is possible to represent all [*super-X*] constructions in a constructional network, where they are vertically (inheritance relations) and horizontally (intra/interparadigmatic links) related and we describe a new type of link representing a reduction process from lexeme to prefix. For instance, we argue that the prefix *super-* of the [*super-X*]<sup>SL</sup> construction is actually a reduction of the lexeme SUPERSYMMETRY and that it carries the meaning of this lexeme and conveys it to the whole construction: in the SL of particle physics, any lexeme formed with the prefix *super-* refers to an entity which is in relation to SUSY.

### 1. Background

#### 1.1. Construction morphology

This study is grounded in CxG (Langacker 1987, Croft 2004, Goldberg 2006, Booij 2010), a cognitive theory whose unit of analysis is the construction. Constructions are generally defined as conventionalized and entrenched symbolic pairings of form and meaning, and language is seen as a network of constructions of different nature (schematic/semi-schematic/substantial; more or less atomic/complex; more or less lexical/grammatical).

---

<sup>1</sup> In this abstract, “SL” refers to the SL of supersymmetry.

Constructions are organized hierarchically and each construction inherits the properties of a “higher”, more schematic construction.

### 1.2. GL vs SL

The main discursive objective of the scientific discourse is to name new referents and to pass on theoretical or practical knowledge. SLs can be seen as subsystems of GL (Castellví 1999). As Sager puts it, “[b]esides containing a large number of items which are endowed with the property of a special reference the lexicon of a special language also contains items of general reference which do not usually seem to be specific to any discipline or disciplines whose referential properties are uniformly vague or generalized” (Sager 1990: 19). We try to show to what extent  $[super\text{-}X]^{GL}$  and  $[super\text{-}X]^{SL}$  patterns can receive a similar morphological description, although they seem to be the product of different processes of lexical creativity (see Section 3).

### 1.3. SUSY in a nutshell

Particle physics is the subdomain of physics that studies the elementary particles our universe is made of and their interactions. In the Standard Model (SM) of particle physics, there are a limited number of elementary particles that are usually described by their *quantum numbers*. The most important quantum numbers are the mass, the charge and the spin of the particles.<sup>2</sup> In the SM, every particle is associated with its antiparticle that has the same mass, the same spin, but an opposite sign of charge. This sign of charge symmetry is better known as the matter/antimatter symmetry. Furthermore, when a particle interacts with its antiparticle, they annihilate each other, turning their mass into energy.

In analogy with the matter/antimatter symmetry, SUSY is a symmetry of spin that associates to every fermion (particle with half-integer spin) a corresponding boson (particle with integer spin), whose name is formed by *s-* prefixation. For example, the corresponding particle of the *muon* is called *smuon*. SUSY also associates to every boson a corresponding fermion, whose name is formed by suffixation by *-ino*. For example, the corresponding particle of the *Higgs boson* is called *Higgsino*. The particles postulated in SUSY are called *superparticles* or *sparticles* and must be much more massive than standard particles, otherwise they would have been already discovered in experiments. The motivation for this additional symmetry is that it provides a natural solution to some difficult problems of the SM. For example, it offers solutions to the *hierarchical problem*<sup>3</sup> and automatically includes gravitation inside the model while the SM fails to do so. It also provides a candidate *dark matter* particle.

## 2. Data and methodology

This study draws on a corpus of lexemes in *super-* and *s-* used in SUSY. We chose not to build a corpus of GL *super-* lexemes but rather to use the data presented in the literature. The SL corpus, by definition, only contains terminological concepts and is smaller than corpora usually gathered in studies on GL phenomena. The corpus of SL *super-* and *s-* lexemes has been collected from the following papers, all aiming at introducing SUSY theory: Miyazawa 1968, DeWitt 1992, Bilal 2001, Haber 2013, Martin 2016). The online Oxford English

<sup>2</sup> The charge and spin of a particle are quantified and can only take integer and half-integer values, respectively.

<sup>3</sup> The hierarchical problem has to do with the relative intensity of the different forces and can only be accounted for in the SM by a huge amount of fine-tuning of parameters.

Dictionary<sup>4</sup> and the Wikipedia entry “Supersymmetry”<sup>5</sup> have also been used. To our knowledge, this data has not been exploited in the linguistic literature so far; studies on *super-*formations have only dealt with their use in GL and no work has been done on *s-* formations. All constructs have been annotated with respect to the criteria summarized in Table 1.

Criteria	GL	SL
Category of the base	N, Adj, V	N
Complexity of the base (polylexical units?)	yes	yes
Semantics of the base	any referent which can be evaluated	mathematical concepts, physical mechanisms, particles
Category of the derived lexeme	N, Adj, V	N
Semantics of the derived lexeme	endocentric	endocentric/exocentric

**Table 1** — Annotation criteria for *super-* and *s-* lexemes

Tests have been designed to differentiate between the GL and SL uses of *super-*, as presented in Table 2.

Test	Description	Example
1	The referent of the base pertains to supersymmetry	SUPER HIGGS MECHANISM
2	The base can also be prefixed by SL <i>anti-</i>	SUPERMATTER/ANTIMATTER
3	The referent of the base and the referent of the derived word are superpartners	SUPERELECTRON
4	The base can also be prefixed by <i>s-</i> , with the same meaning	SUPERELECTRON/SELECTRON
5	The referent of the base is “in a class above” (qualitatively or quantitatively)	SUPERHERO

**Table 2** — Tests for *super-* lexemes

The *s-* pattern is easier to identify since we can rely on a single test: the referent of the derived word must be a superpartner of the base (*smuon*, *selectron*). The test 4 described in Table 2 can also be used (the superpartner of a *muon*, for example, can be called a *smuon* or a *supermuon*), although the variants with the prefix *s-* are much more frequent.<sup>6</sup>

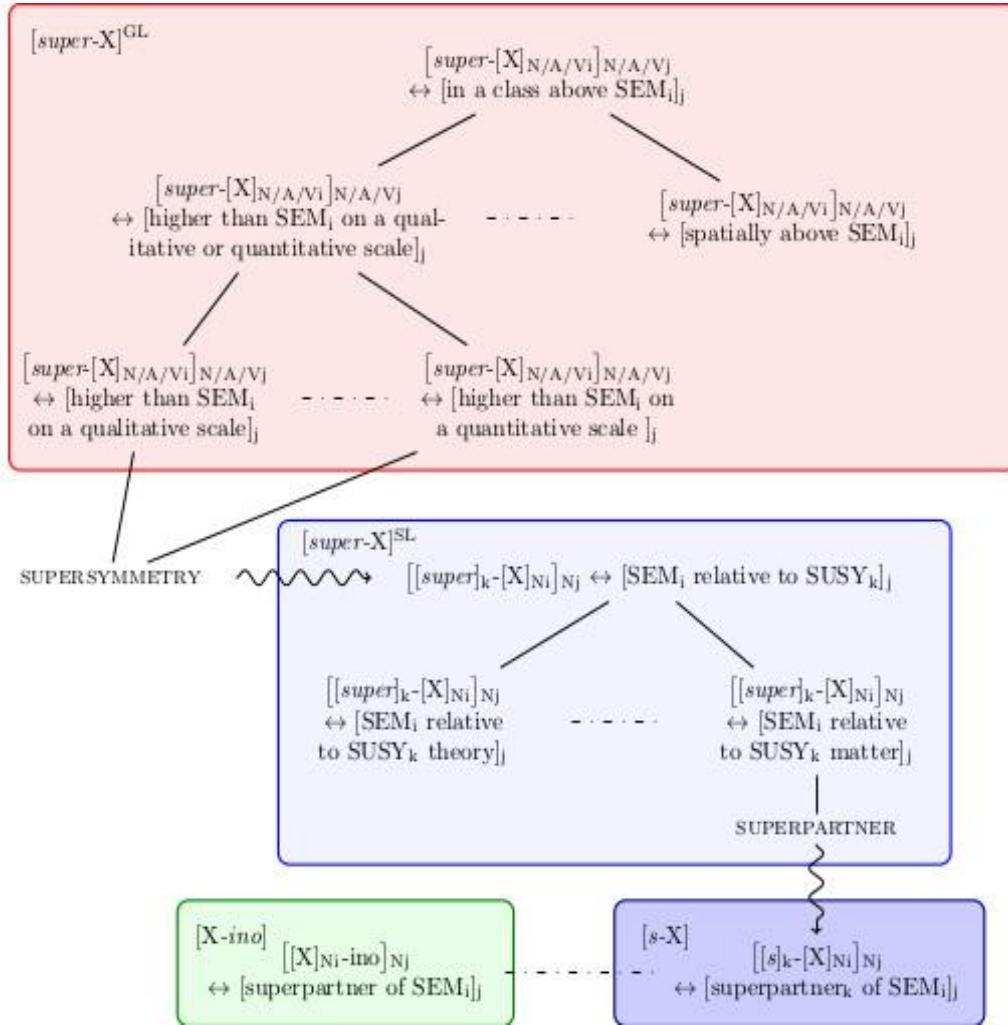
### 3. Results and discussion

Our results are summarized in the constructional network shown in Figure 1. For the ease of presentation, we have used colored boxes to group families of constructions. For example, the box “[super-X]<sup>GL</sup>” groups all the GL uses of the [super-X] pattern, whereas the box “[super-X]<sup>SL</sup>” groups the uses of the [super-X] pattern in the context of SUSY.

<sup>4</sup> <http://www.oed.com>

<sup>5</sup> <https://en.wikipedia.org/wiki/Supersymmetry>

<sup>6</sup> For example, a request on GoogleScholar with *smuon* yields 1900 hits while a request with *supermuon* yields only 3 hits (queries made on May 5th, 2017).



**Figure 1** — Constructional network of  $[super\text{-}X]$  constructions. Construction  $[X\text{-}ino]$  is also included for completeness. Straight lines represent hierarchical links; dash-dotted lines represent paradigmatic links; wavy arrows represent a process of reduction from lexeme to prefix.

In GL, *super-* prefixation is generally described as an instance of evaluative morphology, i.e. a morphological process used to express augmentation, diminution, approval and pejoration (Mel'cuk 1994, Barbaresi 2006, Bauer et al., 2013). Drawing on the analyses provided in the literature (Bauer 1983, Plag 2003, Lieber 2005, Bauer et al. 2013), we argue that the  $[super\text{-}X]^{\text{GL}}$  pattern can have a spatial interpretation or an evaluative one, and that evaluation can be qualitative or quantitative. In Figure 1, the construct SUPERSYMMETRY has been left outside of the boxes because, as we show in our talk, the *super-* prefix used to form SUPERSYMMETRY cannot be analyzed as a GL prefix and is also different from the  $[super\text{-}X]^{\text{SL}}$  construction. One can argue that the *super-* prefix of SUPERSYMMETRY possesses some of the characteristics of evaluative morphology, since, as explained in Section 1.3, SUSY increases the explanatory power of the SM of particle physics, which makes it qualitatively better than other symmetries, such as the matter/antimatter symmetry. Moreover, the superparticles postulated must be more massive than the particles of the SM, which contributes to a quantitative evaluation of SUSY. Therefore, the lexeme SUPERSYMMETRY inherits both from the qualitative and quantitative uses of the  $[super\text{-}X]^{\text{GL}}$  construction, while semantically belonging to the SL lexicon of particle physics. This constitutes an interesting example of lexical creativity.

In Figure 1, wavy arrows are used to represent the link between the construct SUPERSYMMETRY and the construction  $[super\text{-}X]^{SL}$  on the one hand, and between the construct SUPERPARTNER and the construction  $[s\text{-}X]$  on the other hand. In the  $[super\text{-}X]^{SL}$  construction, the prefix *super*- is a reduction of the lexeme SUPERSYMMETRY and carries its meaning: in the SL of particle physics, any lexeme formed with the prefix *super*- refers to an entity which is in relation to SUSY. In the  $[s\text{-}X]$  construction, the prefix *s*- is a reduction of the lexeme SUPERPARTNER and carries its meaning: any lexeme formed with the prefix *s*- refers to a particle which is the superpartner of the particle denoted by the base noun. This reduction process generating new prefixes from lexemes illustrates the lexical creativity of SUSY SL. We discuss the nature of *super*- and *s*- and whether they should be analysed as prefixes, as splinters or as blends.

The  $[super\text{-}X]$  SL construction gives its properties to two subconstructions, namely a construction where the derived lexeme refers to SUSY theory, as in SUPERALGEBRA, and a construction where the derived lexeme refers to SUSY matter, as in SUPERPARTNER. We chose to differentiate between these two constructions mainly because of the analogy between SUSY and the matter/antimatter symmetry. For instance, all constructs that refer to SUSY matter have a corresponding construct prefixed by *anti*- in the SM (they pass test 2 of Table 2), whereas the constructs that refer to SUSY theory do not. The types of bases these two constructions accept are also different, since polylexical base nouns can only occur in the SUSY theory  $[super\text{-}X]^{SL}$  construction (SUPER POINCARÉ GROUP, SUPER HIGGS MECHANISM).

Finally, the  $[s\text{-}X]$  and  $[X\text{-}ino]$  constructions are used to form lexemes referring to particle names such as SFERMION, SQUARK and HIGGSINO, GLUINO, depending on their fermionic or bosonic nature.

## References

- BAUER, L. (1983). *English word-formation*. Cambridge: Cambridge University Press.
- BAUER et al., L. (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- BENVENISTE, E. (1966). *Problèmes de linguistique générale*, volume 2. Paris: Gallimard.
- BOOIJ, G. (2010). *Construction morphology*. New York: Oxford University Press.
- CASTELLVÍ, M.-T. (1999). *Terminology: Theory, methods and applications*, volume 1. Amsterdam/Philadelphia: John Benjamins Publishing.
- CROFT, W. & CRUSE, D. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- GOLDBERG, A. (2006). *Constructions at work: The nature of generalization in language*. New York: Oxford University Press.
- JURAFSKY, D. (1991). *An on-line computational model of human sentence interpretation: A theory of the representation and use of linguistic knowledge*. Technical report, DTIC Document.
- LANGACKER, R. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*, volume 1. Stanford: Stanford University Press.
- LIEBER, R. (2005). “English word-formation processes”. In P. Stekauer & R. Lieber, Eds., *Handbook of word-formation*, pp. 375–427. Berlin: Springer.
- PLAG, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press.
- SCHMID, H.-J. (2010). “Does frequency in text instantiate entrenchment in the cognitive system?” pp. 101–133. Berlin, New York: Mouton de Gruyter.

## Morphomic structure(s) and the Occitan synthetic conditional

Louise ESHER,

CNRS CLLE-ERSS, Université Toulouse Jean Jaurès

louise.esher@univ-tlse2.fr

### 1. Outline

In varieties of Occitan (Gallo-Romance), exponents of the TAM category conventionally labelled ‘synthetic conditional’ (henceforth COND) typically participate in strong and systematic implicational relationships with exponents of the categories ‘synthetic future’ (FUT) and ‘imperfect indicative’ (IPFV.IND) respectively. In this paper I outline the historical origin and subsequent development of these morphological relationships, which I treat as MORPHOMIC in the sense of Aronoff (1994).

The COND-IPFV.IND relationship is of particular interest: firstly, it concerns syncretism of inflectional desinences; secondly, the paradigm cells involved vary according to inflectional class. This behaviour contrasts with other morphomic patterns of paradigmatic distribution (METAMORPHOMES) identified for the Romance verb – including the COND-FUT relationship – which concern syncretism of stem material and which consistently affect the same set of cells independently of inflectional class.

### 2. Origin and traditional descriptions of the Romance COND

The Romance COND has its origin in a periphrasis collocating the infinitive and a ‘past’ form of the verb HABERE ‘have’ (in Gallo-Romance, the IPFV.IND, e.g. CANTARE HABEBAT > Fr. *chanterait* ‘sing.COND.3SG’). A parallel periphrasis, collocating the infinitive and the PRS.IND of HABERE, gives rise to the Romance synthetic future (henceforth FUT) in these varieties (e.g. CANTARE HABET > Fr. *chantera* ‘sing.FUT.3SG’). In most (though not all) modern varieties of Occitan, the COND continues to display inherited formal similarities both with the FUT and with non-first-conjugation IPFV.IND forms; a selection of illustrative FUT, COND and IPFV.IND forms is shown in Table 1.

		Nice	Var	Nontron	Sénaillac-Lauzès
I	COND	canteria	cantarié	parlariá	dintrariá
	FUT	canterà	cantara	parlará	dintrará
	IPFV.IND	cantavi	cantavo	parlava	dintrava
IV	COND	finisseria	fenirié	legiriá	floririá
	FUT	finisserà	fenira	legirá	florirá
	IPFV.IND	finissia	fenissié	legissiá	floriá
III	COND	bateria	metrié	metriá	vendriá
	FUT	baterà	metra	metrá	vendrá
	IPFV.IND	batia	metié	metiá	vendiá

**Table 1** — 3SG COND, FUT and IPFV.IND forms for exemplar continuants of Latin conjugations I, IV and III (*cantar* ‘sing’, *parlar* ‘speak’, *dintrar* ‘enter, return home’, *finir* ‘finish’, *legir* ‘read’, *florir* ‘flower’, *batre* ‘beat’, *metre* ‘put’, *vendre* ‘sell’) in the Occitan varieties of Nice (Toscano 1998), Var (Domènega 2002), Nontron (Reydy 2008), Sénaillac-Lauzès (Sibille 2015). Orthography is that of the source material.

The formal parallels of Gallo-Romance COND forms with FUT and IPFV.IND forms respectively have influenced analysis of the COND both in terms of its semantics (see e.g. Iatridou 2000, Vincent 2013) and in terms of its historical development: Ronjat's detailed historical-comparative survey of Occitan varieties allots a mere half-page to the COND (1937:231), reducing it to the stem found in the future and the desinences found in IPFV.IND forms of class II verbs (i.e. those continuing Latin conjugation III), with occasional local particularities; in the much longer discussion of the FUT, (1937:207-212), and of the IPFV.IND (1937:170-176), reference to the COND is rare (1937:173-175, 211); and a similar approach can be found in classic historical grammars of French, such as that of Pope (1934:365). Such analyses attest to the strength of the implicational relationships between the exponents of the COND, FUT and IPFV.IND: the identities between exponents are so systematic that treating the COND as the sum of its (FUT and IPFV.IND) parts is repeatedly considered by authors to provide an informative and sufficient description of the data.

### **3. The Romance COND/FUT relationship, where it exists, is a METAMORPHOME**

For Romance, Maiden (e.g. 2009a, 2016) identifies a number of abstract yet systematic and recurrent paradigmatic distribution patterns: for example, a set of cells comprising the 1SG.PRS.IND and all PRS.SBJV forms, and a set of cells comprising the reflexes of Latin *perfectum* forms, typically the PFV.PST.IND and IPFV.SBJV. Distributional patterns of this type are termed METAMORPHOMES by Round (2015); for consistency and clarity I will use this term throughout, independently of individual authors' usage. Using comparative and diachronic data, Maiden demonstrates that a range of inflectional phenomena, including novel stem allomorphy, suppletion, defectiveness and so-called empty morphs, are distributed according to existing metamorphomic templates; in one Romanian case, the analogical extension of a 2PL desinence is argued to follow a metamorphomic template (Maiden 2009b). Nevertheless, all established Romance metamorphomes have been identified based on the paradigmatic distribution of root allomorphy, and indeed Maiden (2013:518) argues that the specific advantage of metamorphomes 'lies in minimizing the arbitrariness of the lexical sign' by distributing arbitrary variants according to a consistent schema, and thus ensuring a predictable relationship between [lexical] form and meaning.

The implicational relationship between COND forms and FUT forms in most Occitan varieties readily lends itself to a metamorphomic analysis. In synchrony, FUT and COND forms are observed to share a distinctive stem, which may be suppletive, or display a characteristic phonological shape across lexemes; in diachrony, the set of FUT and COND cells is observed to act as a template for analogical changes, and the shape of the metamorphome is observed to change following regular sound changes which modify the form of exponents (Esher 2013, 2015). These behaviours are entirely parallel with those of the metamorphomes described for Romance by Maiden. A further property observed for metamorphomes in Occitan, including the set of FUT and COND cells, is that these paradigmatic distribution templates are near-exceptionlessly valid for all lexemes, applying independently of inflectional class distinctions.

### **4. The Romance COND/IPFV.IND relationship is morphemic... but of what kind?**

The relationship of COND and IPFV.IND forms concerns inflectional desinences as opposed to roots, stems or other 'lexical' material. Yet its behaviour in other respects recalls that observed for Romance metamorphomes. For the varieties shown in Table 1, and many others, the identity between COND desinences and non-first-conjugation IPFV.IND desinences is systematic and exceptionless. Moreover, there is evidence that this pattern of syncretism

serves as a productive template for analogical change, albeit on a more restricted scale (fewer examples are known, while distribution of root material is not found, nor are cases of suppletion or defectiveness).

In many Occitan varieties, final /i/ has been generalised as a exponent of 1SG and progressively spread by analogy from its etymological source in the PRS.IND and PFV.PST.IND to other paradigm categories. While the extent of spread is highly variable, without exception the COND and the non-first-conjugation IPFV.IND pattern together: either all forms retain their historically regular outcome without final -i (*cantariá* ‘sing.COND.3SG, I’, *finiriá* ‘finish.COND.3SG, IV’, *finissiá* ‘finish.IPFV.IND.3SG, IV’, *batriá* ‘beat.COND.3SG, III’, *batiá* ‘beat.IPFV.IND.3SG, III’) or all forms acquire -i (*cantariái*, *finiriái*, *finissiái*, *batriái*, *batiái*), with no intermediate system being observed (Esher 2017). A further example is found in varieties of Nice and the surrounding area, where a formative characteristic of the first-conjugation IPFV.IND is extended to 1PL and 2PL forms of the COND (all conjugations) as well as of the non-first-conjugation IPFV.IND (Table 2).

<i>parlar</i> ‘speak’			<i>aver</i> ‘have’	
	IPFV.IND	COND	IPFV.IND	COND
1SG	parlavi	parlerii	avii	aurii
2SG	parlaves	parleries	avies	auries
3SG	parlava	parleria	avia	auria
1PL	parlavam	parleriam, parleriavam	aviam, aviavam	auriam, auriavam
2PL	parlavatz	parleriatz, parleriavatz	aviatz, aviavatz	auriatz, auriavatz
3PL	parlavan	parlerian	avian	aurian

**Table 2** — IPFV.IND and COND forms for first- and non-first conjugation verbs, Nice (Gasiglia 1984).

The source of the distribution likewise recalls that of some other Romance metamorphomes: as with the reflexes of Latin *perfectum* forms, which preserve their inherited identity of stem, the syncretism between COND and non-first-conjugation IPFV.IND desinences continues an identity between forms ending in -EBAM, etc.

The relationship of COND and IPFV.IND forms thus presents characteristics which correspond to those of known metamorphomes, and to Blevins’ definition of morphemes as ‘recurrent units of predictiveness’ (2016:106). Furthermore, a notion of MORPHOMIC SUFFIX has already been implemented within the framework of Paradigm Function Morphology (Stump 2016:120–146), providing the formal wherewithal to describe morphemes not involving stem material.

The type of morpheme to be invoked here merits careful consideration. Metamorphomes (whether concerning stems, suffixes or other material) are customarily defined with reference to a single paradigm which is assumed to hold across lexemes, and exist independently of conjugational class. By contrast, syncretism between IPFV.IND and COND exponents for a given lexeme only obtains within a subset of conjugational classes; but the exponents involved also occur outside this subset, appearing in the COND of all conjugational classes as well as in non-first-conjugation IPFV.IND forms. Thus, an analysis which considers only the syncretism relations obtaining within the paradigm of a given lexeme is readily modelled using rules of referral or statements about paradigmatic distribution, but misses a generalisation about the exponents of the COND across conjugations; whereas an analysis describing the exponents in terms of MEROMORPHOMES (morphemic mappings which ‘pertain to sets of word formation operations and [...] derive the pieces of individual word forms’, Round 2015:29–30) captures the intuition of a single set of exponents for the categories COND

and non-first-conjugation IPFV.IND, but assumes a more complex and involved relationship between morphemes of different types, as it requires meromorphemic mappings to take account of inflectional class (RHIZOMORPHOME) distinctions.

## References

- ARONOFF, M. (1994). *Morphology by itself*. Cambridge, Mass.: MIT Press.
- BLEVINS, J. (2016). *Word and Paradigm Morphology*. Oxford: OUP.
- DOMENGE, J.-L. (2002). *Grammaire du provençal varois. Le Verbe*. La Farlède: Association varoise pour l'enseignement du provençal.
- ESHER, L. (2013). Future and conditional in Occitan: a non-canonical morpheme. In S. Cruschina, M. Maiden & J. C. Smith (eds.), *The Boundaries of Pure Morphology: Diachronic and Synchronic Perspectives*. Oxford: OUP. 95–115.
- ESHER, L. (2015). Formal asymmetries between the Romance synthetic future and conditional in the Occitan varieties of the Western Languedoc' (2015). *Transactions of the Philological Society* 113(2). 249–270.
- ESHER, L. (2017). L'aventura de la desinéncia 1SG -i dins les parlars de Lengadòc. *Tenso* 32(1–2), 19–40.
- GASIGLIA, R. (1984). *Grammaire du nissard*. Nice: Institut d'Etudes Niçoises.
- IATRIDOU, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry* 31. 231–270.
- MAIDEN, M. (2009a). From pure phonology to pure morphology. The reshaping of the Romance verb. *Recherches linguistiques de Vincennes* 38. 45–82.
- MAIDEN, M. (2009b). Un capitolo di morfologia storica del romeno: preterito i tempi affini. *Zeitschrift für Romanische Philologie* 125. 273–309.
- MAIDEN, M. (2013). The Latin ‘third stem’ and its Romance descendants. *Diachronica* 30(4). 492–530.
- MAIDEN, M. (2016). Morphemes. In A. Ledgway & M. Maiden (eds.), *The Oxford Guide to the Romance Languages*. Oxford: OUP. 708–721.
- POPE, M. K. (1934). *From Latin to modern French with especial consideration of Anglo-Norman*. Manchester: Manchester University Press.
- REYDY, J.-P. (2008). *Notre occitan: Le dialecte du Périgord-Limousin parlé dans le Parc naturel régional*. Limoges: IEO Limousin.
- RONJAT, J. (1937). *Grammaire istorique des parlers provençaux modernes*, vol. III. Montpellier: Société des Langues Romanes.
- ROUND, E. (2015). Rhizomorphomes, meromorphomes and metamorphomes. In M. Baerman, D. Brown & G. Corbett (eds.), *Understanding and measuring morphological complexity*. Oxford: OUP. 29–52.
- SIBILLE, J. (2015). *Description de l'occitan parlé à Sénaillac-Lauzès (Lot) et dans les communes voisines*. Lemòtges: Editions Lambert-Lucas.
- STUMP, G. (2016). *Inflectional paradigms*. Cambridge: CUP.
- TOSCANO, R. (1998). *Gramàtica niçarda*. Pau: Princi Néguer.
- VINCENT, N. (2013). Compositionality and change in conditionals and counterfactuals in Romance. In S. Cruschina, M. Maiden & J. C. Smith (eds.), *The Boundaries of Pure Morphology: Diachronic and Synchronic Perspectives*. Oxford: OUP. 116–136.

## Lexical categories and semantic primitives: the case of Spanish *-ncia*

**Antonio FABREGAS,**

*Universitet i Tromsø & CNRS (UMR8163)*

antonio.fabregas@uit.no

**Rafael MARÍN,**

*Université de Lille*

rafael.marin@univ-lille3.fr

### 1. Introduction

The study of morphological category-change has always been linked to the question of how lexical categories relate to each other, since at least Chomsky's (1965) featural decomposition (see also Halle 1973, Jackendoff 1975, Lieber 1981). In this context, one of the most suggestive lines of research has been the intuition that lexical categories share a basic vocabulary of primitives, such as a notion of boundedness, that is instantiated in different ways in each case (Mourelatos 1978, Bach 1986, Jackendoff 1997). Aspectual information, being the more discussed instantiation of (un)boundedness, lies at the heart of this question.

Inside the general program of research that explores the hypothesis that lexical categories share some basic vocabulary, Fábregas & Marín (2012) proposed the *Aspect Preservation Hypothesis*: the aspectual information of a deverbal nominalisation is built from the space of possibilities that the base verb allows. In other words: anything that is aspectual in nature, in the syntactic sense of licensing specific types of modifiers, prepositional phrases and auxiliary verb combinations, must come from a verbal structure. Nominalisers signal that the structure has moved from a verbal domain to a nominal domain, and at that point aspect cannot be defined: if anything, the notion of boundedness and its likes are interpreted at that point as count / mass, collective nominals, etc. In Fábregas & Marín (2012) it was claimed that nominalisers such as Spanish *-ción*, English *-(at)ion* and German *-ung* are transparent with respect to aspectual information, and are unable to ignore it or overwrite it. In that work, it was argued that when the nominalisation takes the lexical verb as its base (as opposed to a bigger verbal constituent containing grammatical aspect), state-denoting nominalisations were restricted to those verbal bases that, independently, contained a state subcomponent (typically in the form of a result state).

This kind of contrast brings initial plausibility to the claim that aspect must be defined in the verbal domain, but the hypothesis makes some other predictions that were not discussed in Fábregas & Marín (2012). If it is true that categories share some primitives, we expect also that some nominaliser must be sensitive to the presence or absence of those primitives in the bases they combine with, rather than to their lexically-specific instantiation in V, N or A. This is not visible in affixes like *-ción* / *-ation* / *-ung*, which are default deverbal nominalisers in their respective languages.

Our goal here is to present a case study of the Spanish nominaliser *-ncia*, which we will argue illustrates this situation: it combines with verbs and other bases provided that the interpretation of the base does not involve any change. This, we will argue, extends to all uses of this affix, and allows for a unification between what otherwise seems to be a quality-nominalisation use and an event-nominalisation use.

## 2. The puzzle: a suffix with a weird distribution

The suffix *-ncia* has not been much studied, with few exceptions (Malkiel 1945, Pattison 1975: 75, Pena 2005, NGLE 2009: §6.3p-v, Cano & Jaque 2011, Fábregas 2016: 355-364). Pharies (2002: 70-71) notes that in its Latin origin it was a morphologically complex form consisting on the oblique present participle form *-nt-* and the quality nominaliser *-ia*. Whether the same decomposition applies to the Spanish suffix *-ncia*, as *-nte* ‘-ant’ and *-ia* ‘-ity’ is a controversial matter.

### 2.1. Qualitys and state nominalisations

The most productive use of *-ncia* in contemporary Spanish is as a quality nominaliser glossed as ‘the property of being X’, where X is an adjective in *-nte* related to the base (1).

- (1) arrogancia ‘arrogance’, coherencia ‘coherence’, eficiencia ‘efficiency’, inteligencia ‘intelligence’, negligencia ‘negligence’, prudencia ‘caution’, ...

Next to quality nominalisations, *-ncia* also produces eventuality nominalisations, specifically state-denoting nominalisations from purely stative verbs (2):

- (2) creencia ‘belief’, dependencia ‘dependence’, existencia ‘existence’, ignorancia ‘ignorance’, pertenencia ‘membership’, precedencia ‘precedence’, preferencia ‘preference’, repugnancia ‘disgust’, tendencia ‘tendency’...

The existence of this second class of nominalisations motivates Puzzle #1 with *-ncia*:

- (3) *-ncia* can produce both quality and eventuality nominalisations

However, there is an intuitive relation between a quality nominalisation and a state: both are non-eventive. The minimal distinction, we assume, is that states are after all temporal objects that can be placed in time, while qualities are sets of properties without any temporal variable (cf. among many others Carlson 1977, Bosque 1990, Maienborn 2003). Then, it is not so surprising that *-ncia* produces, in addition to quality nominalisations, eventuality ones if those are restricted to stative verbs.

### 2.2. Eventive bases

However, the situation is not so simple: verbs that are clearly eventive can also be taken as bases by *-ncia*. This is Puzzle #2 of *-ncia* formation.

- (4) Even though most *-ncia* formations are related to bases without eventivity, there are formations with an eventive base.

Here are some of the *-ncia* formations with eventive bases.

- (5)
  - a. vigilar ‘watch’ > **vigilancia** ‘vigilance’, resistir ‘resist’ > **resistencia** ‘resistence’
  - b. exigir ‘demand’ > **exigencia** ‘demand’, coincidir ‘coincide’ > **coincidencia** ‘coincidence’,

- c. absorber ‘absorb’ > **absorbencia** ‘absorbence’,  
adherir ‘adhere’ > **adherencia** ‘adherence’
- d. vagar ‘wander’ > **vagancia** ‘idleness’,  
mangar ‘steal’ > **mangancia** ‘corruption’
- e. comandar ‘command’ > **comandancia** ‘rank of major’,  
resonar ‘echo’ > **resonancia** ‘resonance’
- f. descender ‘descend’ > **descendencia** ‘offspring’,  
ganar ‘earn’ > **ganancia** ‘profit’
- g. comparecer ‘appear-in-court’ > **comparecencia** ‘appearance-in-court’,  
transferir ‘transfer’ > **transferencia** ‘transfer’

All these bases pass at least some standard tests on eventivity (as opposed to stativity): for instance, they can be combined with an event-internal locative (6) and they can be taken as infinitival complements by perception verbs (7); both properties are rejected by purely stative verbs (8).

- (6) Juan observaba la foto en su oficina.  
Juan observed the photo in his office
- (7) Lo vi vigilar la oficina.  
him saw watch the office  
'I saw him watch over the office'
- (8) a. \*Juan sabe inglés en su casa.  
Juan knows English in his house  
b. \*Lo vi saber inglés.  
him saw know English

Note that the apparent counterexamples in (5) are divided in seven classes; the reason for this classification will become clear when we will argue that, far from being real counterexamples, the existence of these formations is predicted by a theory where *-ncia* is a nominaliser that looks for a base that meets the strict subinterval property.

With the possible exception of (5g), in all the other cases, we will argue that the eventive base is either non-dynamic or has a non-dynamic reading. Verbs in (5a) are, in fact, Davidsonian states, while those in (5b) are eventive verbs with a state reading; eventive bases are interpreted as dispositional in (5c), as habituels in (5d) and as potentials in (5e); nominalisations in (5f) denote result objects.

### 3. What this tells us about the nature of *-ncia*

Leaving aside this last group, the results that we have obtained paint the following picture of *-ncia*:

- (9) The suffix *-ncia* always takes bases that are intransformational / non-dynamic  
Thus it is productive in the following cases:
  - a) to produce quality nominalisations, as qualities are non-dynamic
  - b) to derive nouns from both purely stative and D-stative verbs

When combined with a dynamic base, the base is interpreted in a non-dynamic sense, including dispositional, habituality and potentiality, among other possible meanings.

This solves Puzzles #1 and #2 of *-ncia*. Starting from #2, we have seen that when the base is dynamic, the interpretation assigned to the nominalisation is stativised in some way. For reasons of space, let us illustrate with the dispositional reading (eg., *absorbence*). Stativisation can be performed in two ways, structurally: first, it can be obtained ‘by absence’ if the event argument is not introduced in the structure. This would mean that the suffix *-ncia* is introduced below the structural position where such argument is introduced (10a). Second, stativity can be obtained ‘by presence’, namely by merging stativising operators (eg., modals) above the event variable; this implies that *-ncia* can be introduced relatively high in the structure, and embed a quite rich verbal structure (10b).

- (10) a. [NP -ncia [VP ]]  
       b. [NP -ncia [XP Op<sup>stative</sup> [vP <e> [VP ]]]]

We do not have at this moment any empirical argument for or against one of these accounts. For concreteness, however, we will assume (7b). (11) illustrates the structure with the dispositional *absorbencia*; we assume Lekakou’s (2005) ascriptional operator to convey the dispositional semantics: by virtue of its internal properties, an entity will absorb liquids if there are facilitating circumstances.

- (11) [NP -ncia [XP Op<sup>ascrip</sup> [vP <e> [VP -e- [absorb]]]]]

Thus, what is relevant for *-ncia* is that its complement is non-dynamic: that is, as we expect from locality, the information that is embedded below the head of its complement is not relevant to compute the non-dynamicity of the complement.

Interestingly, non-dynamicity is satisfied both by quality nominalisations, stative verbs and stative interpretations of eventive verbs, and at this point we move to Puzzle #1. Even though categorially adjectives expressing qualities and verbs expressing eventualities (dynamic and non-dynamic) are very different, from the perspective of boundedness and their satisfaction of the strict subinterval property, they are identical. The suffix *-ncia* produces both, not caring whether the result is a quality that pairs the nominalisation with an existing adjective or is a stative situation that is somehow related to a verb.

This is expected if there are indeed cross-categorial primitives defining notions like (un)boundedness and *-ncia* happens to be a suffix that is sensitive to these primitives rather than to the category label of its complement. From our examination of the facts, it follows that in order to describe this suffix correctly one must ignore whether the result is an eventuality or a quality, which are lexically-specific notions, and go directly to a lexically-neutral notion of ‘absence of change’. Assuming Piñón’s (1997) primitives, what *-ncia* requires in its complement is an object that lacks an initial or a final boundary, and contains simply a body – an extension without internal limits– which by default will be interpreted as absence of change.

Consequently, Puzzle #1 dissolves: *-ncia* produces both kinds of nominalisation because its selectional requisites are sensitive to the primitives used to build qualities and aspect, not to their lexically-specific instantiation as other kinds of elements. The suffix *-ncia* will combine with bases that do not involve any rate of change, and this includes roots interpreted as qualities (12a), stative verbs (12b) and eventive verbs under the scope of stativising operators (12c).

- (12) a. [NP -ncia [√elega]] ‘elegance’  
       b. [NP -ncia [VP e [√exist]]] ‘existence’  
       c. [NP -ncia [XP Op [vP [VP a [√vag]]]]] ‘idleness’

## References

- CANO, María de los Ángeles & Matías JAQUE. (2011). *On the aspectual interpretation of deverbal formations in Spanish: weak and strong patterns of suffixation*. Unpublished Manuscript, Universidad Autónoma de Madrid.
- FÁBREGAS, Antonio & Rafael MARÍN. (2012). The role of Aktionsart in deverbal nouns: state nominalizations across languages. *Journal of Linguistics* 48. 35-70.
- FÁBREGAS, Antonio. (2016). *Las nominalizaciones*. Madrid: Visor.
- MOURELATOS, Alexander P. D. (1978). Events, processes and states. *Linguistics and philosophy* 2. 415-434.
- PIÑÓN, Christopher. (1997). Achievements in an event semantics. In Andrew Lawson & E. Cho (eds.), *Proceedings of SLT VII*, 273-296. Ithaca (NY): CLC.

## [lo kaze aj-aj-aj]<sup>1</sup>: Haplology in Modern Hebrew plural marking

Noam FAUST,

*Université Paris 8 & CNRS SFL*

faustista@yahoo.com

In Modern Hebrew, nouns appear in two “states”: the Free State, which is the citation form, and the Construct State, which is the form of the noun when it is the head of a nominal compound  $N_{\text{head}} + N_{\text{modifier}}$ . As can be seen in (1), in both States the masculine singular does not carry any overt marking. Feminine singular nouns appear with one of several overt markers; this talk will only be concerned with the most common of these markers, namely *-a*. In the Construct State, this feminine singular marker appears with an additional [t]: *sus-at*. Masculine plural is exponed by a suffix *-im* in the Free State and by a suffix *-ej* in the Construct State. In contrast, feminine plural nouns carry a plural suffix *-ot* in both States. Note further that the feminine plural marker *-ot* is added directly to the stem, rather than to the singular form with the suffix *-at*.

(1)	<i>Singular</i>		<i>Plural</i>	
	<i>Free State</i>	<i>Construct State</i>	<i>Free State</i>	<i>Construct State</i>
<i>masculine</i>	<i>sus</i>	<i>sus ets</i>	<i>sus-im</i>	<i>sus-ej ets</i>
<i>feminine</i>	<i>sus-a</i>	<i>sus-at ets</i>	<i>sus-ot</i>	<i>sus-ot ets</i>

The goal of this talk is to account for two cases of asymmetry between masculine and feminine exponence. The first appears in (2) and concerns possessive suffixes: when a suffix *-ey-*, marking the plurality of the possessed, is added to a masculine noun, it is concatenated to the singular form. But when it is added to a feminine noun, it is concatenated to the plural base. The same plurality is thus marked twice on possessed feminine nouns (marking plural number only once is ungrammatical: \**sus-ot-xem* or \**sus-at-ey-xem*):

(2)	<i>possessed is singular</i>	<i>possessed is plural</i>
<i>masculine</i>	<i>sus-xem</i> ‘your <sub>[mpl]</sub> horse’	<i>sus-ey-xem</i> ‘your <sub>[mpl]</sub> horses’
<i>feminine</i>	<i>sus-at-xem</i> ‘your <sub>[mpl]</sub> mare’	<i>sus-ot-ey-xem</i> ‘your <sub>[mpl]</sub> mares’

The second case of double plural marking exhibits the same asymmetry. It is found in the formation of *new* dual nouns (Schwarzwald 2002). As shown in (3), the dual suffix *-áj-im* attaches to the masculine singular base, but to the feminine plural base.

(3)	<i>singular</i>	<i>dual</i>
<i>masculine</i>	<i>sus</i>	<i>sus-áj-im</i>
<i>feminine</i>	<i>sus-a</i>	<i>sus-ot-áj-im</i>

<sup>1</sup> NEG so [ajajaj], ‘not so good’

The talk presents an analysis of these cases based on haplology, i.e. the omission of one of two adjacent, similar exponents (for ample evidence for haplology, see Nevins (2012)). It is claimed that in both (2) and (3), a masculine plural marker is also expected to be present, but its realization is blocked because it is too similar to the following adjacent marker. Thus, [sus-ej-xem] and [sus-áj-im] are actually /sus-ej-ej-xem/ and /sus-ej-aj-im/ respectively. The first /ej/ is deleted because of haplology. In the feminine cases, haplology does not apply because the exponents are different /sus-ot-ej-xem/ and /sus-ot-aj-im/.

The analysis, conducted within Distributed Morphology (e.g. Embick (2010)), proceeds in two steps. First, a set of Vocabulary Items is proposed, formalizing the generalizations regarding number exponentence in (1). The two important VIs are presented in (4a,b). Masculine is assumed to be the absence of a gender feature. Accordingly, (4a) assumes that /im/ and /ej/ are two allomorphic realizations of a plural feature only, /im/ realizing this feature only at the right edge of the phonological word. (4b) designates /-ot/ as a portmanteau morpheme, realizing both gender and number. I follow Svenonius (2016) in assuming that such VIs take precedence over VIs that are specified for each feature separately (e.g. 4a, 4c).

- (4) a. [plural]                       $\Leftrightarrow$       /-im/                      / \_\_\_\_ ]PhonWord  
     /-ej/  
   b. [plural, gender]               $\Leftrightarrow$       /-ot/  
   c. [gender]                       $\Leftrightarrow$       /-at/

The second step in the analysis establishes the syntactic structures of the two constructions in (2,3). Beginning with the possessive construction, I argue that the plural possessive suffixes are contained entirely under *spec,nP*, as in (5). While this does establish a parallel with the Construct State forms such as [sus-ej ets], note the crucial difference: in [sus-ej ets], /ej/ realizes the feature [pl] on the *num* head, while in (5) it is part of the suffix. As a result, the feature [pl] on the *num* head will seek to be realized. According to (4a) above, the expected exponent is /ej/. However, in the masculine case in (5a) insertion of /ej/ creates haplogy, and so only the second marker survives. In the feminine case (5b), no haplogy comes about.

- |     |   |  |
|-----|---|--|
| (5) | a. <i>sus-ej-xem</i> 'your <sub>(pl)</sub> horses'  | b. <i>sus-ot-ej-xem</i> 'your <sub>(pl)</sub> mares'   |
|     | <pre> graph TD     xP[xP] --&gt; n1[n]     xP --&gt; numP1[numP]     numP1 --&gt; √sus1["√sus"]     numP1 --&gt; numPl1["num[pl]"]     numPl1 --&gt; ej1["ej"]     numPl1 --&gt; xem1["xem"]     ej1 -.-&gt; ejxem1["ej-xem"]     ejxem1 --- ejxem["ej-xem"]     ejxem --- n2[n]     ejxem --- √sus2["√sus"]   </pre> | <pre> graph TD     xP[xP] --&gt; n1[n]     xP --&gt; numP1[numP]     numP1 --&gt; √sus1["√sus"]     numP1 --&gt; numPl1["num[pl]"]     numPl1 --&gt; ej1["ej"]     numPl1 --&gt; xem1["xem"]     ej1 -.-&gt; ejxem1["ej-xem"]     ej1 -.-&gt; ejxem2["ej-xem"]     ejxem1 --- ejxem["ej-xem"]     ejxem1 --- n2[n]     ejxem1 --- √sus2["√sus"]   </pre> |

For opponents of Distributed Morphology, it shows that the analysis in no way depends on that theory's basic assumptions, and can be easily expressed in other theories.

Next, the talk considers two views of the suffix *-aj-* of new duals: i) a marker of dual inflection on *num* or ii) a realization of an additional, derivational nominal layer. Both of these views are shown to make wrong predictions. It cannot be a dual marker specifically because there is no dual inflection in MH; and while this fact can be explained through impoverishment, that explanation is both unconstrained and completely ad-hoc. Conversely, *-aj* cannot be an additional nominal layer because that predicts that all new duals with *-aj-im* should have the same gender, where in fact the gender of the base is preserved.

Instead, it is proposed that *-aj-im* is analogous to a *pluralia tantum* noun such as [nisu<sup>2</sup>-im] ‘marriage’, which in addition always occupies the modifier position of a construct state, namely spec, nP: [[<sub>num</sub>[<sub>p</sub>][[<sub>aj-im</sub>]<sub>xP</sub>[<sub>sus</sub>]<sub>n</sub>P]<sub>nP</sub>]<sub>numP</sub>]. Because of the dual meaning of *-aj*, the *num* head of the base noun must always bear a plural feature (in bold in the structure in the previous sentence).<sup>2</sup> This view gives the following linearizations: /sus-ej-aj-im/ and /sus-ot-ajim/. Only the former is a case of haplology, and the same solution is applied as in (5a), giving /susajim/ and /susotajim/.<sup>3</sup> It is then shown that this analysis of new duals not only predicts correctly with respect to agreement, but can also account for the absence of such duals from the N+N construction: \*[sus-ot-ej ets] ‘two wooden mares’.

To summarize, by formalizing the basic VIs of inflection and setting the syntactic structures of the two constructions, the analysis manages to motivate the two asymmetries under discussion through the use of the concept of haplology. The talk concludes with a call for morphologists to look for good reasons for absent exponents, rather than resort to unconstrained mechanisms like impoverishment.

## References

- BAT-EL, O. (2009). A gap in the feminine paradigm of Hebrew: A consequence of identity avoidance in the suffix domain. In Rice, C. and S. Blaho (eds.), *Modeling Ungrammaticality in Optimality Theory*. London: Equinox. 165-194
- EMBICK, D. (2010). *Localism versus Globalism in Morphology and Phonology*. Cambridge, Massachusetts : MIT Press.
- NEVINS, A. (2012). Haplological dissimilation at distinct stages of exponence. In Trommer, J. (ed.), *The Phonology and Morphology of Exponence*. Oxford: Oxford University Press. 84-116.
- SCHWARZWALD, O. R. (2002). *Studies in Hebrew morphology* vol. 1-4 (in Hebrew). Tel Aviv: Open University Press.
- SVENONIUS, P. (2016). Spans and Words, In Siddiqi, D. and H. Harley (eds.) *Morphological Metatheory*. Amsterdam: John Benjamins. 201-222.

<sup>2</sup> Just like in English one cannot say *a couple of \*week*.

<sup>3</sup> Two adjacent -Vj- suffixes have been shown to be problematic in Bat El (2009), too, regardless of V quality.

## Morphology through the hands and face: evaluative morphology in Italian Sign Language

Elena FORNASIERO,

*Università Ca' Foscari Venezia*

[elena.fornasiero@unive.it](mailto:elena.fornasiero@unive.it)

### Introduction

The issue of evaluative morphology (henceforth: EM) is a quite debated one, being discussed among linguists who try to understand (i) how EM is displayed among world languages, (ii) which kind of morphology (inflectional or derivational) it belongs to, and (iii) which lexical categories it applies to. In general, EM refers to those processes modifying the morphology of nouns in order to convey not only diminutive and augmentative features, but also emotional overtones of approval or endearment and of disapproval or pejoration (Bauer 1997). According to Jurafsky (1996), evaluative features represent the grammatical encoding of the notions of ‘small’ and ‘big’ and of the affective notions of ‘nice/lovely’ and ‘bad/ugly’. In her typological studies, Kortvelyéssy (2012b; 2015) refers to evaluative morphology as the expression of evaluation by means of morphology. She considers evaluative constructions as characterised by an evaluated base (an existing word stem) and an evaluative marker, which assigns new value to the standard of the word stem. Specifically, the new value expresses a semantic value comprised within the semantic primitives ‘good-bad’ and ‘small-big’, as stated by Grandi (2002), who argues that evaluation implies two different perspectives: one objective, represented by the semantic primitives ‘small’ and ‘big’, and one qualitative or subjective, represented by ‘good’ and ‘bad’. World languages vary with respect to the various word-formation processes through which they display evaluative features: affixation, reduplication, compounding or phonological modifications.

Despite being pervasive among sign languages (henceforth: SLs), the phenomenon of evaluative morphology is still quite unexplored, being marginally cited among their morphological properties. However, a deep analysis of the morphological processes adopted by SLs to convey evaluative features could be interesting in order to: (i) improve typological studies defining which languages display evaluative strategies and how; (ii) detect sign language-specific and universal processes shared by both signed and oral languages; and (iii) shed some light on the nature of EM. Sign languages show a marked preference for simultaneous morphological processes, consisting in the modification of the formation parameters of the manual sign (handshape, movement, place of articulation, orientation) to express agreement, number or aspectual information (for e.g. the manual sign for ‘eat’ in Italian Sign Language can be realised with a faster and repeated movement in order to convey the meaning ‘eating a lot’). Crucial for the realisation of simultaneous processes are non-manual markers (NMMs) and classifiers (CLs). NMMs (also called non-manual signs) are a distinct feature of SLs and comprehend facial expressions (raised or furrowed eyebrows, open or squinted eyes, inflated cheeks, mouth movements, eye-gaze), head and body movements, which are simultaneously articulated with the manual sign(s) to fulfil lexical, adverbial and syntactic functions. Thus, they are considered bound morphemes (Liddel 1980; Bickford,

Fraychineaud 2006; Sandler, Lillo-Martin 2006; Tomaszewski, Farris 2010). Classifiers, instead, are morphologically complex constructions, consisting of a handshape associated to a movement, which combine to express information about certain features of an entity: its visual-geometric characteristics (classifiers defining the external properties of an entity are called Size and Shape Specifiers (SASSs)), its abstract semantic category, its handling or manipulation (a. o. Supalla 1982; Corazza 1990; Zwitserlood 2012). They can realize sequential processes when occurring after the manual sign for nouns to convey specific characteristics of the entity. Schuit (2007) argues that SLs are agglutinative languages, since morphemes are easily segmented, even though they display both sequential and simultaneous processes. The relevance of taking sign languages into consideration in linguistic research is to provide a great opportunity to confirm whether the theoretical frameworks that have been mainly developed for spoken languages can also account for sign languages, considering that they are visual-gestural languages, conveyed through manual and non-manual signs and visually perceived.

## Goals

The present paper aims at (i) offering a preliminary description of the display of evaluative features in Italian Sign Language (LIS), (ii) analysing the morphological behaviour of the main linguistic elements involved, i.e. non-manual markers (NMMs) and size and shape classifiers (SASSs), and (iii) trying to find a position for LIS evaluative morphology within the inflectional-derivational continuum.

## The study

In order to collect a considerable amount of data, I have analysed a corpus of spontaneous data coming from 22 LIS fairy-tales, and elicited data produced by three LIS native signers involved in a picture-description task, narration and grammaticality judgements. The items created for the tests aimed at eliciting specific characteristics defining the referent quality, size and shape (for e.g. a small ruined house, a big sweet dog) in order to elicit both objective and subjective features. The starting point of the present study was the analysis developed by Petitta, Di Renzo, Chiari (2015) who detect three main morphological strategies to convey evaluative features in LIS: (i) *Manual sequential/simultaneous evaluation*: sequential evaluation consists in the articulation of the sign for the noun followed by specific SASSs defining its size and shape, both marked by specific NMMs; manual simultaneous evaluation, instead, refers to the modification of the phonological features of the manual sign (such as movement, handshape), whose articulation can be enlarged or reduced to convey the meaning ‘big’ or ‘small’. (ii) *Non-manual simultaneous evaluation*: evaluative features are conveyed by adding specific NMMs to the articulation of the manual sign; (iii) *reduplicative evaluation* (sequential and simultaneous), which consists in the partial or full reduplication of the manual sign, or a portion of it, with possible variation in manual patterns.

## Description of the data

Within the fairy-tales corpus, I have detected a total number of 154 elements containing diminutive, augmentative, endearment and pejorative features: 77 examples of manual sequential/simultaneous and non-manual simultaneous evaluation, 69 realised through the articulation of the sign for the noun followed by classifiers and adjectives such as ‘nice’, ‘lovely’, ‘bad’, ‘big’, ‘small’. Within elicited data, I detected 187 instances of evaluative strategies: 125 examples of manual sequential/simultaneous and non-manual simultaneous

evaluation, and 62 consisting in the production of the noun followed by classifiers and adjectives. For both groups, manual sequential/simultaneous and non-manual simultaneous evaluation were the preferred strategies, and they were mostly used to convey diminutive (examples 1, 7, 8) and augmentative (2, 5, 6) features. Endearment and pejorative features were mainly conveyed through the articulation of the adjectives mentioned above, however I detected some instances of non-manual simultaneous evaluation as well, which are reported in (3) and (4).

(1) to (4) show examples of non-manual simultaneous evaluation in which the manual signs for nouns are modified in their articulation (restricted for the diminutive as in (1), enlarged for the augmentative as in (2)) and through the articulation of specific non-manual markers for each evaluative feature (the occurrence of NMMs is signalled by a line spreading over the English glosses, and their function is indicated through abbreviations written above the lines).



(1) CHAIR  
‘a small chair’



(2) CHAIR  
‘a big chair’



(3) BEAR<sup>Pej</sup>  
‘a bad/angry bear’



(4) BEAR  
‘a sweet bear’

(5) to (8) are examples of manual sequential evaluation in which the size and shape classifiers following the sign for the nouns convey augmentative (5, 6) and diminutive (7, 8) features. As the examples shows, specific non-manual markers for each feature occur with the articulation of the classifiers.



(5) BOOK CLsize CLdepth  
Book big thick  
‘a very big book’



(6) CAR CLsize  
car long/big  
‘a very big car’



(7) BOOK CLsize  
Book small  
‘a small book’



(8) TABLE CLsize  
table small  
‘a little table’

From the analysis of both elicited and naturalistic data, it results that the preferred strategies adopted by signers to convey evaluative features are: (i) specific NMMs produced simultaneously to the manual sign for the noun; (ii) the articulation of size/shape classifiers following the sign for the noun, both marked by specific NMMs. SASSs were selected considering the size and shape of the entity, as can be seen in (5) vs. (7). Evaluative features are encoded differently in the data: (i) the diminutive feature is conveyed through a restricted articulation of the sign for the noun or the classifier defining size and shape and it is associated with NMMs consisting of squinted eyes and tensed mouth or furrowed eyebrows and lips/tongue protrusion. (ii) augmentative features are conveyed through an enlarged articulation of the sign for the noun or the size/shape classifier, associated with open eyes and mouth shaping O/A or furrowed eyebrows and inflated cheeks or teeth biting the inferior lip. (iii) endearment features are conveyed through specific NMMs consisting of relaxed eyebrows and lips protrusion. (iv) pejorative features are conveyed through furrowed eyebrows and tensed cheeks or tongue/lips protrusion.

## Discussion

As the examples show, non-manual markers play a crucial role in conveying evaluative features in LIS, by adding specific morpho-syntactic features to the manual signs, whose presence, or absence, changes the meaning of the manual sign. They are bound morphemes since they cannot occur alone, and they define properties of the noun with which they are articulated. NMMs can also occur with the classifier defining size and shape of the noun: in such cases, the meaning of the classifier would not be complete without the specific NMMs for the evaluative feature considered. The role of SASSs, instead, is quite controversial: even though they have been included among the strategies to convey evaluative features in LIS, and they actually define if an entity is smaller or bigger with respect to the standard value, it seems that their role is to convey additional information about the external characteristics of the entity involved, whose evaluative features could be conveyed just through the production of non-manual markers with the manual sign. A possible explanation could be that SASSs are employed only with entities of which size and shape could be further specified. In fact, NMMs alone can also change the meaning of manual signs referring to abstract entities, as in (9), an example of non-manual simultaneous evaluation as the ones in (1) to (4), in which the sign for ‘green’ becomes ‘greenish’ thanks to the addition of the NMMs dedicated to pejorative features.



(9) GREEN  
‘greenish’

However, an analysis of other sign languages of the world provided the same results, showing that the articulation of size and shape classifiers and non-manual markers is the preferred strategy to convey evaluative features (AmericanSL: Bickford, Fraychineaud 2006; Schnepf 2011; BritishSL: Sutton-Spence, Woll 1999; GermanSL: Pfau, Quer 2010; FrenchSL: Cuxac, Salandre 2007; PolishSL: Tomaszewski, Farris 2010; IsraeliSL: Fucks 2014; Sandler 2009;

AdamorobeSL: Nyst 2007). Thus, the nature and function of SASSs is an open issue which needs further research, taking into account that SASSs occurring in strictly evaluative descriptive contexts seem to be ordered among the other nominal modifiers, thus having a specific distribution within the noun phrase.

## Conclusions

From a typological perspective, the analysis of the data collected shows that: (i) LIS belongs to that group of languages displaying both augmentative and diminutive features (Kortvélyessy 2012b; 2015); (ii) it displays evaluative features through the articulation of both manual and non-manual morphemes modifying the morphology of nouns, and (iii) evaluative morphology in LIS belongs to the domain of derivational morphology. The present study provides a detailed description of evaluative strategies in LIS, showing that each evaluative feature is associated to specific non-manual markers and modifications in the articulation of manual signs, together with the occurrence of specific classifiers. The phenomenon of evaluative morphology allows, once again, to account for the richness of sign languages, whose possibility to convey many information simultaneously increases the knowledge about the functioning of the language faculty.

## References

- BAUER, L. (1997). "Evaluative morphology: in search of universals". *Studies in Language*, 21: 3, 533–75.
- CORAZZA, S. (1990). "The Morphology of Classifier Handshapes in Italian Sign Language (LIS)". In: C. Lucas (a cura di) *Sign Language Research: Theoretical Issues*, 71-82. Washington D.C.: Gallaudet University Press.
- GRANDI, N. (2002). *Morfologie in contatto: le costruzioni valutative nelle lingue del Mediterraneo*. Milan: Franco Angeli.
- KÖRTVÉLYESSY, L. (2012b). *Evaluative Morphology from a Cross-Linguistic Perspective*. Budapest, ELTE: Habilitationsschrift.
- KÖRTVÉLYESSY, L. (2015). *Evaluative Morphology from a Cross-Linguistic Perspective (1)*. Cambridge Scholars Publishing.
- PETITTA, G., DI RENZO, A., CHIARI, I. (2015). "Evaluative morphology in sign languages". In: Grandi, N., Kortveleyssey, L. *Edinburgh Handbook of Evaluative Morphology*, 155-169, Edinburgh University Press.
- SANDLER, W., LILLO-MARTIN, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- SCHUIT, J. (2007). *The typological classification of sign language morphology*. Master's Thesis, Universiteit van Amsterdam.
- SUTTON-SPENCE, R., WOLL, B. (1999). *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press.

## Deverbal nominalizations denoting places

Bernard FRADIN,

Université Paris Diderot & CNRS

[bernard.fradin@linguist.univ-paris-diderot.fr](mailto:bernard.fradin@linguist.univ-paris-diderot.fr)

**1.** Nouns derived from verbs, which are not *nomina agentis*, have generally an eventive interpretation e.g. eng *arrival* ‘action of arriving’ ← *arrive*. Some of them may also denote the place where an event takes place e.g. Italian *entrata* ‘entrance’ ← *entrare* ‘to enter’. Apresjan (1974) analyses this phenomenon as a type of regular meaning extension involving a process akin to metonymy. Even though metonymic extension is a well-attested and widespread phenomenon, the presentation argues that the reason why some nominalizations (NZNs) can have a locative meaning is tied to the fact that the semantics of the verb they derive from crucially includes a spatial relationship: in the absence of the latter, no locative interpretation occurs. It also claims that the base verb is generally telic, a condition not imposed by suffixation in *-oir*, which is a process partly specialized in deriving spatial nouns in French e.g. *lavoir* ‘wash house’ ← *laver* ‘to wash’ (Namer & Villoing, 2008). The discussion focuses on French nominalizations in *-age*, *-ment* and, more rarely *-ion*, and uses data collected come in Frantext, FrWaC and the Web.

**2.** Following (Melloni 2011: 116-120), it is argued that the ‘locative nominalizations’ dealt with in the present paper have to be distinguished from NZNs derived from verbs of ‘degree achievement’ such as ita *prolungamento* ‘extension’ ← *prolungare* ‘to extend’. The semantically these nouns are interpreted as ‘means’ (or ‘effector’, in the spirit of Van Valin and Lapolla (1997)). A distinctive feature of ‘means’ is their capacity to head a NP that functions as subject of a sentence whose predicate is the base verb e.g. ita *un prolungamento prolunga* (*Y*) / eng *an extension extends* (*Y*), *an extension is what extends* (*Y*). Nominalizations with a locative interpretation do not share this property.

I assume that the semantics of verbs such as GARDER ‘to park’, RANGER<sup>2</sup> ‘to clean up, to store’, etc. obligatorily specify a spatial relationship. A rough semantic representation of GARDER would involve the following elements (i): **park**(x,y,e<sub>1</sub>) ∧ CAUSE(e<sub>1</sub>,e<sub>2</sub>) ∧ LOC(y, INESS(z)) e<sub>2</sub>) ∧ CAUSE(e<sub>2</sub>,e<sub>3</sub>) ∧ ¬LOC(y, SUPERS(w),e<sub>3</sub>) ∧ **way**(w) in words ‘the result of the parking event e<sub>1</sub> is in the fact e<sub>2</sub> of y being within z and of not being on the way (where traffic moves)’ (INESS = inessif, SUPERS = superessif). The important thing here is that this representation makes available variable z, which corresponds to the Ground (Talmy 2000)(or landmark (Langacker 1987)) in the spatial relationship. When the variable selected in the nominalization process is *e*, the derived noun denotes an event; when this variable corresponds to the ground, it denotes a place. Hence GARAGE with the locative meaning has roughly interpretation (ii), which is a generalization of (i): (ii)  $\lambda z. \exists xy \exists e_1 e_2 [\text{park}(x,y,e_1) \wedge \text{CAUSE}(e_1,e_2) \wedge \text{LOC}(y, \text{INESS}(z)) e_2] \wedge \text{way}(w)$ , a garage is a place whose function (or characteristic feature) is to have y parked inside (y = vehicle). To tell the truth, representation (ii) does not account for the functional part of the meaning, which presumably stems from the nominalization process itself (cf. *couper* ‘an x whose function is to cut (y), x = object’). The same story holds true for the original meaning assigned to the converted noun GARE ‘station’ in the vocabulary of railways. The locative interpretation arises because advantage is taken of

the possibilities built-in the semantics of the verbal lexeme. Resorting to metonymy is an unnecessary move. In a similar way, the interpretation of RANGEMENT ‘place where things can be stored in a neat way’ e.g. *Les rangements permettant un minimum d'encombrement sont plus que bienvenus dans ce type de configuration* ‘storage spaces that are compact are very welcome in this type of configuration’ (Web, 5.2017). Among the nominalizations exhibiting a similar behavior, we find ABRI ‘shelter’, ENTRÉE ‘entrance’, ESSART ‘cleared ground’, APPUI ‘support’ and many others.

In representation (i), the verb is telic as shown by the fact that the spatial relationship is satisfied at the end of the parking process (final location). It seems that most verbal lexemes expressing a spatial relationship that give rise to nominalizations denoting a place are telic. However, PASSAGE ‘passage’ and PROMENADE ‘promenade’ show that this requirement is not mandatory since the verbal lexeme they derive from is atelic. SURVOLER ‘hover’, which literally means ‘X voler (sur | au dessus de) Y’ = ‘fly over Y’, and thereby describes an activity. The fact that the nominalizations that derive from it, SURVOL and SURVOLEMENT, do not denote a place but a flying event could not therefore be attributed to telicity. A possible explanation, which will be discussed in the presentation, is that the place denoted by the derived noun has to be spatially bound, insofar as only bound spaces can easily be associated with a function. This requirement is met whenever the verb involves a final location (GARAGE, ABRI), or a place identified by an activity which is limited to a portion of space (PROMENADE, PASSAGE). It is not met in the case of SURVOL.

3. Not all telic verbs incorporating a spatial relationship are suited however. Those derived from nouns denoting a place (Huyghe 2009: 77-80) e.g. APPONTER ‘land on an aircraft carrier’ ← PONT ‘deck’ do not seem to be appropriate, as shown by nominalization APPONTAGE, whose interpretation can only be ‘action of landing on the deck of an aircraft carrier’ and never ‘place (on an aircraft carrier) where planes land’. Following this pattern, we find ATTERRIR ‘to land’ ~ ATTERRISSAGE ‘landing’, PERCHER ‘to perch’ ~ PERCHAGE ‘action of perching’, etc. Finally, spatial telic verbs heading a construction where the Ground argument is also the Patient do not do the job either. For instance, GARNISSAGE ‘fill, lining’ ← GARNIR ‘to fill, to garnish’, besides the ‘means interpretation’ i.e. the stuff that fills, has only an eventive interpretation because its direct argument (variable y) corresponds to the ground e.g. *nous garnissons les fauteuils de matériaux naturels* ‘we fill armchairs with natural materials’.

## References

- APRESJAN Juri. 1974. "Regular Polysemy". *Linguistics* 142 (1):5-32.
- HUYGHE Richard. 2009. *Les noms généraux d'espace en français. Enquête linguistique sur les noms de lieu*. Bruxelles: De Boeck / Duculot.
- LANGACKER Ronald W. 1987. *Foundations of Cognitive Grammar. Theoretical Prerequisites*. Vol. 1. Stanford: Stanford University Press.
- MELLONI Chiara. 2011. *Event and Result Nominals*. Bern: Peter Lang.
- NAMER Fiammetta & Florence VILLOING. 2008. "Interpréter les noms déverbaux: quelle relation avec la structure argumentale du verbe base? Le cas des noms en -oir". In *Congrès Mondial de Linguistique Française - CMLF'08*, Durand J., B. Habert & B. Laks (eds). 1539-1557. Paris: ILF - EDP Sciences.
- TALMY Leonard. 2000. *Toward a Cognitive Semantics*. 2 vols. Vol. 1. *Concept Structuring Systems*. Cambridge / London: The MIT Press.
- VAN VALIN Robert D. Jr. & Randy J. LAPOLLA. 1997. *Syntax. Structure, meaning and function*. Cambridge: Cambridge University Press.

## English *ie*-hypocoristics: A DM analysis

**Edward Göbbel,**

*University of Wuppertal*

goebbel@uni-wuppertal.de

In this paper it is argued that there are two different strategies of forming *ie*-hypocoristics in English (*comfy*, *fattie*), one resorting to regular word-formation processes and a second, essentially syntactic one, operating on phrasal constituents.

Due to phonological restrictions on their formation, *ie*-hypocoristics have during the last decades been mainly discussed within the framework of Prosodic Morphology. For example, McCarthy & Prince (1998) assume that IE is a Class II suffix attached to a word tailored down to the size of a monosyllabic foot. On the other hand, on the basis of crosslinguistic data hypocoristic suffixes have been argued to belong to a separate class of evaluative affixes, with individual properties, which sets them apart from derivational and inflectional affixes (Scalise 1986).

The paper examines the morphological properties of *ie*-suffixation in order to (i) establish the nature of the base to which IE is attached (i.e., stem, word or phrase) and (ii) provide an analysis of the internal structure of these words.

It will be argued that there is no evidence that *ie*-suffixation is stem-based. Secondly, only hypocoristics based on common nouns, compounds and a small number of adjectives (e.g., *daffy* < daffodil, *nightie* < nightdress, *comfy* < comfortable) undergo regular word-formation processes. As such, they can be the input to compounding as well as to derivational and inflectional processes, as shown in (1).

- (1) Root + Class I suffix + IE + Class II / inflectional suffix
  - a. lun+atic+IE+ly (loonily), leftwing+IE+ness (leftiness)
  - b. comfort+able+IE+er (comfier)
  - c. left-hander+IE + friendly (leftie friendly)

However, the vast majority of *ie*-hypocoristics are based on noun phrases. In this case, IE is a phrasal affix, cliticising onto a modifier of the noun, which has been deleted. As such, they exhibit a phrasal syntax and semantics. The analysis is supported by the following facts:

- (i) only modifiers of nouns form the base of the affix, the modifiers themselves are, with a few exceptions, always nontruncated monosyllabic adjectives (2a). If truncation is possible at all, it is an (adjectival) affix that is deleted (2b).
- (ii) the interpretation of the hypocoristic cannot be derived from the adjectival base (2c)
- (iii) they give rise to irregular compounds, by allowing inflectional affixes in modifier position (2d)
- (iii) IE can be sandwiched between two inflectional affixes (3)

- (2) a. cutie < cute woman, oldie, newie, fattie, baddie, greenie, etc.  
b. talkie > talking picture  
c. shorty < short person, drink, story, film, article, etc.  
d. a shorties film festival

(3) Root + inflectional suffix + IE + inflectional suffix  
best + ie + s (Pl) (< best friend)

The structural analysis is couched within the framework of Distributed Morphology. It will be argued that IE is a head in the extension of a noun or adjective, the suffixed form being derived by head movement from structures like (4a). The phrasal hypocoristics are derived by ellipsis of the noun and cliticisation of IE onto the adjective (4b). Their structure includes at least NumP, even in root compounds. All types of ellipsis in English are licensed by functional heads (T, D, v) and the analysis here provides the missing type, licensed by the nominal suffix IE, an instance of small *n*.

- (4) a. [<sub>aP</sub> IE [<sub>a</sub> -able √COMFORT ]] (comfy)  
 b. [NumP [plural] [<sub>nP</sub> short [<sub>nP</sub> IE film ]]] (shorties)

## References

- MCCARTHY, J.J. & A. PRINCE. (1998). Prosodic Morphology. In: Spencer, A., Zwicky, A.M. (eds.), *The Handbook of Morphology*. Wiley-Blackwell, pp. 283-305.

SCALISE, S. (1986). *Generative Morphology*. Walter de Gruyte.

## Morphophonology and paradigms in the Démonette derivational database

**Nabil HATHOUT,**

*UMR 5263 CLLE-ERSS, CNRS & Univ de Toulouse-Jean Jaurès, France*  
*nabil.hathout@univ-tlse2.fr*

**Stéphanie LIGNON,**

*UMR 7118 ATILF, CNRS & Univ de Lorraine, France*  
*stephanie.lignon@univ-lorraine.fr*

**Fiammetta NAMER,**

*UMR 7118 ATILF, CNRS & Univ de Lorraine, France*  
*fiammetta.namer@univ-lorraine.fr*

### **Introduction**

Démonette is a morphological derivational database (MDB) of French characterized by an original relational structure. It does not describe the properties of individual morphological complex words but the properties of derivational relations between pairs of lexemes such as LANCER ‘launch’<sub>V</sub> → LANCEUR ‘launcher’ or LANCEUR → LANCEMENT ‘launch’<sub>N</sub>. In Démonette, properties are assigned to lexemes by the morphological relations that connect them. This conception of the morphological lexicon directly determines the structure of the MDB. It makes it redundant because the relations are bi-directional: for instance, the MDB contains a relation LANCER → LANCEUR (a launcher is one that launches) and a relation LANCEUR → LANCER (launching is what a launcher does); each lexeme is involved in many relations and it is assigned properties by each of them; the network includes indirect relations (LANCEUR → LANCEMENT (a launch is the act performed by a launcher)) in addition to the classical direct ones (LANCER → LANCEMENT (to launch is to perform the act of launching)). In Démonette, derivational relations define derivational families (connected graphs of derivational relations), which in turn are organized into paradigms (stacks of derivational families). Previous work on the resource mainly focused on the morphosemantic content of the database. We here address a different question, the morphophonological aspects of the resource, and show that the relational structure of Démonette provides an original way to represent these properties and to describe various morphophonological regularities.

### **Morphological derivational databases**

Basically, derivational morphology is lexical and a large part of the information that determines the analysis of complex lexemes is memorized. Many efforts have recently been made on improving morphological analysis based on large corpora (Cotterrel & Schütze, 2017, Lazaridou et al., 2013) but it is not the same for the storage of morphological information and its harmonization. Yet, hoarding this knowledge is essential for research in description morphology, lexicology, psycholinguistics, language teaching, etc.

Psycholinguists have created the first MDBs in order to build experimental material. The best known is CELEX (Baayen et al., 1995) whose first version dates back to the 1990's. It provides a wide range of phonological, morpho-syntactic, inflectional and derivational information for English, German and Dutch and is still a reference that has no real equivalent even if its coverage is reduced in view of the corpus we have today. Other resources have been created for English, such as the CatVar lexicon of derivational families (Habash & Dorr, 2003), designed primarily for NLP. However, this BDM cannot be used directly by the morphologists because the CatVar families have not been validated manually. A recent similar resource has been created for German: DerivBase (Zeller et al., 2013) has also been built automatically from corpus using distributional semantics techniques. For Italian, let us mention DerivaTario (Talamo et al. 2016), a derivational dictionary whose annotations are based on strong assumptions about allomorphy and suppletion. For instance, BELLICOSE 'bellicose' is analyzed as derived from GUERRA 'war'. For French, the only similar resource is Démonette whose main characteristics are presented in the following section.

## Démonette

The Démonette MDB (Hathout & Namer 2014) aims precisely to fill the void that exists for French in terms of reliable morphological resources with sufficient coverage. It is characterized by an original structure and designed to accommodate morphological descriptions resulting from research work such as theses and TAL resources such as VerbAction (Hathout & Tanguy 2002). Therefore, the list of fields used to describe relationships is not closed and can be expanded as needed. Among the fields present in the current version of the database, the most original ones are probably those used for the semantic description which include morphosemantic types such as "feminine agent name", concrete glosses that define the meaning of the second element of the relation with respect to the meaning of the first one as 'she who walks' for MARCHEUSE 'walker'<sub>F</sub>, but also abstract glosses which generalize the concrete ones and allow them to be inserted in derivational series such as the series of names of feminine agents which contains all the nouns whose abstract gloss is 'she who @' where @ represents the meaning of an action verb.

## The morphophonological descriptions in Démonette

In this talk, we present how Démonette entries are supplemented by morphophonological descriptions: phonological representations of the connected lexemes, the properties of their radicals, their exponents, the regularities they present and the variations that affect them. These descriptions are associated with the other properties of derivational relations, in particular the morphological and morphosemantic ones. Indeed, we extend to the morphophonological level the approach adopted for morphosemantic properties: distinction of concrete and abstract representations and attribution by the derivational relations of the morphophonological properties of the lexemes they connect.

Morphophonology is both simpler and more complex to describe than morphosemantics. On one hand, it is simpler because IPA transcripts can be used to represent the phonological content of the words. On the other hand, it is more complex because lexemes have no formal properties of their own. Lexemes are indeed realized by their inflected forms and only these have phonological properties, which include inflectional stems and exponents (Baerman & Corbett, 2009).

In Démonette, we have chosen to describe only the morphophonological properties relevant to word-formation (Roché, 2010, Plénat, 2009). We therefore describe only the stems and exponents involved in the derivational relations, which we calculate from the inflected forms

present in the GLÀFF lexicon (Sajous et al., 2013), completed if needed with data coming from Lexique3 (New, 2006). For nouns, we only consider the representation of the singular (FILLE ‘girl’ → FILLETTE ‘little girl’). For adjectives, the relevant stems are the ones of the masculine (BEAU ‘beautiful’ → BEAUTÉ ‘beauty’) and of the feminine (GRAND ‘big’ → GRANDEUR ‘size’) (Bonami & Boyé, 2005). For verbs, 6 stems are represented: present indicative 3SG, 1PL and 3PL, present participle, masculine and feminine past participles (Bonami & Boyé, 2003).

Tables 1 and 2 illustrate the fields of Demonette describing the morphophonological properties of the derivational relations and of the lexemes they connect. Lexemes M1 and M2 are listed with transcriptions of the forms we just presented and of their stems (Rad1, Rad2). The stems depend on the derivational relation (see Tab2): for a direct relation (for example, if M1 is the base of M2 as in line 1), Rad1 is one of the inflectional stems of M1 and Rad2 is what remains in the form of M2 once the exponent has been removed; if M1 and M2 are in an indirect relation (as in line 2), their stems are the sequences obtained when the derivational exponents are removed from the forms of the two lexemes. Tab1 shows that the description of M1 and M2 also includes stem size (number of syllables), and the properties of stems last syllable: onset, nucleus and coda.

			M1							M2						
	M1	M2	Rad1	Size1	LastOns1	LastNuc1	LastCod1	Suf1	Rad2	Size2	LastOns2	LastNuc2	LastCod2	Suf2		
1	boire ‘to drink’	buveur ‘drinker’	byv	1	b	y	v	--	byv	1	b	y	v	œr		
2	admirateur ‘admirer’	admiratif ‘admiring’	admirat	3	r	a	t	œr	admirat	3	r	a	t	if		

Table 1

In Tab2, the morphophonological relation (*Concr.Phon*) between M1 and M2 involves the Rad<sub>i</sub> sequences, concatenated with Suf<sub>i</sub> if necessary. This relation normally brings out (field Z) a sequence common to the two stems, which (i) can be identical to the two radicals (as in line 1), (ii) can correspond to a part of each of the radicals (as in line 2) or (iii) be void (as in line 6). When Rad1 and Rad2 share a common string, their difference is presented as an alternation (*Altern* field). The *Phon.Rule* field completes this description by indicating, when relevant, the involved phonological rule (sibilantization, sound, assibilation, insertion, etc.). When Z exists, it is used to form an abstract representation in *Abstr.Phon* that generalizes *Concr.Phon* and highlights the morphophonological organization of the lexicon, especially in terms of stem (and possibly exponent) variation. This abstract schema can also be used as label for morphophonological relations.

Alternations define morphophonological classes of lexemes and derivational sub-families: for instance, (COMPOSER ‘to compose’, COMPOSITEUR ‘composer<sub>M</sub>’, COMPOSITRICE ‘composer<sub>F</sub>’, COMPOSITION ‘composition’) and (INHIBER ‘inhibit’, INHIBITEUR ‘inhibitor<sub>M</sub>’, INHIBITRICE ‘inhibitor<sub>F</sub>’, INHIBITION ‘inhibition’) display the same set of stem variations that can be represented by a list of pairs such as Z ↔ Zit (INHIBER ↔ INHIBITEUR), Z ↔ Zis (INHIBER ↔ INHIBITION), Z ↔ Z (INHIBITEUR ↔ INHIBITRICE), Zt ↔ Zs (INHIBITEUR ↔ INHIBITION). The -eur, -rice and -ion exponents show no variations in the lexemes of the two sub-families. Notice that having indirect morphological relations (in addition to the direct ones) in the database allows to highlight regularities such as between EXTINCTEUR ‘(fire) extinguisher’ and

EXTINCTION ‘extinguishment’ that are not brought out at the level of each individual direct relations ÉTEINDRE ‘extinguish’ → EXTINCTEUR and ÉTEINDRE → EXTINCTION.

This representation allows us to distribute Démonette’s entries into four types of morphophonological families, according to the identity or variation of their stems and the nature of the formal relation that exists between them. We can distinguish, on the basis of the values of the *Altern* and *Abstr.Phon* fields: (i) entries for which there is no stem variation (lines 1, 2), (ii) entries with a stem variation that can be described by a phonological rule (lines 4, 5, 7), (iii) entries with a stem variation represented by an alternation that cannot be described phonologically (line 3), (iv) entries with no common sequence (lines 6, 8).

Filling in the morphophonological fields of the database entries is work in progress. It has been completed for some of Démonette entries. The main tasks that have been carried out to calculate the contents of these fields consist of: producing M1 and M1 stem spaces; re-syllabizing stems; comparing M1 and M2 stem spaces to identify Rad1 and Rad2; determining the alternations between Rad1 and Rad2; constructing concrete and abstract variations from these alternations; attributing the variations to morphophonological rules. Some of these operations are fully automatic such as the acquisition of stem spaces or the identification of common stems; others need some manual verification such as the identification of the relevant phonological rule.

	M1	M2	Rad M1	Suf M1	Rad M2	Suf M2	Z	Altern	Concr Phon	Abstr Phon	Phon Rule
1	boire	buveur	byv	--	byv	œr	byv	=	byv / byvœr	Z / Zœr	=
2	admirateur	admiratif	admirat	œr	admirat	if	admira t	=	admiratœr / admiratif	Zœr / Zif	=
3	admirer	admirateur	admir	--	admirat	œr	admir	/at	admir / admiratœr	Z/ Zœr	NONE
4	admirateur	admiration	admirat	œr	admiras	jō	admira	t/s	admiratœr / admirasjō	Zœr/ Zsjō	+sib
5	extincteur	extinction	ekstēkt	œr	ekstēks	jō	ekstēk	t/s	ekstēkœr / ekstēksjō	Zœr/ Zsjō	+sib
6	éteindre	extincteur	etē	--	ekstēkt	œr	e	--	etē / ekstēkœr	etē / ekstēkœr	NONE
7	aliment ‘food <sub>N</sub> ’	alimen- taire ‘food <sub>Adj</sub> ’	alimā	--	alimāt	er	alimā	/t	alimā / alimäter	Z/ Zter	+C
8	cheval ‘horse <sub>N</sub> ’	hippique ‘horse <sub>Adj</sub> ’	ſəval	--	ip	ik	--	--	ſəval / ipik	ſəval / ipik	NONE

Table 2

## Conclusion

With this organization, the network of morphological, morphosemantic and morphophonological relations that make up Démonette is able to capture paradigmatic patterns at several levels. On the morphophonological level, having concrete and abstract descriptions allows for observations that depend on the perspective chosen. At concrete level, the pairs EXTINCTEUR ‘extinguisher’ ↔ EXTINCTION ‘extinguishment’, ADMIRATEUR ‘admirer<sub>M</sub>’ ↔ ADMIRATION ‘admiration’ et PRÉDATEUR ‘predator’ ↔ PRÉDATION ‘predation’

involve the same mechanisms; at abstract level, (ADMIRER, ADMIRATEUR, ADMIRATION) and (ÉTEINDRE, EXTINCEUR, EXTINCTION) will be grouped into distinct series. When crossing the morphophonological and morphological plans, (PRÉDATEUR, PRÉDATION) is identified as an instance of the sub-paradigm of a larger paradigm, which, for example, includes (ADMIRATEUR, ADMIRATION, ADMIRER). Finally, comparing the morphosemantic and the morphophonological descriptions allows us to include (ADMIRER, ADMIRATEUR, ADMIRATION) and (CONSPIRER ‘to conspire’, CONSPIRATEUR ‘conspirateur’, CONSPIRATION ‘conspiracy’) in the same morphophonological paradigm but to place each family in a different morphosemantic paradigm. Conversely, (ENSEIGNER ‘to teach’, ENSEIGNANT ‘teacher’, ENSEIGNEMENT ‘teaching’) belongs to the same morphosemantic paradigm as (CONSPIRER, CONSPIRATEUR, CONSPIRATION), but both differ in terms of morphophonological paradigms.

## References

- BAAYEN, R. H., PIEPENBROCK, R. & GULIKERS, L. (1995). The CELEX lexical database. *Linguistic Data Consortium, Philadelphia*.
- BAERMAN, M. & CORBETT, G. (2009). “A typology of inflectional exponence: stems, affixes and what lies between”. Communication au *7th Mediterranean Morphology Meeting*. Nicosie, 10-13 septembre 2009.
- BONAMI, O. & BOYE, G. (2003). « Supplétion et classes flexionales dans la conjugaison du français ». *Langages* 152 : 102-126.
- BONAMI, O. & Boyé, G. (2005). « Construire le paradigme d'un adjectif ». *Recherches Linguistiques de Vincennes* 34 : 77-98.
- COTTERELL, R. & SCHÜTZE H. (2017). Joint Semantic Synthesis and Morphological Analysis of the Derived Word. *Transactions of the Association for Computational Linguistics*.
- HABASH, N. & DORR, B. (2003). A categorial variation database for English. In *Proceedings of NAACL 2003* pp. 17-23.
- HATHOUT, N. & NAMER, F. (2014). Démonette, a French derivational morpho-semantic network. LiLT, 11.
- TANGUY, L. & HATHOUT, N. (2002). Webaffix: un outil d'acquisition morphologique dérivationnelle à partir du Web. In *9e Conférence TALN*.
- LAZARIDOU A., MARELLI M., ZAMPARELLI R. & BARONI M. (2013). Compositional-ly derived representations of morphologically complex words in distributional semantics. In *Proceedings of ACL 2013*, 1517–1526.
- NEW, B. (2006). Lexique 3 : Une nouvelle base de données lexicales. In *Proceedings of TALN*, 892-900.
- PLÉNAT, M. (2009). « Le conditionnement de l'allomorphie radicale en français ». *Mémoires de la Société de Linguistique de Paris*, Nouvelle série, n° 17:119-140.
- ROCHE, M. (2010). « Base, thème, radical ». *Recherches linguistiques de Vincennes*, 39,(1), 95-134.
- SAJOUS, F., HATHOUT, N. & CALDERONE, B. (2013). Gläff, un gros lexique à tout faire du français. In *Actes de la 20e conférence TALN* (pp. 285-298).
- TALAMO, L., CELATA, C. & BERTINETTO, P. M. (2016). DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure*, 9(1), 72-102.
- ZELLER, B. D., SNAJDER, J. & PADÓ, S. (2013). DErivBase: Inducing and Evaluating a Derivational Morphology Resource for German. In *ACL 2013* pp. 1201-1211.

## Form-form relations in the paradigm

**Borja HERCE,**

*University of the Basque Country & University of Surrey*  
**borja.herce@ehu.eus**

It is well known that elements of form (whether they belong to affixal inflectional material or to stems) are often distributionally aligned with morphosyntactic features. For example, every element in Table 1 (e.g. m-, -on, -ij) correlates perfectly with some morphosyntactic value.

	SG	DU	PL
1	mon	muäna	mij
2	ton	tuäna	tij
3	son	suäna	sij

**Table 1** — Skolt Saami personal pronouns, nominative forms (Feist 2011:251)

When this happens, morphologists have usually little trouble in arguing that those particular pieces of form express or realize a certain morphosyntactic property. Realizational models, in fact, start with the assumption that this is exactly how morphology works by default. If inflectional morphology is there in order to express morphosyntactic features, correlation of form and meaning is only to be expected. Cases like the Skolt Saami pronouns suggest that morphosyntactic and semantic features must be able to serve as a sort of mould that shapes the distribution of formatives. There is, however, a footnote to this empirical finding which is not frequently discussed or granted theoretical import, which is that formal elements themselves can do exactly the same.

Formal elements do exist which are not coextensive with any particular morpho-syntactic or semantic feature. These elements of pure form (so-called 'morphemes', Aronoff 1994) have been the object of analysis of many morphologists (most prominently Martin Maiden) and a topic for debate in recent years (e.g. Luís & Bermúdez-Otero 2016) but continue to be controversial to this day both theoretically (how should they inform our models?) and empirically (how should we identify them in a language?). This remains, therefore, one of the most important tasks for research in morphology. Obtaining a clear picture of the variation which can be found within and around morphomicty (see the framework of canonical typology, Corbett 2005) and arriving at quantitative measures or correlates of the phenomenon will be one of my most important goals (see, by way of initial approaches, the measure of morphosyntactic coherence introduced by Esher 2014 or the concept of false positives and false negatives as raised by Trommer & Bank 2017).

Concrete elements of form ('meromorphemes' under Round's (2013) terminology) often have the same unmotivated paradigmatic distribution. In many cases, this may be just due to the continuation of the historical *status quo* (e.g. rhizotomy and a certain vowel alternant in the stem, which correlate in many Romance languages because the former was a source for the

latter). However, in other cases, the distribution of one formal element (e.g. augments like -isc- or suppletive stems, for example in Italian) has been analogically rearranged to match the distribution of another formal element, whether in the same or in another comparable paradigm.

This shows that, the same as morphosyntactic or semantic features, formal elements can also act as 'attractors' of other formal elements. Two affixes or two stem alternants with similar distributions will strive to become still more similar in distributional terms. Distributional similarity, therefore, tends to generate more similarity. In addition, formal similarity also generates more similarity (see Burzio's 2001:664 'gradient attraction'). Stem affinities may give rise to greater affixal similarity or even identity (Table 2) and conversely affixal affinity may sometimes promote greater stem similarity or identity (Table 3).

	1SG.PRES	2PL.IMP(I)	INF	INF	GER(II)	3PL.PAST
'know'	sé	sabed > saber	saber	saber	sabiendo > supiendo	supieron
'put'	pongo	poned > poner	poner	poner	poniendo > pusiendo	pusieron
'tell'	digo	decid > decir	decir	decir	diciendo > dijendo	dijeron
'want'	quiero	quered > querer	querer	querer	queriendo > quisiendo	quisieron
'make'	hago	haced > hacer	hacer	hacer	haciendo > diciendo	hicieron

**Table 2** — Spanish analogical change (I)

**Table 3** — Spanish analogical change (II)

The stem identity of the 2PL imperative and the infinitive, shown in Table 2, facilitates the analogical replacement of etymological *sabe-d* by the innovative *sabe-r* in substandard varieties of Spanish. Conversely, the affixal similarity (stressed -ie-) of the gerund and many of the cells with a so-called PYTA root (exemplified here by the 3PL past), has been at least partially responsible for the analogical developments displayed in Table 3 (e.g. *sab-ie-ndo* becomes *sup-ie-ndo* in some local Peninsular varieties, see Pato & O'Neill 2013).

Along with the development of clear measures for the morphomicity of different formal elements (whether stem alternants, theme vowels, affixes, stress patterns etc.), another goal of mine is, therefore, to analyze the interrelation of these various formal elements within the paradigm space. The orthogonality or isomorphy of formatives of every kind will be analyzed in various languages to answer the question of whether and to what extent can elements of pure form shape grammar.

## References

- ARONOFF, Mark. (1994). *Morphology by Itself: stems and inflectional classes*. Boston: MIT Press.
- BURZIO, Luigi. (2001). Zero derivations. *Linguistic Inquiry* 32.4: 658-677.
- CORBETT, Greville G. (2005). The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (Eds.) *Linguistic diversity and language theories*: 25-49. Amsterdam: John Benjamins.
- ESHER, Louise. (2014). Autonomous morphology and extramorphological coherence. *Morphology* 24, 4: 325-350.
- FEIST, Timothy Richard. (2011). *A grammar of Skolt Saami*. PhD Dissertation.
- PATO, Enrique & Paul O'NEILL. (2013). Los gerundios analógicos en la historia del español (e iberorromance). *Nueva Revista de Filología Hispánica* 61, 1: 1-27.

- ROUND, Erich R. (2013). Rhizomorphomes, meromorphomes, and metamorphomes. In Matthew Baerman, Dunstan Brown & Greville Corbett (Eds.), *Understanding and Measuring Morphological Complexity*: 29-52. Oxford: Oxford University Press.
- LUÍS, Ana & Ricardo BERMÚDEZ-OTERO (Eds.). (2016). *The Morphome Debate*. Oxford: Oxford University Press.
- TROMMER, Jochen & Sebastian BANK. (2017). Inflectional learning as local optimization. *Morphology* 27, 3: 383-422.

## Predicate highlighting in Yiddish and the question of “pseudo-infinitives”

**Alain KIHM,**

*CNRS-Université Paris-Diderot (Laboratoire de Linguistique formelle)*

alain.kihm@linguist.univ-paris-diderot.fr

### 1. Predicate highlighting: the syntactic and semantic evidence

I use “highlighting” as a cover term for “topicalization” and “focussing”, in order to avoid the often tricky issue of distinguishing one from the other from a discourse-oriented perspective.

Not only may arguments be highlighted in Yiddish by being realized in first position (thereby triggering verb-second) and related to a gap as in *A briv shraybt er* {a letter writes he} ?‘A letter he is writing’ (‘What he’s writing is a letter’), but predicates may as well (Cable 2004):

- (1) *Shraybn shraybt er a briv der mamen.*  
 write writes he a letter to.the mother  
 As for writing, he’s writing a letter to his mother.

- (2) *Shraybn a briv shraybt er der mamen.*  
 write a letter writes he to.the mother  
 As for writing a letter, he’s writing to his mother.

Only the head of the predicate is highlighted in (1), while the head and its first complement is in (2).

There being two occurrences of the verb in (1) and (2) is crucial in order to distinguish predicate highlighting (PRED-HI) from verb phrase highlighting (VP-HI) as in (3) (Jacobs 2005:260).,

- (3) *Esn an apl vil der man.*  
 eat an apple wants the man  
 To eat an apple is what the man wants.

VP-HI does not convey the same speech act as does PRED-HI. In (3) eating an apple is implicitly contrasted with other things the man could have wished to do. As such it constitutes a new discourse referent. In (1) and (2), in contrast, writing or writing a letter was already under discussion when the current speaker specifies the mother as the goal of the activity. Also see (4) where eating fish is clearly the matter under discussion:

- (4) *Esn fish est Maks hekht.*  
 to.eat fish eats Max pike  
 As for eating fish, Max eats pike.

## 2. Predicate highlighting: the morphological issue

In PRED-HI as in VP-HI the higher verb has to be a nonfinite form, i.e. an infinitive as in (1)-(2) or a past participle as in (4)-(5):

- (4) *Gegesn fish hot Maks gegesn.* (PRED-HI — Cable 2004:2)

eaten fish has Max eaten  
As for having eaten fish, Max ate (it).

- (5) *Gegesn fish hot Maks.* (VP-HI — Jacobs 2005:260)

eaten fish has Max  
Eat fish is what Max did.

In the PRED-HI construction, if the first occurrence is a past participle, the second must be one too: \**Gegesn est Maks fish* (*est* ‘eats’). On the other hand a higher infinitive may be paired with a lower past participle, as in (6):

- (6) *Esn hot Maks gegesn a fish.*

to.eat has Max eaten a fish  
As for eating, Max has eaten a fish.

Past participles raise no formal issue. Infinitives do, however. Two types of infinitives are distinguished in Yiddish. Regular infinitives show the same stem as the present tense (the only nonperiphrastic finite form with the imperative): cf. *shraybn* {ʃrajb-ŋ} ‘to write’ and *ikh shrayb* ‘I write’, *du shraybst* {ʃrajb-st} ‘you<sub>sg</sub> write’, *zi shraybt* {ʃrajb-t} ‘she writes’, *mir/zey shraybn* {ʃrajb-ŋ} ‘we/they work’, *ir shraybt* {ʃrajb-t} ‘you<sub>pl/pol</sub> write’. Infinitive and present tense share the one stem {ʃrajb}. Syllabic /ŋ/ realizes the feature [VFORM inf] or the feature set {[VFORM fin] [TENSE prs] [AGR 1/3pl]}. If the stem ends in a nasal (i.e. /n/, /m/, /ŋg/, /ŋk/), syllabic /l/, or a stressed vowel or diphthong, the ending is *-en* instead of *-ŋ*.

Irregular infinitives involve distinct stems in the infinitive and the present: cf. *visn* {vis-ŋ} ‘to know’ vs. *ikh veys* ‘I know’, etc. Six verbs belong to this class. Their full paradigms are given below:

	<b>zayn ‘be’</b>	<b>hobn ‘have’</b>	<b>gebn ‘give’</b>	<b>visn ‘know’</b>	<b>ton ‘do’</b>	<b>veln ‘want’</b>
<b>1SG</b>	ikh bin	ikh hob	ikh gib	ikh veys	ikh tu	ikh vil
<b>2SG</b>	du bist	du host	du gist	du veyst	du tust	du vilst
<b>3SG</b>	zi iz	zi hot	zi git	zi veyst	zi tut	zi vil
<b>1PL</b>	mir zaynen	mir hobn	mir gibn	mir veysn	mir tuen	mir viln
<b>2PL</b>	ir zayt	ir hot	ir git	ir veyst	ir tut	ir vilt
<b>3PL</b>	zey zaynen	zey hobn	zey gibn	zey veysn	zey tuen	zey viln

*Zayn* ‘to be’ is the most irregular as it involves five stems: {zaj} for the infinitive and 2PL; {bin} for 1SG; {bi} for 2SG; {iz} for 3SG; and {zajn} for 1/3PL. *Gebn* ‘to give’ comes next with three stems: {geb} for the infinitive; {gib} for 1SG and 1/3PL; and {gi} for 2/3SG and 2PL. *Hobn* ‘to have’ shows two stems: {hob} for the infinitive, 1SG, and 1/3PL; {ho} elsewhere. *Visn* ‘to know’, *ton* ‘to do’ and *veln* ‘to want’ also involve two stems, but their distribution is different: {vis}/{to}/{vel} in the infinitive vs. {vejs}/{tu}/{vil} elsewhere.

If PRED-HI involves one of the six verbs above, the fronted infinitive is not the “true” one, but it is a “pseudo” one using the present stem, as shown by (7) (Jacobs 2005:225):

- (7) *Veysn veyst er alts.*  
 know knows he everything  
 As for knowing, he knows everything.

Similarly, *Tuen tut er alts* ‘As for doing, he does everything’, *Viln vil er alts* ‘As for wanting, he wants everything’. With *gebn* and *hobn* the ‘long’ stems {gib} and {hob} are used: *Gibn git er alts* ‘As for giving, he gives everything’, *Hobn hot er alts* ‘As for having, he has everything’. In the latter case, the PRED-HI infinitive cannot be distinguished from the ordinary infinitive.

*Zayn* ‘to be’ builds its pseudo-infinitive on the three stems {bin}, {iz}, and {zajn}. Below are examples for the first two stems (Jacobs 2005:225):

- (8) *Binen bin ikh a kalifornyer.*  
 be am I a Californian  
 As for what I am, I am a Californian.
- (9) *Izn iz er a kamtsn.*  
 be is he a cheapskate  
 As to what he is, he is a cheapskate.

With the third stem the pseudo-infinitive is *zaynen*, identical with the finite 1/3PL forms: *Zaynen zaynen mir kalifornyer* ‘As for what we are, we are Californians’, *Zaynen zaynen zey kamtsns* ‘As for what they are, they are cheapskates’.

The 2SG/PL forms are special. The 2SG stem {bi} is unusable, so that recourse must be had to the true infinitive: *Zayn bistu a kamtsn* ‘As for what you are, you are a cheapskate’. Likewise, {zajn} has to be used instead of 2PL {zaj}: *Zaynen zayt ir kamtsns* ‘As for what you are, you are cheapskates’.

### 3. Pseudo-infinitives: a tentative account

Given the evidence presented by the six verbs above, the generalization can hardly be escaped that PRED-HI always involves pseudo-infinitives, even in the majority case where the latter are indistinguishable from regular infinitives owing to stem sharing. What exactly are pseudo-infinitives?

*Binen* and *izn*, whose morphological analyses are [bin-n] and [iz-n], give us an answer: pseudo-infinitives are infinitives that include tense and possibly agreement (person-number) features. Tensed infinitives are rather commonplace, such as e.g. the Latin perfect and future infinitives *amauisse* ‘to have loved’ and *amaturum esse* ‘to be about to love’, the latter periphrastic like English ‘to have loved’ (Ernout 1914/2002: 219, 229). Agreeing infinitives are also attested, for instance Portuguese so-called “inflected” infinitives.

Only *binen* and *izn* show both types of features because the stems and word-forms *bin* and *iz* are inherently present as well as 1SG and 3SG. That is to say, the features [TENSE *prs*] and [AGR 1sg]/[AGR 3sg] belong to their lexical entries, they are not morphously expressed. Given syncretism which makes -(e)n ambiguous, the same may be true of 1/3PL *zaynen* in the sense that -en is not recognized as a morph, so that *zaynen* may be considered another free stem like *bin* and *iz*.

The second persons singular and plural, in contrast, do show dedicated morphs, /-st/ and /-t/, expressing the person-number features. From which we conclude that inflected infinitives may only be formed on bases coextensive with bare stems, not on bases including morphs in addition to the stems. With the copula it excludes *bist* and *zayt*. With the other five verbs it excludes all forms but 1SG.

The ill-formedness of \*⟨bi-st-n⟩, \*⟨zaj-t-n⟩, \*⟨gi-t-n⟩, etc. may therefore be related to the exclusion of iterative inflection in Yiddish, a predominantly non-agglutinating language. Even when iterative inflection is the case, the TMA morph regularly precedes the person-number morph (Bybee 1985). Clearly \**bistn*, \**zaytn*, \**gitn*, etc. would run afoul of this general constraint, since the infinitive (mood) exponent follows the person-number exponents. But why not then inflect the true infinitive yielding \**zaynst*, \**zaynt*, \**gebnst*, etc. and respecting the constraint? (Compare Portuguese *cantarmos* /kānt-ar-mus/ ‘us to sing’.)

The question takes us back to the issue of the difference between (1)-(2) (PRED-HI) and (3) (VP-HI). In (3) the complement of the head verb *vil* ‘wants’ is a gap whose filler is the extracted VP *esn an apl* ‘eat an apple’. PRED-HI, in contrast, involves “copying”. Two occurrences of the same lexeme are thus copresent in the same sentence. (PRED-HI triggers verb-subject inversion, which shows the first occurrence to occupy the sentence’s initial field — see Weinreich 1949; Kathol 2000.) Such copresence is akin to resumption phenomena as in the following example (Jacobs 2005:237):

- (10) *Der yid vos mit im hot er geredt iz mayn shokhn.*  
 the man that with him has he spoken is my neighbour  
 The man he spoke to is my neighbour.

Resumptive *im* ‘him’ is token-identical with *der yid* ‘the man’, though they are lexically and morphosyntactically distinct — *im* is dative, whereas *der yid* is nominative. Likewise *shraybn* and *shraybt* in (1) or *binen* and *bin* in (8) are token identical in the sense that pseudo-infinitive *shraybn* and *binen* do not refer to writing and being in general as the true infinitives *shraybn* and *zayn* would, but to the same writing and being situations as are denoted by *shraybt* and *bin*. They share the same value of INDEX, in other words (Sag & Wasow 1999:106).

This is what rules out \**zaynst* and \**zaynt*. The true infinitive on which such counterfactual forms are built does not share the value of INDEX with the finite forms *bist* and *zayt*, while the pseudo-infinitives *binen*, *izn*, *zaynen*, *veysn*, etc. do, since they use free stems with inherent inflectional values —full sets in *binen* and *izn*, tense and number in *zaynen*, tense only in the remainder. Remarkably, as we saw, the bound stems ⟨bi⟩, ⟨zaj⟩, ⟨ho⟩, and ⟨gi⟩ are not available for pseudo-infinitive formation (no \*⟨bi-en⟩, \*⟨zaj-n⟩, \*⟨ho-en⟩, \*⟨gi-en⟩), implying they are not morphologically active. With the copula, *zaynen* can be mobilized at 2PL. There seems to be no remedy, in contrast, at 2SG, thereby forcing recourse to the true infinitive *zayn*. Further investigations with native speakers are necessary, however, to make sure this is the only solution.

In the vast majority of cases where true and pseudo infinitives are formally indistinguishable, as already pointed out, only the semantics of the construction is there to tell us that the infinitive “copy” actually is a sham.

Note that the true infinitive is used when the highlighted predicate is in the past tense: *Visn hob ikh es gevust* ‘As for knowing, I knew it’ (Cable 2004:5). This is expected, as the pseudo-infinitive *veysn* built on a present stem couldn’t share INDEX with a past form.

#### 4. Conclusion

Yiddish pseudo-infinitives represent a fascinating case of morphological creativity in situations of language contact. PRED-HI seems to have no direct equivalent in Middle High German or the Slavonic languages. It is likely to be a calque of the Hebrew infinitive absolute construction “used... to strengthen the verbal idea, i.e. to emphasize ... either the certainty... or the forcibleness and completeness of the occurrence” (Gesenius 2006:342). Transparent forms such as *binen* or *veysn* support a morphous view of morphology (Crysman & Bonami

2015) as they show that the Yiddish initiators did recognize /n/ as the infinitive exponent. Being clearly analogical, they do not force us to step outside the realizational framework (Blevins & Blevins 2009; Ackerman et al. 2009).

## References

- ACKERMAN, Farrell, BLEVINS, James P. & Robert MALOUF (2009). Parts and wholes: implicative patterns in inflectional paradigms. In J.P. Blevins & J. Blevins (eds), *Analogy in Grammar: Form and Acquisition*, 54-82. Oxford: Oxford University Press.
- BLEVINS, James P. & Juliette BLEVINS (2009). Introduction: Analogy in grammar. In J.P. Blevins & J. Blevins (eds), *Analogy in Grammar: Form and Acquisition*, 1-12. Oxford: Oxford University Press.
- BYBEE, Joan (1985). *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.
- CABLE, Seth (2004). Predicate cleft and base generation: evidence from Yiddish and Brazilian Portuguese. MIT, ms. (Downloadable from the author's web page)
- CRYSMANN, Berthold & Olivier BONAMI (2015). Variable morphotactics in Information-based Morphology, *Journal of Linguistics*.
- ERNOUT, Alfred (1914/2002). *Morphologie historique du latin*. Paris: Klincksieck.
- GESENIUS, Wilhelm (2006). *Gesenius' Hebrew Grammar. Edited and Enlarged by E. Kautzsch*. Mineola (NY): Dover Publications.
- JACOBS, Neil G. (2005). *Yiddish. A Linguistic Introduction*. Cambridge: Cambridge University Press.
- KATHOL, Andreas (2000). *Linear Syntax*. Oxford: Oxford University Press.
- Sag, Ivan A. & Thomas Wasow (1999). *Syntactic Theory. A Formal Introduction*. Stanford: CSLA Publications.
- WEINREICH, Uriel (1949). *College Yiddish*. New York: YIVO.

## French nominals in -ant: Semantic properties

**Marie Laurence KNITTEL,**  
*Université de Lorraine & ATILF*  
 marie-laurence.knittel@univ-lorraine.fr

### 1. Introduction

Our study focuses on French nominals ending in *-ant(e)* (henceforth Nant) converted from verbs in the present participle form (1a) and/or adjectives in *-ant* (1b), most of these being themselves arguably converted from present participles (1c). Our primary data is the list of 530 (381 masculine + 149 feminine) forms ending in *-ant/-ante* and registered as nouns in Lexique 3 (New et al. 2001, New 2006). After elimination of irrelevant forms (morphologically simple / out of use / belonging to specialized vocabulary / with an unclear semantic relationship with the base / prefixed after nominalization), our final list consists in 292 nouns, among which 84 are feminine forms alternating with masculine, that is 208 bases, comprising 6 feminine nouns only.

- |  |  |
|--|--|
| (1) a. ASSISTANT <sub>N</sub> < ASSISTER <sub>V</sub> 'to assist':<br>b. PUISSANT <sub>N</sub> < PUISSANT <sub>Adj</sub> 'powerful':<br>c. DÉCAPANT <sub>N</sub> < DÉCAPER <sub>V</sub> 'to strip':<br>< DECAPANT <sub>Adj</sub> 'stripper': | <i>J'ai vu les infirmiers assistant le chirurgien.</i><br>'I saw the nurses assisting the surgeon'<br><br><i>Cette drogue est très puissante</i><br>'This drug is very strong'<br><br><i>J'ai vu Paul décapant le plancher</i><br>'I saw Paul stripping the floor'<br><i>Ce produit est un excellent décapant</i><br>'This substance is an excellent stripper' |
|--|--|

We show that these nouns, apart from a few exceptions (*expectorant* 'expectorant') regularly refer to the subject of the underlying verbal or adjectival predicates (Roy & Soare 2012, 2014), and tend to follow two major and one minor patterns. (i) When unambiguously converted from verbs (1a), they nominalize the [+H(uman)] subject of the verbal predicate, regardless of its semantic role, and exhibit Masc./Fem. variation according to the gender properties of their referent. (ii) If they have an adjective as a possible base, and that the adjective predicates on [-A(nimate)] subjects, they are strictly masculine and tend to refer to substances, that is [-Animate] mass nouns. The function of nominalizing the subject of the underlying predicate can consequently be regarded as the main use of these Nant, their reference varying accordingly. Finally, (iii) if the adjective admits [+H] subjects, Nant also refers to [+H]. In such cases however, the direction of the conversion can be questioned.

### 2. Observations

#### 2.1. Nant reference

In order to determine the reference of Nant, we have used to Test 1:

- |   |                                      |
|---|--------------------------------------|
| T1.<br>(i) Max <sub>[+H]</sub> est un Nant<br>(ii) Ca, c'est {un / du} Nant | 'Max is a Nant<br>'This is (a) Nant' |
|---|--------------------------------------|

- (2) a. *Max est un {enseignant / habitant / \*calmant / \*édulcorant}*  
           Max is {a teacher / an inhabitant / sedative / sweetener}  
       b. *Ca, c'est {??un enseignant / ??un habitant / un calmant / de l'édulcorant}*  
           This is {a teacher /an inhabitant / (a) sedative / (a) sweetener}

If a Nant can describe a [+H] referent, such as *Max* (T1(i)), it qualifies as [+H]; conversely, if it describes a [-A], it occurs felicitously in T1(ii), the pronoun *ça* 'this' referring to inanimates. As shown in Table 1, although about 40% of Nant have inanimate referents, more than half refer to [+H], all having feminine counterparts. Note also that ambiguous [ $\pm$ H] are infrequent.

Referents	Number/	Examples
+ Animate	3	<i>rampant</i> (crawling insect)
+ Human	116	<i>commandant</i> (commander), <i>habitant</i> (inhabitant), <i>adoptant</i> (adopter)
± Human	6	<i>dominant</i> (dominant)
- Animate	82	<i>colorant</i> (dye), <i>gélifiant</i> (gellant), <i>adoucissant</i> (softener)

**Table 1** — Nant(e) reference

## 2.2. Category of the bases

The category of the base has been determined by the means of Test 2:

- |     |   |  |                           |   |
|-----|---|--|---------------------------|---|
| T2  | X est un Nant 'X is a Nant' =>                            | (i) X V /                                | (ii) X est (très) Adjoint | 'X is (very) Adjoint'   |
|     |   |  | (iii) un N Adjoint        | 'a Adjoint X'   |
| (3) | a. <i>X est un surveillant</i> =><br>'X is a oversee-ant' | (i) <i>X surveille</i> /<br>'X oversees' | (ii)                      | * <i>X est (très) surveillant</i><br>'X is (very) overseeing <sub>Adj</sub> ' |
|     |   |  | (iii)                     | * <i>un X surveillant</i><br>'an overseeing <sub>Adj</sub> X'                 |
| b.  | <i>X est un malveillant</i> =><br>'X is a malevolent'     | (i) [no verbal base] /                   | (ii)                      | <i>X est (très) malveillant</i><br>'X is (very) malevolent' <sub>Adj</sub> '  |
|     |   |  | (iii)                     | <i>un X malveillant</i><br>'a malevolent' <sub>Adj</sub> X'                   |
| c.  | <i>X est un stimulant</i> =><br>'X is a stimul-ant'       | (i) <i>X stimule</i> /<br>'X stimulates' | (ii)                      | <i>X est (très) stimulant</i><br>'X is (very) stimulating <sub>Adj</sub> '    |
|     |   |  | (iii)                     | <i>un X stimulant</i><br>'a stimulating <sub>Adj</sub> X'                     |

Nouns answering positively to T2(i) have been analyzed as deverbal (3a), those verifying either T2(ii)-T2(iii) are considered as deadjectival (3b) (see 3.3.). Finally, if they answer positively both, they are analyzed as potentially derived from verbs and adjectives (3c). Table 2 shows that, apart from 8 forms, all the bases are potentially (96) or exclusively (104) verbal.

Bases			
Verb	200	104	<i>gouvernant</i> (ruler), <i>assiégeant</i> (besieger)
Verb + Adjective		96	<i>démêlant</i> (conditioner), <i>marrant</i> (funny)
Adjective	8		<i>malveillant</i> (malevolent), <i>savant</i> (erudite)

**Table 2 —** V/Adjant(e) bases

### 2.3. The relation between the category of the base and the referent properties

Next, we have crossed the above parameters so as to examine the referent properties in the light of the category of the base. Our results are shown in Table 3:

Referent / Base	Verb	Verb + Adjective	Adjective
+Animate	2	1	
+Human	<b>81</b>	<b>29</b>	8
± Human	2	3	
- Animate	1	<b>81</b>	
<b>Total</b>	<b>86</b>	<b>114</b>	<b>8</b>

**Table 3** — Repartition of the referent in relation to the category of the base

Three major observations surface from Table 3 (see figures in bold).

- (i) strictly verbal bases give rise mostly to [+H] Nant (81/86); the same holds for strictly adjectival Nant.
- (ii) Nant with both verbal and adjectival bases tend to refer to [-A] (81/114, about 70%);
- (iii) in about 1/4 of the cases (29/114), Nant with ambiguous bases refer to [+H].

In what follows, we restrict our attention to these classes.

## 3. Analysis

### 3.1. [+Human] N with verbal bases

The [+H] properties of the referents of deverbal Nant are in most cases inherited from those of the prototypical subjects of the corresponding verbs. In other words, the base verbs tend to select primarily [+H] subjects (*surveiller* 'to oversee', *manifester* 'to demonstrate'), even though animates are generally possible (*arriver* 'to come', *migrer* 'to migrate'). Although agents-subjects are predominant, the agentivity tests T3 (Huyghe & Tribout 2015), confirm Roy & Soare's (2012, 2014) observation that Nant are not necessarily derived from agentive verbs; 1/4 of Nant refer to Themes (*arrivant* '(new)comer') or Patients (*mutant* 'mutant'), etc.

- |        |   |   |
|--------|---|---|
| T3     | (i) Nant a décidé de V  | 'Nant decided to V'   |
|        | (ii) Nant a {volontairement / consciencieusement / intentionnellement} Vé   | 'Nant {voluntarily / deliberately / conscientiously / intentionally} Ved' |
| (4) a. | <i>Le surveillant a décidé de surveiller / le communiant a décidé de communier / *l'arrivant a décidé d'arriver / *le mutant a décidé de muter</i><br>'The overseer decided to oversee / the communicant decided to communicate / the (new)comer decided to come / the mutant decided to mutate'  |   |
| b.     | <i>Le surveillant a consciencieusement surveillé / le communiant a volontairement communié / *l'arrivant est volontairement arrivé / *le mutant a intentionnellement muté</i><br>'The overseer oversaw conscientiously / the communicant communicated voluntarily / the (new)comer came voluntarily / the mutant mutated intentionally' |   |

Like [+H] Neur (Huyghe & Tribout 2015), these Nant can be implied in events in three manners, which are not mutually exclusive. First, they can refer to professions (*Paul est commandant / correspondant de presse* 'Paul is a commander / a press correspondent') or statuses (*Paul est étudiant (en histoire) / coopérant* 'P. is a (history) student / co-operant'). In such cases, even though the referents are designated by the means of the verbal base, they do

not need to have performed the action to be designated as such (*cet étudiant n'a jamais vraiment étudié* 'This student never studied'; *ce commandant n'a commandé personne* 'this commander never commanded anyone'). Statuses and professions being S-Level, these Nant can be associated with *futur* (*futur dirigeant* 'future ruler') or *ancien* 'former' (*ancien opposant* 'former opponent'). Second, Nant can refer to [+H] in designating them by their participation in an occasional specific event, which has to be performed for the designation to be valid (*Paul est un passant / le suppléant de Max* 'Paul is a passer-by / Max's substitute'). Finally they can refer to the participant of a series of events (habitual reading), in which case they admit modification by *petit* 'small, petty', *grand*, *gros* 'big' to denote frequency or intensity (*gros trafiquant* (de drogue) 'big (drug) dealer', *petit épargnant* 'small saver'). They also combine with *ancien / futur* for former or future habits (*ancien trafiquant* 'former dealer', *futur étudiant* 'future student').

### 3.2. [-Animates] with categorially ambiguous bases

This class of nominals can be derived either from verbs or adjectives, the latter being homophonous with present participles.

About 70% of the verbal bases present a regular alternation between [+H] and [-H] subjects (5a,b). Yet, as noted by Kupferman (1992) and Fradin (1997), the corresponding -ant adjectives admit only [-H] subjects (5c,d):

- (5) a. *Max a {collé le timbre / calmé le bébé / désherbé l'allée / coupé le rôti}*  
'Max has {stuck the stamp / calmed the baby / weeded the path / cut the roast}'
- b. *Le sirop a {collé dans le verre / calmé le bébé} / ce produit a bien désherbé l'allée / ce couteau coupe même les os*  
'The syrup has {stuck in the glass / calmed the baby} / this substance well weeded the path / this knife even cuts bones'
- c. \**Max est {collant / calmant / désherbant / tranchant}*  
'Max is{stick-ant / calm-ant / weed-ant / cut-ant}'
- d. *ce sirop est {collant / calmant} / ce produit est désherbant / ce couteau est coupant*  
'This syrup is {stick-ant / calm-ant} / this substance is weed-ant / this knife is cut-ant'

To sum up, the strictly [+dynamic] construal of the verbs is entailed by [+H] subjects only (5a); [-H] subjects combine with [±dynamic] verbs (5b), and the properties described by the corresponding adjectives only apply to [-A] (5c,d). The [-A] reference of these Nant suggests that they nominalize the external argument of the adjectival predicates.

The question that arises is that of the semantic value of these nominals. Since they refer to subjects of adjectival predications, they arguably nominalize Themes, or, more specifically Holders of Properties. However, they can also refer to participants of events (5b), and, due to their [-A] nature, can correspond to either Instruments, Causes or Means.

The Means role is straightforwardly excluded since Means occur only as subjects of [-dynamic] sentences (Fradin 2012). Note also that these nominals answer positively to tests T4 put forward by Huyghe & Tribout (2015) for Neur instruments.

- |        |   |                    |
|--------|---|--------------------|
| T4 (i) | <i>J'ai V {grâce à / avec / au moyen de / à l'aide de} Instr</i>  |                    |
|        | 'I have Ved {thanks to / with / by the means of / with the help of} Instr'  |                    |
| (ii)   | <i>J'utilise Instr pour V</i>   | 'I use Instr to V' |
| <br>   |   |                    |
| (6) a. | <i>J'ai {isolé / nettoyé / décapé} le sol {grâce à / avec / au moyen de / à l'aide de} {(de) l'isolant / du nettoyant / du décapant}'</i> |                    |

'I have {insulated / cleaned / stripped} the floor {thanks to / with / by the means of / with the help of}{insulator / cleaner / stripper}'

- b. *J'ai utilisé {de l'isolant/du nettoyant/du décapant} pour {isoler/nettoyer/décapier} le sol*  
 'I used {insulator / cleaner / stripper} to {insulate / clean / strip} the floor'

However, contrary to Neur Instruments, these Nant are not (necessarily) artefacts; their possible participation in the corresponding events is primarily due to their inherent properties (cf. (5b)).

Moreover, they do not need to be permanently controlled to perform the action; once applied, they act on their own. Their inherent properties enable them to behave as natural forces, as noted by Alexiadou & Schäfer (2006), and thus qualify them as Instrument Causers (Kamp & Rossdeutscher 1994) in the event described by the verbal predicate, which, in this case, is necessarily transitive. The same characterization applies to [-A] nouns converted from deverbal adjectives suffixed by -if (*répulsif* 'repellant' <sub>N[-A]/Adj</sub> > *repousser*<sub>V</sub> 'to repel') and built on verbal bases; conversely, nominals in -oir (*arrosoir* 'watering can' > *arroser* 'to water'), which have no corresponding adjectives, refer only to simple Instruments (Namer & Villoing 2008).

### 3.3. [+Humans] with categorially ambiguous bases

The 29 [+H] nouns with categorially ambiguous bases crucially differ from the former since both their verbal and adjectival bases combine with [+H] subjects (7a,b). Moreover, most of the corresponding adjectives, although they can occur as attributes (7c) (cf.T2ii), reject degree modification. It seems in fact that their bases are better analyzed as denoting states, like verbs, rather than properties (d).

7. a. *Paul {meurt / adopte / débute}*      'Paul {dies / adopts / begins}'
- b. *Paul est (\*très) {mourant / adoptant / débutant}*      'Paul is (very) {dying<sub>Adj</sub> / adopting<sub>Adj</sub> / begining<sub>Adj</sub>}'
- c. *Un soldat mourant / une famille adoptante / un pianiste débutant*  
             'A dying soldier / an adopting family / a beginning pianist'
- d. *mourant:*      = who is dying  
                           ≠ who has the property of dying (= mortel 'mortal')

To account for these properties, we hypothesize that these Nant are strictly deverbal, and that the corresponding adjectives are in fact derived by conversion.

### References

- ALEXIADOU, A. & F.SCHÄFER, (2006). Instrument subjects are Agents or Causers. Proceedings of the 25th WCCFL ed. D.Baumer et al. eds. Somerville, MA: Cascadilla Proceedings Project, 40-48.
- ANSOMBRE, J.C. (2001). A propos des mécanismes sémantiques de formation de certains noms d'agent en français et en espagnol. *Langages* 143, 28-48.
- FRADIN, B. (1997). Esquisse d'une sémantique de la préfixation en -anti. RLV 26: 87-112.
- FRADIN, B. (2012). Les nominalisations et la lecture 'Moyen' *Lexique*: 125-152.
- HUYGHE, R. & TRIBOUT, D. (2015). Noms d'agents et noms d'instruments : le cas des déverbaux en -eur. *Langue Française* 185-1.
- KAMP, H. & ROSSDEUTSCHER, A. (1994). Remarks on lexical structure and DRS construction. *Theoretical Linguistics*, 20, 97-164.

- KUPFERMAN, L. (1992). Une morphologie dispositionnelle ? In L. Tasmowski & A. Zribi-Hertz (eds) *De la musique à la linguistique, hommages à Nicolas Ruwet*. Ghent: Communication & Cognition. 345-359.
- NAMER F. & F. VILLOING (2008). Interpréter les noms déverbaux: quelle relation avec la structure argumentale du verbe de base? le cas des noms en -OIR du français, 1er Colloque Mondial de Linguistique Française, ILF, Paris: 1539-1557.
- NEW, B., PALLIER C., FERRAND L. & R. MATOS (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE, *L'Année Psychologique* 101, 447-462.
- NEW, B. (2006). Une nouvelle base de données lexicales. Actes de la conférence TALN, Louvain : 10-13 avril 2006. (<http://www.lexique.org>).
- ROY, I. & E. SOARE, (2012). L'enquêteur, le surveillant et le détenu : les noms déverbaux de participants aux événements, lectures événementielles et structure argumentale. *Lexique 20 : Nominalisations : nouveaux aspects*. R. Marín & F. Villoing (eds.), 207-231.
- ROY, I. & E. SOARE (2014). Les noms d'humains dérivés de participes : dérivés nominaux en ant et é/i/u. CMLF 2014.

## (Dis)favoring polycategoriality in word formation? Evidence from Palestinian Arabic and Hebrew

**Lior LAKS,**  
*Bar-Ilan University*  
 Lior.Laks@biu.ac.il

This study examines the tendency to (dis)favor polycategoriality and ambiguity, and the conditions that allow or block it (see Mithun 2010, Vapnarsky & Veneziano, forthcoming). I examine differences between two closely related languages, Palestinian Arabic (PA) and Modern Hebrew (MH). The former tends to avoid polycategoriality, while the latter largely allows it. In addition, even in Hebrew, polycategoriality is not completely productive. I argue that this is dictated by the component of the grammar where words are formed, the lexicon and the syntax. Specifically, I claim that the morphological mechanism aims at creating a distinction between categories only when both are stored in the lexicon.

Semitic morphology relies highly on non-concatenative morphology, where words are formed in patterns (Berman 1978, Bolozky 1978, Bat-El 1994, Benmamoun 1999, 2003, Ussishkin 1999). PA and Hebrew participle patterns have a special status with regard to the lexical categories that they host (Berman 1978, 2017, Bat-El 2008): present tense verbs, nouns and adjectives. The Hebrew participle pattern *maCCiC*, for example consist of ambiguous forms that function as verbs and nouns (*madrix* 'is guiding/a guide') or verbs and adjectives (*maksiim* 'is charming/charming').<sup>1</sup>

While the option for such ambiguity exists both in Hebrew and in PA, only Hebrew demonstrates a tendency to employ it productively. Morphological variation in Hebrew derived nominals provides evidence for it. Instrument nouns (INs) and agent nouns (ANs) tend to receive additional forms in participle patterns, making them identical to verbs. For example, *masnen* 'a filter' is formed in the *maCCeC*, pattern, which is used only for noun formation. However, it has a doublet in the *meCaCeC* participle pattern (*mesanen*), which also denotes the present form of the verb *sinen* 'filtered'. While *masnen* denotes an IN and *mesanen* denotes both a verb and an IN, *mesanen* is used more productively, similarly to other such ambiguous forms that denote both verbs and INs/ANs. Not all INs and ANs undergo such variation, but when they do, the change is always into participle patterns. The change is motivated by semantic and morphological transparency between derived nominals and verbs, based on systematic guidelines, as discussed in previous studies (Laks 2015, 2017). The main issue here is that Hebrew, which is far less polycategorial than languages like English (Berman 2017), tends to tolerate polycategoriality and trigger its formation.

Palestinian Arabic (PA), in contrast, largely avoids polycategoriality. Participle patterns that can denote verbs, adjectives and nouns are rarely ambiguous. The form *ka:teb*, for example,

---

<sup>1</sup> Participle forms are used for other tenses as well, but this is irrelevant for the purposes of this study. I address only lexical category distinctions.

denotes both 'a writer' and 'is writing', the present form of the verb *katab* 'wrote'. However, such cases are rather rare. Unlike Hebrew, PA INs/ANs tend to retain their exclusive nominal patterns and do not have doublets in participle patterns. For example, the verb *dahan* 'painted' has a derived nominal *dahha:n* 'a painter' in the *CaCCa:C* nominal pattern, and the participle *da:hen* only denotes 'is painting'. Moreover, when ANs or INs are formed in participle patterns, these forms are used exclusively as nouns and the verbal meaning is blocked. For example, the verb *ħaras* 'guarded', has a derived AN in the participle pattern *Ca:CeC* (*ħa:res* 'a guard'), which cannot mean 'is guarding'. This meaning is expressed via the imperfective pattern (*byuħrus* 'is guarding').

In addition to the differences between PA and Hebrew, the current study shows that even languages that tend polycategoriality like Hebrew, tend to restrict it based on systematic guidelines. When Hebrew allows polycategoriality, it is mostly when at least one of the forms is not stored/derived in the lexicon, but is derived in the syntax. In ANs/INs doublet formation, only the nouns are assumed to be stored in the lexicon, while the present tense verbs are inflected forms of the base past forms. Assuming that inflected forms are mostly derived in the syntax (Aronoff 1976, Anderson 1982, 1992, Scalise 1984, Perlmutter 1988 among others), the two meanings of the ambiguous form are represented in different components of the grammar. This is also evidenced in the fact that there are hardly any cases of participle forms that denote both nouns and adjectives, but are mostly used for inflected verbs and another category. This is because such nouns and adjectives are not inflected forms, but are assumed to be stored entries in the lexicon. Past forms of verbs are also assumed to be stored in the lexicon as basic entries. Indeed, there are only rare cases where past forms of verbs are identical to nouns or adjectives (e.g. *iver* 'blind/made X blind').

It follows that the grammar tends to either avoid polycategoriality and ambiguity, as much as possible (PA), or tend to allow it, as long as only one form is stored in the lexicon (Hebrew). It should be noted that this reflects strong tendencies rather than clear-cut dichotomies. The talk will highlight the typological differences between PA and Hebrew with respect to (dis)allowing polycategoriality. Such a distinction could lead to some hierarchy of the degree to which languages tolerate polycategoriality, which can be further examined cross-linguistically.

## References

- ARONOFF, Mark. (1976). *Word formation in Generative Grammar*. Cambridge: MIT Press.
- ANDERSON, Stephen.R. (1982). Where is morphology? *Linguistic Inquiry* 13, 571-612.
- ANDERSON, Stephen R. (1992). *A-morphous morphology* (*Cambridge Studies in Linguistics*, 62. Cambridge: Cambridge University Press).
- BAT-EL, Outi. (1994). Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory* 12, 572-596.
- BAT-EL, Outi. (2008). Morphologically conditioned V-O alternation in Hebrew: Distinction among nouns, adjectives and participles, and verbs. In S. Armon-Lotem, G. Danon & S. Rothstein (eds.), *Current Issues in Generative Hebrew Linguistics*. Amsterdam: John Benjamins. 27-60.
- BENMAMOUN, Elabbas. (1999). Arabic morphology: The central role of the imperfective. *Lingua* 108: 175–201.
- BENMAMOUN, Elabbas. (2003). The role of the imperfective template in Arabic morphology. In J. Shimron (ed.), *Language Processing and Acquisition in languages of Semitic, root-Based, morphology*.Amsterdam: John Benjamins. 99-114.

- BERMAN, Ruth A. (1978). Modern Hebrew structure. Tel Aviv: University Publishing Projects.
- BERMAN, Ruth A. (2017). Word Class Distinctiveness versus Polycategoriality in Modern Hebrew: Typological and Psycholinguistic Perspectives. In Valentina Vapnarsky and Edy Veneziano (eds.), *Lexical Polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches*. Amsterdam: John Benjamins. 343-377.
- BOLOZKY, Shmuel. (1978). Word formation strategies in Modern Hebrew verb system: denominative Verbs. *Afroasiatic Linguistics* 5, 1-26.
- LAKS, Lior. (2015). Variation and change in instrument noun formation in Hebrew and its relation to the verbal system. *Word Structure* 8.1, 1-28.
- LAKS, Lior. (2017). Morphological and semantic transparency in Hebrew agent noun formation. In M. Bloch-Trojnar & A. Malicka-Kleparska (eds.), *Aspect and Valency in Nominals, Studies in Generative Grammar* 134. Berlin/Boston: De Gruyter Mouton. 229-252.
- MITHUN, Marianne. (2010). Polycategoriality and conversion: Categorical distinction or cline of productivity? A paper presented at the International Conference on Polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches. École Normale Supérieure, Paris.
- PERLMUTTER, David. (1988). The split-morphology hypothesis: evidence from Yiddish. In M. Hammond and M. Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*. Orlando: Academic Press. 79-99.
- SCALISE, Sergio. (1984). *Generative Morphology*. Dordrecht: Foris.
- USSISHKIN, Adam. (1999). The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output-output correspondence. *Phonology* 16, 401-442.
- VAPNARSKY, Valentina & Edy VENEZIANO (eds.). (2017). *Lexical Polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches*. Amsterdam: John Benjamins.

# Compétences morphologiques compositionnelles en production chez des sujets sains francophones dans une tâche de dénomination d'images

**Alice MILLET,**

*Université de Neuchâtel, Suisse*

[alice.millet@unine.ch](mailto:alice.millet@unine.ch)

**Noémie AUCLAIR-OUELLET,**

*McGill University, Montreal, Canada*

[noemie.auclairouellet@mcgill.ca](mailto:noemie.auclairouellet@mcgill.ca)

**Marion FOSSARD,**

*Université de Neuchâtel, Suisse*

[marion.fossard@unine.ch](mailto:marion.fossard@unine.ch)

## 1. Introduction

Les études sur les trois domaines de la morphologie ont montré que les capacités en flexion, en dérivation et en composition peuvent être perturbées suite à une lésion cérébrale. Cette étude porte spécifiquement sur la morphologie compositionnelle.

Du point de vue linguistique, la morphologie compositionnelle, comme la morphologie dérivationnelle, participe à la création lexicale de la langue [1]. La dérivation fonctionne selon un processus d'affixation tandis que la composition suit un processus de combinaison de deux (ou plus) unités lexicales dotées de l'autonomie reconnue aux mots [2]. En français, il existe différentes manières de réaliser des mots composés : ils peuvent être écrits en un mot, séparés d'un espace, reliés par un tiret, une apostrophe ou une préposition [3].

Les mots composés entiers comme leurs constituants peuvent appartenir à diverses catégories grammaticales et les processus de composition de mots en français suivent différentes règles. En raison de cette hétérogénéité, le statut de « mot » n'est pas unanimement reconnu ; toutefois, un mot composé présente une cohésion interne que le syntagme ne possède pas [2].

En psycholinguistique, les mots composés présentent un intérêt particulier pour leur contribution dans le débat sur la façon dont les mots sont représentés et récupérés. Les études sur des patients cérébrolésés indiquent que les mots composés sont récupérés à la fois de manière décomposée (i.e. par leurs constituants) et en mots entiers [4 ; 5]. L'un des effets qui a été mis en avant est le *compound effect*, c'est-à-dire la tendance des patients aphasiques à substituer des mots composés par d'autres formes composées [4 ; 6]. Plusieurs autres effets ont également été investigués : les effets de position en faveur du premier constituant [7], de tête [8] ou de transparence [9 ; 10] ont été observés.

Les études sur les capacités morphologiques en composition sont en pleine expansion. Toutefois, bien que diverses langues aient été investiguées dans le domaine de la composition, les études sur le français sont rares. Ayant observé l'existence de troubles morphologiques en composition, une question d'intérêt est de savoir comment les mettre en évidence en clinique auprès de patients cérébrolésés et surtout de savoir comment les interpréter. Comparé à l'anglais, le nombre d'outils d'évaluation pour les francophones est bien plus restreint [11].

Or, chaque langue varie dans sa fréquence d'usage de mots composés, des règles pour les construire et un mot composé dans une langue ne l'est pas nécessairement dans une autre (ex.: *swimming pool* vs. *piscine*). Une traduction d'une langue à une autre n'étant pas envisageable, il est nécessaire de construire une tâche spécifique aux caractéristiques du français, à même de mettre au jour les compétences compositionnelles de sujets sains et d'établir des normes. La dénomination sur images s'avère être une tâche pratique à proposer en clinique. Les outils actuels ne proposent pas de tâche qui cible cet aspect et qui manipule des variables psycholinguistiques pertinentes.

## 2. Objectifs

Dans le cadre de cette étude, une tâche de dénomination d'images ciblant la production de noms composés (NC par la suite) a été soumise à un petit échantillon de sujets francophones, sans troubles cognitifs avérés. Leurs performances ont été étudiées pour identifier l'effet potentiel de variables psycholinguistiques.

## 3. Méthodologie

**Population :** 37 personnes âgées entre 46 et 75 ans ( $M = 60,63$  ;  $ET = 7,04$ ) (19 hommes et 18 femmes) de langue maternelle française ont accepté de participer à l'étude. Pour être retenues, les personnes ne devaient pas présenter de troubles du langage ni d'antécédents neurologiques.

**Matériel :** une tâche de dénomination d'images visant la production de mots composés a été créée sur logiciel PowerPoint. Seuls des items nominaux ont été sélectionnés. Les items consistaient en des photographies en couleurs issues de différentes sources (la BOSS [12], Pixabay et Common.wikimedia.org). Les NC ( $n = 39$ ) ont été comparés à des noms non composés ( $n = 36$ ). La longueur de ces noms non composés a été manipulée (une à quatre syllabes). Pour ces noms non composés, la variable de fréquence a été contrôlée ( $< 35$ ) à partir de la base de données de fréquence de Lexique.org [13]. En revanche, les données sur la fréquence pour les NC sont lacunaires. Pour pallier à ces données manquantes, en plus de la fréquence pour les items mentionnés sur Lexique.org, une estimation du *name agreement* des items a été réalisée en se référant à des études sur le *name agreement* de sujets francophones [12 ; 14] et en effectuant un prétest auprès de 30 étudiants à partir d'une tâche de dénomination d'images. Les NC ont été manipulés en fonction de leur structure interne (six Nom-Nom ; sept Nom-Adjectif ; six Adjectif-Nom ; huit Verbe-Nom ; six Nom-Préposition-Nom ; six Nom-Préposition-Verbe). Pour définir la transparence de la forme globale des NC, 30 étudiants ont été soumis lors d'un prétest à une tâche de jugement de transparence : 54 NC leur ont été présentés oralement et à l'écrit sur PowerPoint. Ils devaient estimer, sur une échelle de Likert, à quel point il était facile de se représenter le sens transmis par le NC à partir de ses constituants. Trente-neuf NC ont été retenus et répartis selon trois degrés de transparence (transparent, partiellement transparent, et opaque). Etant donné le nombre réduit de NC opaques inclus dans l'étude ( $n = 3$ ), ceux-ci n'ont finalement pas été pris en compte dans les présentes analyses, lesquelles ont donc été menées sur les NC transparents ( $n = 18$ ) et partiellement transparents ( $n = 18$ ).

**Procédure :** Les items étaient présentés sur un écran d'ordinateur et leur enchaînement était automatique. Leur présentation suivait un ordre aléatoire (pas plus de deux items simples ou composés à la suite, et pas plus de deux structures internes de NC identiques à la suite). La tâche était précédée de deux items de pratique. Le participant produisait sa réponse oralement. Dans certaines conditions définies au préalable, une relance pouvait être faite par l'expérimentatrice.

Cotation et analyses: les réponses ont été transcrives et cotées comme bonne réponse ou mauvaise réponse. Un score strict (seule la première réponse était prise en compte) et un score large (incluant les autocorrections) ont été calculés. Les scores ont été entrés sur fichier informatique Excel pour effectuer les analyses descriptives (score brut, moyenne, écart-type, pourcentage). Les analyses statistiques inférentielles ont été effectuées à l'aide du logiciel SPSS. En raison de l'absence d'une distribution normale des résultats, ce sont des tests non paramétriques (Friedman et Wilcoxon) qui ont été utilisés. Le seuil de signification statistique a été ajusté pour les comparaisons multiples (correction de Bonferroni).

#### 4. Résultats<sup>1</sup>

##### Moyenne et pourcentage de réussite en tâche de dénomination d'images en fonction de la structure globale des items, de la structure interne et de la transparence des NC

Variables	Valeurs	Moyenne	Ecart-type	Pourcentage
Structure globale	Noms non composés (36)	31,19	2,99	86,63%
	NC (36)	31,38	2,81	87,16%
Structure interne des NC	NN (6)	5,3	0,7	88,29%
	NA (6)	4,08	1,21	68,01%
	AN (6)	4,97	0,99	82,88%
	VN (6)	5,54	1,09	92,34%
	NpN (6)	5,7	0,57	95,05%
	NpV (6)	5,81	0,4	91,85%
Transparence des NC	Transp. (18)	17,11	0,94	95,05%
	Part. trans. (18)	14,27	2,38	79,28%

La différence de performance entre noms non composés et NC n'est pas significative ( $Z = -0,625$ ,  $p = 0,532$ ). Le test de Friedman montre que la structure interne des NC a un effet sur le taux de réussite ( $X^2(5) = 76,844$ ,  $p = 0,000$ ) : globalement, les NC qui contiennent une préposition et/ou un constituant verbal tendent à être mieux réussis que les autres structures. Les NC sont surtout substitués par des noms simples mais ils sont toutefois plus souvent substitués par une structure composée (existante ou néologique) en comparaison des noms simples. Les erreurs d'omission ou de substitution d'un constituant concernent essentiellement le deuxième constituant. Sur les erreurs respectant la structure composée des NC ayant une tête (endocentriques<sup>2</sup>), une grande majorité concerne le modifieur (dernier constituant). Un effet de transparence est observé, les NC transparents étant mieux réussis que les NC partiellement transparents ( $Z = -4,268$ ,  $p = 0,000$ ).

<sup>1</sup> Seuls les résultats pour le score strict sont présentés ici.

<sup>2</sup> Les NC endocentriques sont des NC dont l'un des constituants est la tête qui permet de définir les propriétés sémantiques (ex.: un *château de cartes* est un type de château) et syntaxiques (ex.: un *château de cartes* est au masculin singulier) [16]

## 5. Conclusion

Les résultats montrent que les performances des participants sont influencées par la structure interne des NC et leur degré de transparence. Il se peut que l'effet de structure interne en faveur des NC contenant une préposition soit dû au fait que cette structure, la plus productive en langue française [15], soit aussi de nature à disposer d'un degré de transparence plus élevé. La différence significative entre les NC transparents et partiellement transparents peut s'expliquer par le fait que les NC partiellement transparents reçoivent moins de support de la part des représentations de leurs constituants au niveau sémantique contrairement aux NC transparents [10], ou que les NC partiellement transparents sont plus coûteux à traiter, en raison d'un conflit entre la signification du mot entier et celui de la forme composée construite ad hoc à partir des constituants [9]. Plusieurs erreurs (substitution ou manque du mot sur un constituant) sont cohérentes avec le *compound effect* observé chez les patients cérébrolésés. Ces erreurs indiquent que les participants avaient conscience du caractère composé des items, ce qui soutient l'hypothèse que les connaissances sur le statut composé des mots seraient stockées indépendamment de celles sur leur forme phonologique [6]. Les erreurs de substitution ou d'omission concernent le plus souvent le dernier constituant. Par ailleurs, ces erreurs qui respectent la structure composée de la cible concernent surtout les NC endocentriques et c'est notamment le constituant tête qui est produit. Toutefois, le nombre d'erreurs est restreint et pour réellement distinguer l'effet de tête de celui de position, il faudrait comparer un nombre suffisamment important d'erreurs produites sur des structures où la tête est à gauche (ex.: structure Nom-Adjectif) et d'autres à droite (ex.: structure Adjectif-Nom).

La réalisation de cette étude a mis en évidence l'influence de facteurs psycholinguistiques sur la performance de participants sains en morphologie compositionnelle. L'élargissement de l'échantillon de participants serait nécessaire afin d'obtenir des données supplémentaires qui permettent davantage d'investigations sur ces effets, ainsi que l'établissement de normes en vue d'une utilisation de cette tâche dans la clinique. Plus généralement, la poursuite d'études dans le domaine encore négligé de la composition est à encourager, en envisageant d'autres mesures telles que les temps de réponse, pour permettre de trancher en faveur de certains modèles théoriques et de répondre à des questionnements nouveaux que cette étude a fait naître.

## Références

- [1] BAUER, L. (2008). Derivational Morphology. *Language and Linguistics Compass* 2/1, 196)210.
- [2] MOESCHLER, J. & AUCHLIN, A. (2009). Chapitre 5 : Morphologie. Mot et morphème. In J. Moeschler & A. Auchlin (Eds). *Introduction à la linguistique contemporaine* (pp. 58-67). Paris : Armand Colin.
- [3] LIEBER, R. (2009). Chapter 3: Lexeme formation: the familiar. In R. Lieber (Ed). *Introducing morphology* (pp. 51-58). Cambridge University Press.
- [4] BADECKER, W. (2001). Lexical composition and the production of compounds: evidence from errors in naming. *Language and cognitive processes*, 16(4), 337-366.
- [5] KUPERMAN, V., BERTRAM, R. & BAAYEN, R. (2008). Morphological dynamics in compound processing. *Language and cognitive processes*, 23 (7/8), 1089-1132.
- [6] CHIARELLI, V. MENICHELLI, A. & SEMENZA, C. (2005). Naming compounds in aphasia and in Alzheimer's disease. *Brain and Language*, 95, 137-138.
- [7] MÄKISALO, J. NIEMI, J. & LAINE, M. (1999). Finnish compound structure: Experiments with a morphologically impaired patient. *Brain and Language*, 68, 249-253.
- [8] BLANKEN, G. (2000). The production of nominal Compounds in aphasia. *Brain and Language*, 74, 84-102.

- [9] JI, H., GAGNÉ, C. L. & SPALDING, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, 65, 406-430.
- [10] ZWITSERLOOD, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9(3), 341–368.
- [11] MACOIR, J. GAUTHIER, C., JEAN, C. & POTVIN, O. (2016). BECLA, a new assessment battery for acquired deficits of language: Normative data from Quebec-French healthy younger and older adults. *Journal of the Neurological Sciences*, 361, 220-228.
- [12] BRODEUR, M. B., DIONNE-DOSTIE, E., MONTREUIL, T. & LEPAGE, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS One*, 5 (5), 1-13.
- [13] NEW B., PALLIER C., FERRAND L. & MATOS R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE, L'Année Psychologique, 101, 447-462.
- [14] GUERARD, K., LAGACE, S. & BRODEUR, M. B. (2015). Four types of manipulability ratings and naming latencies for a set of 560 photographs of objects. *Behavior Research Methods*, 47(2), 443-470.
- [15] MONDINI, S., LUZZATTI, C., SALETTA, P., ALLAMANO, N. & SEMENZA, C. (2005). Mental representation of prepositional compounds: evidence from italian agrammatic patients. *Brain and Language* (94), 178-187.
- [16] BOOIJ, G. (2014). Chapter 7: Morphology: The structure of words. In K. Allan (Ed). *The Routledge Handbook of Linguistics* (pp. 104-107). London & New York : Routledge.

## The Continuum from Epenthesis to Linking Elements

**Sedigheh MORADI,**  
*Stony Brook University*

**Mark ARONOFF,**  
*Stony Brook University*

**Lori REPETTI,**  
*Stony Brook University*

We show that there is a continuum in the factors conditioning the insertion of semantically empty phonological material, running from purely phonologically conditioned epenthesis through morphologically conditioned epenthesis to linking elements (LE) (Table 1). This continuum is a one-directional cline of grammaticalization of semantically empty material: from phonologically (/phonetically) motivated epenthesis, to added morphological conditions, to dropping the phonological motivation altogether (linking elements<sup>1</sup>).

	Canon. Ep.	Non-Canon. Ep.	Linking Element <sup>2</sup>
Presence required by phonology	+	+	- <sup>3</sup>
Realization motivated by phonology	+	-	-

**Table 1** — Semantically empty elements

The term *canonical epenthesis*, in Table 1, refers to the insertion of a segment whose occurrence and quality are both motivated entirely by phonology (Kitto & de Lacy 1999). The Japanese data in (1) are examples of canonical epenthesis in loan words.

- (1) a. allergy [æləɪdʒɪ] > [ærəwɪgɪ:] [u]: default epenthetic vowel  
 b. match [mætʃ] > [mat̪t̪ɪ] [i]: after palato-alveolar affricates  
 c. eight [eɪt] > [eɪt̪o] [o]: after alveolar stops

<sup>1</sup> In other words, we assume that linking elements are epenthesized segments without phonological conditions.

<sup>2</sup> Linking elements are marked by full phonemes, e.g. /s/ in German. Basque, on the other hand, contains a featural linking element whose phonological content is [-voice, -continuant] and is inserted at the beginning of the second element of the compound (Labrune 2014). Japanese and Korean show linking features too (Labrune 2014, Koike 1996).

<sup>3</sup> Nübling and Szczepaniak (2008) discuss examples of German compounds where LEs act as a prosodic means of PW optimization.

In *non-canonical epenthesis* the motivating factor for the presence of the segment is phonological, but the *quality* of the segment is not phonologically driven. The cases we focus on here are all morphologically conditioned.<sup>4</sup> For the most part, they have been described as allomorphy in the literature. Looking at the phenomenon as a case of epenthesis, however, helps us understand why we choose one form rather than another, and why the alternation exists.

We review a number of such cases, drawn from typologically diverse languages, including Persian (2), Arabic (3), Mohawk (4), many Romance varieties (Italian (5), San Marino, Veneto, Catalan, etc.), Hungarian, and Seri. These languages repair an ill-formed structure with a default epenthetic segment in most contexts, but a special epenthetic segment is used to repair the same ill-formed structure in a particular morpho-syntactic environment.

(2) Persian: *VV	[?]	default	/batʃe -i/ → [batʃe <u>?</u> i] child INDF → ‘a child’
	[g]	before noun marker /-i/	/batʃe -i/ → [batʃegi] child N → ‘childhood’
(3) Arabic: *#CC	[i]	default	/smaʃ/ → [ʔismaʃ] hear.2SG.IMP → ‘hear!’
	[u]	person + verb (class 2, 3, 4)	/t-darrisu/ → [tu <u>d</u> arrisu] 2SG-teach (class 2) → ‘you.sg teach’
	[a]	person + verb (other classes)	/t-smaʃu/ → [ta <u>s</u> maʃu] 2SG-hear (class 1) → ‘you.sg hear’
(4) Mohawk: *CC <sup>5</sup>	[e]	default	/s-wa-nuhwe <u>?</u> -s/ → [se <u>w</u> anú:we <u>?</u> s] 2-PL-like-HAB → ‘you.pl like it’
	[a]	C <sub>noun</sub> + C <sub>verb</sub>	/hr-atʌ-yen-rho-s/ → [ratʌyen <u>á</u> rhos] MA-SRF-oil-spread-HAB → ‘he is greasing up’
(5) Italian: *CC	[i]	default	‘psychology’ → [pisikologo]
	[o]	word-finally	/l specchio/ → [l <u>os</u> .pæk.kjo] <sup>6</sup> DEF mirror

There are at least three other types of semantically empty inserted elements, whose properties overlap with our broad definition of non-canonical epenthesis.

<sup>4</sup> For instances of syntactic epenthesis see Cardinaletti and Repetti (2008).

<sup>5</sup> The specific illegal consonant sequences are not discussed here. See Rawlins (2006).

<sup>6</sup> For a discussion on the underlying form of the definite article in Italian, see Repetti (ms) and the references therein.

## **1. Augment (infix) /isk/ in Italian and Maltese**

In Italian, <isc> has no semantic role, and is found in certain 4<sup>th</sup> conjugation class verbs (7).

- (7)    \*fín-o              fin-ísc-o              'I finish'  
       cfr. fin-í              \*fin-isc-í              'I finished'

Some argue that the presence of <isc> is phonologically motivated. It is used with verbs whose stems cannot bear stress, so the <isc> instead is stressed. However, according to DiFabio (1990: 4), stress is a consequence of the presence of <isk>.

Hoberman & Aronoff (2003: 74) show that the same augment is found in Italian verbs borrowed into Maltese. The augment appears “in just those cases where the stem would otherwise be stressed... This is a case of borrowing a phonological condition on a morphological rule.” Notice in the data in (8) that the infix (<ixx> in Maltese) may be present in different cells in the paradigms of the two languages.

- |     |   |  |
|-----|---|--|
| (8) | Italian   | Maltese                                |
|     | +SK sugger-ísc-o (*suggér-o)<br>'I suggest'     | +SK ni-ssugger-íxx-i (*ni-suggér-i)    |
|     | +SK sugger-ísc-o (*suggér-o)<br>'I suggest'     | -SK i-ssugger-í (*)i-sugger-ixx-í      |
|     | -SK sugger-í (*sugger-isc-í)<br>'she suggested' | -SK i-ssugger-íet (*i-ssugger-ixx-íet) |
|     | -SK sugger-í (*sugger-isc-í)<br>'he suggested'  | +SK i-ssugger-íxx-a (*i-ssuggér-a)     |

## 2. Stem extenders in Catalan

In some dialects of Catalan, second person singular imperatives of conjugation classes II and III have an “extension” included when enclitics are added; the “extension” has no semantic role while the addition of an extension is driven by a prosodic constraint: a moraic trochee is built at the right edge of the verb+enclitic unit. (Bonet and Torres-Tamarit 2009). The form of the extender is taken from other forms of the verb.

- (9) [búλ] 'boil!'      [buλ-iyé-lə] 'boil it (fem)!' (with [iyə] extension)

### 3. Ante-suffixal interfixes in Spanish<sup>7</sup>

Spanish interfixes are inserted elements, e.g. <ec>, <ol>, etc., whose distribution is morphologically conditioned (i.e., after certain stems and before certain suffixes). These elements have no meaning or connotative value, but their distribution is phonologically influenced (short bases favor insertion of the interfix). (Dressler & Merlini Barbaresi 1994; Prati 1942).

- (10) madr-e ‘mother’ madr-ec-ita ‘(dim)’  
      comadr-e ‘godmother’ comadr-ita ‘(dim)’ (\*comadr-ec-ita)

<sup>7</sup> The same kind of semantically empty ante-suffixal interfixes can be found in Russian, Polish, and Italian (Dressler & Merlini Barbaresi 1994).

We go on to compare the properties of morphologically conditioned epenthetic segments (non-canonical epenthesis) with linking elements. While linking elements may or may not be phonologically motivated, their presence (and quality) play a morphological role in marking compounds (6a) and other morphological boundaries (6b) (Neef 2015).

- (6) a. German: [s], [n], etc.
- |   |               |
|---|---------------|
| Arbeits <u>z <td style="width: 70%;">‘workroom’</td> </u> | ‘workroom’    |
| Blumen <u>n</u> stangel                                   | ‘flower stem’ |
- b. Italian: [o] and [i]
- |                      |                 |
|----------------------|-----------------|
| musico <u>l</u> ogo  | ‘musicologist’  |
| colori <u>f</u> icio | ‘paint factory’ |

True LEs are not inflectional morphemes (e.g., Modern Greek (Ralli 2007, 2009); German (Neef 2009, 2015); Dutch (Don 2009)). Regardless of the historical source of LEs, they are semantically empty (Bauer 2009: 346). LEs are sometimes described as allomorphic variations of the first element of the compound (Neef 2009, 2015).

The comparison among the semantically empty inserted elements discussed above leads us to conclude that the factors that condition these inserted elements lie along a cline from phonological to morphological. We predict that this cline follows a diachronically directional grammaticalization. One advantage of this typology is that it allows us to include semantically empty elements larger than a single unit whose presence is phonologically motivated, but whose realization is not.

## References

- BAUER, L. (2009). Typology of compounds. In *The Oxford Handbook of Compounding*.
- BONET, E. & F. TORRES-TAMARIT. (2009). Allomorphy in pre-clitic imperatives in Formenteran Catalan: An output-based analysis. *Romance linguistics*: 337-352.
- CARDINALETTI, A. & L. REPETTI. (2008). The Phonology and Syntax of Preverbal and Postverbal Subject Clitics in Northern Italian Dialects. *Linguistic Inquiry* 39: 523-563.
- DE LACY, P. (2002). *The Formal Expression of Markedness*. Ph.D. dissertation, University of Massachusetts Amherst.
- DIFABIO, E. (1990). *The morphology of the verbal infix/-isk in Italian and in Romance*. PhD dissertation, Harvard University.
- DRESSLER, W.U. & L.M. BARBARESI. (1994). *Morphopragmatics: Diminutives and intensifiers in Italian, German, and other languages* (Vol. 76). Walter de Gruyter.
- HOBERMAN, R.D. & M. ARONOFF. (2003). The verbal morphology of Maltese. *Language Acquisition And Language Disorders*, 28: 61-78.
- KITTO, C. & P. DE LACY. (1999). Correspondence and epenthetic quality. In Catherine Kitto & Carolyn Smallwood (eds.), *Proceedings of AFLA VI*. Holland: Holland Academic Graphics.
- KOIKE, S. S. (1996). Sequential voicing in Japanese and adjacency. *Proceedings of ConSOLE IV*, 143-50.
- LABRUNE, L. (2009). Featural linking elements in Basque compounds. *Morphology* 24: 377-405.
- LIEBER, R. & P. STEKAUER (Eds.). (2009). *The Oxford handbook of compounding*. Oxford University Press.

- MICHELSON, K. (1989). Invisibility: vowels without a timing slot in Mohawk. In Donna B. Gerdts & Karin Michelson (Eds.), *Theoretical Perspectives on Native American Languages*. SUNY Press.
- NEEF, M. (2015). The status of so-called linking elements in German: Arguments in favor of a non-functional analysis. *Word Structure*, 8(1): 29-52.
- NEEF, M., (2009). IE, Germanic: German. In *The Oxford Handbook of Compounding*.
- NÜBLING, D. & R. SZCZEPANIAK. 2008. On the way from morphology to phonology: German linking elements and the role of the phonological word. *Morphology*, 18(1): 1-25.
- PRATI, A. (1942). Antisuffissi. *L'Italia dialettale* 18: 75-166.
- RALLI, A. (2009). IE, Hellenic: Modern Greek. In *The Oxford Handbook of Compounding*.
- RAWLINS, K. (2006). *Stress and epenthesis in Mohawk*. Ms., UC Santa Cruz.
- REPETTI, L. Ms. The Definite Article in Italian.

## Permutations in the Lexicon: an Automated Approach

**François NEMO,**  
*Université d'Orléans,*  
francois.nemo@univ-orleans.fr

**Flora BADIN,**  
*Université d'Orléans,*  
flora.badin@univ-orleans.fr

### 1. Introduction

Our starting point will be that even though the existence of permutation is not new in morphology, it has been studied mostly at word level as a morphophonological phenomenon but so under-tested in the overall lexicon (but for *verlan*) that permutative relationships have generally been overlooked. Even the *form/morph* relationship is absent from morphological literature altogether, including dedicated databases (notably the Ohio State University Metathesis database). As a result, no measurement has been made of the frequency of permutative pairs such as the pair (*star*, *astr-onomy*) and their existence has not been systematically tested. Our main goal in this respect will be to show that looking for semantically relevant permutative relationships in the lexicon can be extensively automated and shows that such relationships are not only frequent but explanatory of many apparently unpredictable listemes (in the sense of Di Sciullo & Williams, 1987).

#### 1.1. Permutations at word-level

Permutations can occur at word-level, with a single word having different forms or realizations in which at least two phonemes are permuting. In some cases, such as the pronunciation of French *aéroport* as *aréoport*, it concerns only the realization of the form and only one of the two forms can be considered as lexicalized. In some rare cases, permutation is associated with inflection as is arguably the case for French *œil* (eye), whose plural is *yeux* (eyes).

Much more frequently, permutations are associated with lexical allomorphy, i.e. with the coexistence in a single grammatical position and with an identical meaning of two mutually exclusive forms (complementary distribution), as is the case of French *vér-ité* (truth) and *vrai* (true) in which the form *vér-* occurs in the adjectival slot of a [Adj-ité]<sub>N</sub> construction, with the same meaning as in *vrai*.

All other cases, in which the criteria of grammatical or semantic identity are not met, and for which testing complementary distribution is hazardous, do not concern individual words but relations between words and thus require a completely different methodology. We shall hence refer to *lexical allomorphy* in the first case and to *morphemic polymorphy* for the last case, which includes for instance the relationship between French adjectives *rude* (hard) and *dur* (hard) and nouns *tor-sion* and *rot-ation* in French and English which share a **morphemic meaning** but have no identical meaning.

Not conflating lexical allomorphy and morphemic polymorphy is essential, since lexical allomorphy concerns only units in complementary distribution with the **same grammar**, the

**same meaning** and different but most often related sound shapes, while morphemic polymorphy concerns units which are not in complementary distribution, have potentially distinct grammar, partly identical and partly distinct meanings and related sound shapes: lexical allomorphy will thus concern the *-iple* alternate of the *-uple* suffix, whereas morphemic polymorphy will concern the relationship between the *-uple* suffix and the *plu-* and *pul* components of words like *pluriel* (plural) or *plupart* (most of, majority) or *pulluler* (proliferate, multiply), or between French adjectives *rude* and *dur*.

## 1.2. Permutations in the lexicon

Morphemic permutation as a morphosemantic relationship, if tested, can be shown to be very frequent in many languages, including languages, such as English, French or Spanish, in which they have not been documented yet. Proving such morphosemantic relationships can be done using seven different types of semantic tests, some of which can be fully automated, given that lexicographical glosses massively use, whenever possible, one of the permutations in order to describe others, or use the same terms to describe both, as shown by the American Heritage glosses of English *lock* and *occlude* (both of which using the permutative form *close*):

### lock (lök)

1. A device operated by a key, combination, or keycard and used, as on a door, for holding, **closing**, or securing.
2. A section of a waterway, such as a canal, **closed** off with gates, in which vessels in transit are raised or lowered by raising or lowering the water level of that section.

### oc·clude

- v.*tr*
1. To cause to become **closed**; obstruct: *occlude an artery*.
  2. To prevent the passage of: *occlude light*; *occlude the flow of blood*.
  3. *Chemistry* To absorb or adsorb and retain (a substance).
  4. To force (air) upward from the earth's surface, as when a cold front overtakes and undercuts a warm front.
  5. To bring together (the upper and lower teeth) in proper alignment for chewing.
- v.*intr.*
- To **close** so that the cusps fit together. Used of the teeth of the upper and lower jaws.

[Latin *occlūdere* : *ob-*, intensive pref.; see *ob-* + *cludere*, to **close**.]

## 2. Sharing a morphemic layer of interpretation: semantic and methodological issues

Speaking of permutative relationships in the lexicon does not entail that one word was created from the other, or vice versa, it only means that these two or more words can be proven to share an identifiable semantic layer of morphemic presupposition. The problem hence is not to know whether for instance the verb to *mock*, the adjective *com-ic* and the noun *com-edy* have the same meaning but to know if they all presuppose a certain type of situation, for instance a situation in which someone or something is associated with other people laughing, either unwillingly or on purpose. Morphemic layers have proven to always impose a single situational context, leaving to other layers the determination of what exactly take place in that context. What is important in this respect is to realize that as far as semantics is concerned, there is actually no way to account for the polysemy of *to mock* without accounting for the

polysemy of *comedy* or *comic*, no way to account for the diversity of situations in which the verb can be used without accounting for the diversity of situations in which *comic* or *comedy* is used.

Permutative interlexical relationships are thus established whenever it proves impossible to describe what goes on in the situation associated with one (use of a) word without introducing something which is also necessary in the description of the situation associated with another word (with a permutative form). The simplest case are words such as French *quarte* or *quarteron* (quadroon) and *quatre* (four) or *tré-*, *ter*, *tern-*, *tertio*, *tierce*, *tierce*, *tiers*, *tri-*, etc. which have different meanings that all includes at some point having to mention the existence of three X or four X, exactly as *habl-ar* and *blabla* in Spanish cannot be described without pointing to a single situational element, and exactly as French prefix *in-* (*un-*), conjunction *ni* (neither) and verb *ni-er* (deny) have distinct meanings but share the necessity to introduce at some point the fact that something is not the case.

The advantage of morphemic presupposition (when present) is that it is not possible for lexicographers to describe the situations associated with (various uses of the) lexemes without having at some point to mention the situational elements at stake, thus allowing to confirm the permutative relationship in much less intuitive cases than those presented so far. Lexicographers have thus typically produced glosses using a similar or identical metalanguage and keywords, allowing for an automated detection of the elements at stake. In the most straightforward cases, the permutative form will be present in the gloss, or the two permutative forms will be described using the same hyperonym/term or the two permutative forms will be described using synonyms, etc.

Complementing such direct lexicographical tests, translation tests and synonymy tests can also be automated, and confirmed by other available semantic tests, all of which converge when the interlexical relationship does exist and can further undergo psycholinguistic tests.

### **3. Permutative interlexical relationship, morphemic archiphonemes and expansions**

Permutation does not always occur alone and is often combined with other modifications of form. A good illustration of such cases is observed when a phonological feature is neutralized, i.e. when an archiphoneme in a morpheme is found with both its voiced and non-voiced realizations (for a consonant) or nasal and oral realization (for a vowel).

An illustration of the latter is French permutative pair *blanc* and *alb-* (*alb-inos*, *alb-umine*, *alb-ion*), both meaning white, in which the oral/nasal feature is neutralized.

Illustrations of the former are the English pair *stop* and *obst-* (*obst-acle*, *obst-ruction*) - in which the permutation is associated with a neutralization of the voiced/non voiced feature – and French pair of adjectives *complet* (complete) and *comble* (full, packed), used for instance to say that a show is *complet* (no places available left) or that the theater is *comble* (fully packed, no places available left).

Permutations can also be associated with expansion, i.e. addition of an extra phoneme to the sound form (which ultimately can be analyzed as a minimal form of cranberry morpheme) as in French *tour* and *tourn-er* or English *ear* and *hear*. Illustrations of expanded permutations are the relationship between *star* and *aster-oid* in English or *rot-ation* and *torn-ade* in French. Expansions are important because they imply that permutative pairs may very often vary in size, as is the case of French *se ru-er* (to rush) or *une ruée* (e.g. a gold rush) and adjective *urgent*, which all share a morphemic presupposition that something move or act with speed or haste.

Finally, considering that the various types of modifications can co-occur in association with permutations, it follows that permutative relationships do not relate only pairs but lexical networks (into morphemic webs).

### **4. Permutative interlexical relationship and diachrony**

When semantically relevant permutative relationships are established in synchronic terms, it becomes possible to question the diachronic status of these relationships. As for the French and English examples provided so far, what can be observed is that there are cases in which: i) the same permutative relationship is arguably already present in Latin; ii) the origin of bimorphy is pre-romance (e.g. *morph/form*); iii) the permutative relationship appears to emerge at a given time with the resemanitization of semantically unrelated pre-existing units; iv) some words (and not only meanings) appears to emerge out of the blue (e.g. French *pingre* meaning *scrooge* or *skinflint*); v) a new meaning in one form/word leads to a new meaning of the permutative form (e.g. French *amorphe*).

As for the third case, it is important to note that the semantically unexpected emergence appears to be directly explainable by the permutative relationship. This implies that what would typically be considered as a semantic drift leading to the listing of such words must in fact be analyzed as a (predictable) formal drift with no semantic drift (and no listing). If initially the standard analysis is H1:

H1. Origin of B is A, with B keeping the form of A but not its meaning;  
it appears that H1, when confronted with H2:

H2 Origin of B is C, with B keeping the meaning of C, but changing its linear form.  
must be reformulated as H1b:

H1b. Origin of B is A with B keeping the form of A but taking incidentally the  
meaning of C;

Given that H1 and H1b are making claims that are either unjustified (when it comes to ruling out formal permutative drift), untestable and hence unfalsifiable (when it comes to the postulation of a semantic drift/jump) or improbable (when it comes to postulating that an unpredicted change of meaning will accidentally lead to the meaning of a word with a permutative sound form), it is clear that the classical analysis must be ruled out in such cases.

## 5. Automating the search for permutative interlexical relationships in lexicons

Implementing the search for semantically relevant permutative relationships requires only to have access to fully electronic versions of existing dictionaries such as TLFi, the dictionary of the Académie française, the American Heritage dictionary, etc. and to program specific scripts. They must include phonetic forms, serving as a base to a phonetization of all the glosses, and can be complemented by dictionary of synonyms, automated translation searches on bilingual dictionaries (e.g. wordreference), etc.. Various strategies can be implemented and will be described in detail.

Meaning-based strategies consist in scanning a set of words for which domain-related keywords have been used (in the lexicographical glosses) and testing the existence of permutations among them.

Sign-based and listeme-based strategies consist in producing a set of permutative alternates for a form, to extract all words associated to these forms, and to test a spectrum of possible semantic relationship between them. Starting from a word such as *mark*, the form *gram* for instance is produced and all words with a *gram* segment, such as *monogram* for instance, have their gloss (i.e. *a design composed of one or more letters, typically the initials of a name, used as an identifying mark*) tested.

The third approach consists in the systematic testing of a given type of permutations or phonological distance in the whole lexicon. It allows measuring the frequency of each type of permutations and of permutative relationships in the lexicon.

Automated tests can be said to be reliable because they do not produce fake-positive and simply make explicit what otherwise would remain unnoticed. They can however produce fake-negative by missing untested semantic relationships. Complementing them by human

expertise, by proposing formal pairs (e.g. *proximity* and *promiscuity*) to annotators, can be automated.

The whole implementation and its results (mostly based on the TLFi) will be presented.

## 6. Permutative relationship and morphological theory

The idea that signs had a forcefully linear *signifiant* has been considered as axiomatic since at least Saussure. Considering that permutations and permutative relationships can be proven to exist, it is not possible for morphological theory to postulate such mechanisms only as local ad hoc hypotheses, and it is compulsory to test them systematically and exhaustively on entire lexicons. Fortunately, lexicographical glosses, among others data, allow for an automated testing of such permutative relationships, free of semantic or graphical intuition and fear of being ridiculed. Such tests show that permutative relationships are very frequent and can allow to predict parts of the lexicon which until now were simply listed. Given that the only recognized interlexical relationships in standard morphology have been the relations of inclusion of one word in another, i.e. derivation and composition, the existence of semantically related words sharing only a morphemic presupposition enlarges our view of what lexicons are and how they are organized. It can be considered as problematic only if a monist attitude is adopted; the coexistence of mechanisms based on lexical inheritance (derivation, composition) and morphemic inheritance is in fact completely unproblematic for the same reason that gravitation and electromagnetic forces are not contradictory in physics. The provable existence of permutative morphemic relationships in the lexicon allows morphological theory to account for parts of the lexicons which have remained out of reach due to the axiomatization of the hypothesis that the *signifiant* was linear. As importantly, it allows to map entire lexicons and to question atomistic representations of the lexicon according to which most of the lexicon is to be learned item by item.

## References

- BLEVINS, J. & A. GARRETT (1998). The origins of consonant-vowel metathesis. *Language* 74(3):508-555.
- BYBEE, J. L. (1985). *Morphology. A Study of the Relation between Meaning and Form*, John Benjamins Publishing Company, Amsterdam.
- CREUTZ, M. & K. LAGUS (2002). Unsupervised discovery of morphemes. In *Proceedings of the ACL Workshop on Morphological and Phonological Learning*, pp. 21–30. ACL, Philadelphia, Penn.
- DURAND, J. (2009). On the scope of linguistics: data, intuitions, corpora. In Y. Kawaguchi, M. Minegishi, et J. Durand, éds., *Corpus and Variation in Linguistic Description and Language Education*, 25–52. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- FUCHS, C & B. VICTORRI (1994). *Continuity in linguistic semantic*. Amsterdam/Philadelphia: Benjamins.
- GAUME, B., K. DUVIGNEAU, O. GASQUET & M.D. GINESTE (2002). Forms of meaning, meaning of forms. *Journal of Experimental and Theoretical Artificial Intelligence*, 14(1): 61–74.
- GOLDSMITH, J. (2001). Unsupervised learning of the morphology of natural language. *Computational Linguistics*, 27(2):153–198.
- HATHOUT, N. (1996). Pour la construction d'une base de connaissances lexicologiques à partir du *Trésor de la Langue Française*. Les marqueurs superficiels dans les définitions spécialisées. *Cahiers de Lexicologie*, 68(1):137–173.
- HATHOUT, N. (2002). From WordNet to CELEX: Acquiring morphological links from dictionaries of synonyms. In *Proceedings of the Third International Conference on*

- Language Resources and Evaluation*, pp. 1478–1484. ELRA, Las Palmas de Gran Canaria.
- HOCK, Hans Henrich. (1985). Regular metathesis. *Linguistics* 23. 529-546.
- HUME, E. (1998). Metathesis in phonological theory: The case of Leti. *Lingua*. 104. 147-186.
- NEMO, F. (2005). Morphemes and Lexemes versus Morphemes or Lexemes. In G.Booij & alii (eds.)*Morphology and Linguistic Typology. Siculorum Gymnasium.*, University of Catania. 57, 253-272.
- PLOUX, S. & B. VICTORRI (1998). Constructions d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement automatique des langues*, 39(1):161–182.
- ULTAN, R. (1978). A Typological View of Metathesis. In J. Greenberg (ed.), *Universals of Human Language* vol. 2. Stanford, CA: Stanford University Press. 368-399.

## Stem alternation at the morphology-phonology interface: examples from Skolt Saami

**Flore PICARD,**  
*Université Paris-Sorbonne*  
florepicard@yahoo.fr

Modern Skolt Saami, a Finno-Ugric language spoken in central Lappland, makes use of significant stem-internal alternations in its inflectional morphology. The Uralist tradition tends to look at Saami inflection only in diachrony or as purely phonological processes, which led to Skolt Saami being perceived as "extremely complex" (Aikio 2008) with no less than 55 noun types and 29 verb types according to Sammallahti & Mosnikoff (1991). Feist (2010) in his dissertation reduces this number to a more reasonable 4 verb classes (10 subgroups) and 12 noun classes, but the language still presents a real challenge to any attempt at morphological reduction.

To try to make sense of this complex morphology, we use Paradigm Function Morphology, a highly formal Word-and-Paradigm model based on realization and inferential morphology, developed by Gregory Stump (Stump 2001a Stewart & Stump 2007a Bonami & Stump 2014). It presumes that "a word's inflectional markings are determined by the morphosyntactic properties which it carries" (Stewart & Stump 2007, 387) and that word forms are deduced from roots by means of rules associating morphological processes with these morphosyntactic properties. It can thus handle non-canonical phenomena such as stem alternations and empty affix slots.

It is particularly well adapted for the analysis of highly inflectional languages, to the point where it is usually admitted that it does not face many challenges when modeling these languages. We will show that, despite that, Skolt Saami forces us to make some hard decisions in regards to how to formalize its inflection, and especially the link between stem alternations and morphophonological processes.

Table 1 is a paradigm for the verb *kuulləd*, "to wash", in Skolt Saami.

		Present		Past	
		Singular	Plural	Singular	Plural
		1st	2nd	3rd	4th
		<b>kuul-am</b>	<b>kuull-əp</b>	<b>ku'l:l-em</b>	kuul-im
		kuul-ak	kuull-ve'ted	ku'l:l-ič	kuul-id
		<b>kooll</b>	<b>ko'l:l-e</b>	kuul-i	ku'l:l-e
		kuul-ət		ku'l:l-eʃ	

**Table 1** — Present and past forms of the verb *kuulləd* "to wash" (Feist 2010)

Stem alternations in this paradigm are made clear by bold characters: there are five different stems, combining four alternation processes:

- Vowel length (kuul- vs ku'l:l-),
- Consonant gradation (kuul-, kuull, ku'l:l-),

- Vowel quality or ablaut (kuull- vs kooll-),
- Palatalization of the second consonant (ku'l:l-)

In this paradigm (Feist 2010's Class 1A), palatalization as well as vowel reduction appear to be linked to consonant length: they only appear when the consonant is overlong. This does not hold true for all paradigms: some verbs are not affected by any of these phenomena, or only by one of them, typically consonant gradation, and some have palatalized stems by default, only losing their palatalization in a few forms. For this reason, we cannot consider palatalization to be a result of either vowel lengthening or consonant strengthening, or the other way around.

Admitting, as it is likely the case, that the vowel lengthening is a direct result of the presence of an overlong consonant and is phonological, we still have three separate alternations of morphological nature making up five different stems. All three alternations arise from old morphophonological processes: consonant gradation is still fully morphophonological in languages like Finnish, triggered by the openness of the syllable following the affected consonant (Sammallahti 1998a), the ablaut originally comes from metaphony, still present in the western Saami languages (Wilbur 2014), and palatalization was once linked to a palatal vowel in the following syllable (Feist 2010). The evolution of the personal affixes have made them lose their original purpose, and they have become fixed phenomena inside the paradigms, although they are still productive to some extent.

Paradigm Function Morphology (PFM) separates in its sets of rules between Rules of Stem Choice (RSC), Rules of Exponence (RE) for affixes and Morphophonological Metageneralizations (MPM). The MPMs are closely tied to the realizations, being the last to apply to a form, while the RSCs (we will not deal with REs here) are closer to the lexeme itself and further from the realizations. MPMs are metageneralizations: they cannot use a set of morphosyntactic properties as one of their argument, but only set a context where they apply. The rule linking the shortening of the vowel to the third consonantal length is a MPM as in (1).

(1) Long stem internal vowel is shortened before an overlong consonant.

The other processes, being morphological in nature, cannot be formalized by MPMs. This means that they have to be taken into account at the level of the Rules of Stem Choice. Example (2) is a formalization of the five stems by RSCs. Each rule specifies the value of Stem for a set of morphosyntactic properties  $\sigma v$ . The rules apply according to Pān.ini's principle: the narrowest rule applies first, and rule (2 a) serves the default stem, for all the remaining forms.

- (2) (a) Stem(< kuulləd ,  $\sigma v : \{ \}$ ) = <kuul,  $\sigma v$ >  
 (b) Stem(< kuulləd ,  $\sigma v : \{ \text{TNS:prs PERS:1/2 NUM:pl} \}$ ) = <kuull,  $\sigma v$ >  
 (c) Stem(< kuulləd ,  $\sigma v : \{ \text{TNS:pst } \{ \text{PERS:3 NUM:sg} \} \{ \text{PERS:1/2 NUM:pl} \} \}$ ) = <ku'l:l,  $\sigma v$ >  
 (d) Stem(< kuulləd ,  $\sigma v : \{ \text{TNS:prs PERS:3 NUM:sg} \}$ ) = <kooll,  $\sigma v$ >  
 (e) Stem(< kuulləd ,  $\sigma v : \{ \text{TNS:prs PERS:3 NUM:pl} \}$ ) = <ko'l:l,  $\sigma v$ >

There is no theoretical limit to the number of stems a paradigm can contain, but in most languages it does not go beyond two or three. In addition, this paradigmatic complexity presented by Skolt Saami causes another problem : can these alternations truly be put on the same level?

Consonant gradation is the result of a phonological process of strengthening in the Saami family which has become morphological early on, resulting in an identical distribution over all affected verbs across inflectional classes (Sammallahti 1998b). In contrast, apophony is neither present in all Skolt Saami verbs nor always affecting the same set of cells. It is thus closer to the individual inflectional classes (in its distribution) and the cells' realizations (in the actual phonological change). Palatalization, as we said before, is at least as much linked to

individual verbs as to the inflectional classes they belong to. For this reason, it does not seem fully appropriate to analyze those three processes on the same level, thus implicating that they are interconnected and morphological equals.

One of the solution we can propose for this problem is to bring the consonant gradation to before even the stem selection step by adding a GRAD feature with the values of lenis, medium and fortis associated to a general template such as table 2, valid for the whole inflectional system.

	Present		Past	
	Singular	Plural	Singular	Plural
1st	<i>Lenis</i>	<i>Medium</i>	<i>Fortis</i>	<i>Lenis</i>
2nd	<i>Lenis</i>	<i>Medium</i>	<i>Fortis</i>	<i>Lenis</i>
3rd	<i>Medium</i>	<i>Fortis</i>	<i>Lenis</i>	<i>Fortis</i>
4th	<i>Lenis</i>		<i>Fortis</i>	

**Table 2 —** Template for the GRAD feature

Through a set of examples like the one presented above, we wish to look at the options PFM provides us to formalize the stem alternations in Skolt Saami verbal inflection, at the interface between morphology and phonology, and where its limits lay, and to propose a comprehensive analysis of the processes mentioned above across inflectional classes.

### Abbreviations

GRAD	Grade
MPM	Morphophonological Metageneralizations
NUM	Number
PERS	Person
PFM	Paradigm Function Morphology
pl	Plural
prs	Present
pst	Past
RE	Rules of Exponence
RSC	Rules of Stem Choice
sg	Singular
TNS	Tense

### References

- AIKIO, Ante. (2008). Extreme Morphophonological Complexity As A Challenge To Grammatical Description – The Case Of Skolt Saami (Presentation). In Cognitive and Functional Perspectives on Dynamic Tendencies in Language, Tartu.
- BONAMI, Olivier & Gregory STUMP. (2014). Paradigm Function Morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, Cambridge: Cambridge University Press.
- FEIST, Timothy Richard. (2010). A Grammar of Skolt Saami. Manchester: University of Manchester dissertation.
- SAMMALLAHTI, Pekka. (1998a). The Saami languages: An introduction. Kárášjohka: Davvi Girji.
- SAMMALLAHTI, Pekka. (1998b). Saamic. In Daniel Mario Abondolo (ed.), *The Uralic Languages*, Taylor & Francis.

- SAMMALLAHTI, Pekka & Jouni MOSNIKOFF. (1991). *Suomi-koltansaame sanakirja =: Lää'dd-sää'm sää'nnke'rjj.* Ohcejohka: Girjegiisá oy.
- STEWART, Thomas & Gregory STUMP. (2007). Paradigm Function Morphology and the Morphology-Syntax Interface. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 383–421. Oxford: Oxford University Press.
- STUMP, Gregory T. (2001). Inflectional morphology: a theory of paradigm structure (*Cambridge studies in linguistics* 93). Cambridge; New York: Cambridge University Press.
- WILBUR, Joshua. (2014). *A grammar of Pite Saami* (Studies in Diversity Linguistics 5). Berlin language science press edn. <http://langsci-press>

## From Manner to Result Verbs: the Evolution of *-izar* Complex Verbs in Old Spanish

**Isabel Pujol PAYET,**  
*Universitat de Girona (Spain)*  
 isabel.pujol@udg.edu

Among current Spanish complex verbs, verbal formations with *-izar* are one of the most widely used patterns. Although synchronic studies have shown that they are not a homogeneous class (since some significant semantic, syntactic and aspectual differences can be pointed out within this group), it is true that most of them are causative verbs which involve a change of state. Hence they have a telic aspect (e.g. *carbonizar* ‘to carbonize’, *criminalizar* ‘to criminalize’, *fluorizar* ‘to fluoridate’, *hospitalizar* ‘to hospitalize’), —Pena (1993), Batiukova (2008 and 2016), NGRAE (2009) and Fábregas (2015), among others. This situation contrasts with verbal formations with *-ear*, which are usually intransitive and atelic (e.g. *bobear* ‘to fool about, to do silly things’, *cabecear* ‘to nod’, ‘to nod off’, ‘to shake one’s head’, ‘to head’, *bordear* ‘to border’, *martillear* ‘to hammer’, *parrandear* ‘to go on a binge’), —Pena (1993), Martín García (2007), NGRAE (2009).

Both verbs with *-izar* and *-ear* are attested from Early Spanish. The two suffixes, the learned *-izar* and the popular *-ear*, come from Latin *-izare* (with the *-issare* and *-idiare* variants) adapted from the Greek verbal suffix *-ίζειν*. As Cockburn (2012) states, “the Latin suffix was recurrent in medical and political language (e.g. *cauterizare* ‘to cauterise’, *tyrannizare* ‘to act as a tyrant’) and is found in abundance in the Latin of the Christians (*baptizare* ‘to baptise’, *prophetizare* ‘to prophesy’). It then continued to develop in Medieval Latin and in the Romance Languages, where it is still an extremely productive verbal suffix.” According to this autor, the majority of Latin verbs whith *-izare* are intransitive o transitive which specify a manner of carring out an action. Moreover, some of these transitive verbs can also shown a result reading (e.g. *pulverizare* ‘to make dust’ and ‘to reduce to powder’).

The aim of this paper is to analyze the aspectual change of the verbs with *-izar* in the history of Spanish, from manner verbs to result verbs. For this reason the analysis focuses on the Old Spanish verbs. If we make a comparison between the most frequent verbs of the thirteenth century and those of the fifteenth (according to the corpus of Mark Davies), we can observe how in the first stage of Spanish most of the verbs with *-izar* are inherited from the Late Latin (eg. *bautizar*, *profetizar*, *escandalizar*, *catequizar*, *martirizar*) versus some few neologisms (*temporizar*, *subtilizar*). In contrast, as the following table describes, the fifteenth century shows a higher productivity of new forms of adjectival base, as well as some parasyntetics ones.

Verbs attested in Latin (Cockburn 2012)	New forms in Old Spanish
bautizar	autorizar
profetizar	sutilizar
cauterizar	moralizar
polvorizar	atemorizar
gargarizar	descuartizar
escandalizar	temporizar
tiranizar	granizar
canonizar	vulgarizar
martirizar	sincopizar
clisterizar	solemnizar
silogizar	gigantizar
tesorizar	esterilizar
sabatizar	
judaizar	
evangelizar	

**Table 1** — Verbs with *-izar* in the 15th c.

In this research it will argue that the change from manner verbs with *-izar* to result verbs is motivated by several factors: a) the transitive structures of these verbs; b) the variety of bases that can form this type of verbs, in particular, the appearance of neologisms of internal creation with an adjectival base throughout the fifteenth century; c) the regression of the pattern of result verbs with *-ecer*—Malkiel (1941), Batllori and Pujol (2012), Batllori (2015), among others—; and d) the consolidation of verbs with *-ear* as manner verbs.

## References

- BATIUKOVA, Olga (2008). “Morfología: del léxico a la sintaxis oracional”. In Antonio Moreno Sandoval (coord.), *Actas del VIII Congreso de Lingüística General*. Madrid: Universidad Autónoma de Madrid.
- BATIUKOVA, Olga (2016). “Restricciones léxico-semánticas y mecanismos composicionales en la morfología derivativa: el caso de *-iza(r)*”. In Cristina Buenafuentes de la Mata, Gloria Clavería Nadal and Isabel Pujol Payet (eds.), *Cuestiones de morfología léxica*. Madrid / Frankfurt: Iberoamericana / Vervuert, 101-165.
- BATLLORI DILLET, Montserrat (2015). “La parasíntesis a la luz de los datos históricos de los verbos en *a-* y *-escer*”. In José María García Martín (ed.), *Actas del IX Congreso Internacional de Historia de la Lengua Española. Cádiz, 10-14 de septiembre de 2012*. Madrid / Frankfurt: Iberoamericana / Vervuert, 617-638.
- BATLLORI DILLET, Montserrat and Isabel PUJOL PAYET (2012). “El prefijo *a-* en la formación de derivados verbales”. In Emilio Montero (ed.), *Actas del VIII Congreso Internacional de Historia de la Lengua Española. Santiago de Compostela, 14-18 de septiembre de 2009*. Santiago de Compostela: Meubook-Unidixital, S.L., 659-671.
- COCKBURN, Olivia Claire (2012). *Los verbos latinos en -izare (-issare, -idiare). Adaptación, uso y desarrollo del morfema griego -ίζειν en el latín antiguo. Resultados lexémáticos en iberorromance*. Madrid: Universidad Autónoma de Madrid. PhD thesis.
- DAVIES, Mark. *Corpus del Español* <<http://www.corpusdelespanol.org>>
- FÁBREGAS, Antonio (2015). “Sobre el sufijo *-iza(r)* y sus propiedades internas”. *Lengua y Habla* 19, 166-187.

- MALKIEL, Yakov (1941). “*Atristar – entristar*: adjectival verbs in Spanish, Portuguese and Catalan”. *Studies in Philology*. 38, 429-461.
- MARTÍN GARCÍA, Josefa (2007). “Verbos denominales en -ear: caracterización léxico-sintáctica”. *Revista Española de Lingüística* 37, 279-310.
- NGRAE = Real Academia Española (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- PENA, Jesús (1993). “La formación de verbos en español: la sufijación verbal”. In Soledad Varela (ed.), *La formación de palabras*. Madrid: Taurus, 217-281.

## Does French have verbal-nexus Noun+Noun compounds? A corpus-based study.

**Jan RADIMSKÝ,**  
*University of South Bohemia, České Budějovice (CZ)*  
 radimsky@ff.jcu.cz

In comparison to Germanic languages, such as English or German, Noun+Noun (N-N) compounds in Romance languages still represent a peripheral phenomenon, since complex naming units are rather formed by syntactic rules following the Noun+Preposition+Noun pattern. Nevertheless, N-N compounding in Romance languages can be characterized as an emerging pattern which has been developing in a dynamic way especially in the past 50 years (cf. Dardano, 2009:228). This makes the study of N-N structures particularly interesting in that the data are more likely to reflect actual regularities and tendencies with little interference from anomalies resulting from the diachronic evolution of the language. Romance N-N structures lie on the edge between morphology (i.e., “compounding”) and syntax (“N-N noun phrase”), and even the most recent sources (such as Arnaud, 2015) point out that in spite of some descriptive knowledge accumulated recently, their status is still far from being clear. In this paper, Romance N-N structures (such as Fr. *exposition photos* – “photography exhibition”) will be treated as compounds (cf. Gaeta-Ricca, 2009), and the implicit relationship between the two nouns will be described following the classification of the linguists of the Bologna group (Scalise-Bisetto, 2009).

In the classification of Scalise-Bisetto (2009), verbal-nexus (VNx) compounds are subordinate structures, such as *book seller* or *pickpocket*, featuring a predicate-argument relationship between a (verbal or deverbal) head and a non-head element. Roughly speaking, VNx compounds correspond to the group traditionally labelled in English as “synthetic compounds”. The conception by Scalise-Bisetto (2009), however, is broader in the sense that even compounds in which the non-head element corresponds to a complement or an adjunct of the deverbal head can be considered as verbal-nexus as, for instance, *tree eater* interpreted as “a person eating on a tree”. Within the Romance N-N compounds, VNx compounds are those featuring a deverbal noun in the leftmost position and its argument/complement/adjunct in the rightmost position, such as in It. [trasporto]<sub>N</sub> [latte]<sub>N</sub> (“milk transport”), *attacco hacker* (“hacker attack”), and *uscita autostrada* (“highway exit”).

Although in modern French N-N compounding as such seems to represent a productive pattern (Villoing, 2012:41) characterized by an exponential growth (Noailly, 1990:12-13), previous studies did not come to agreement on the question whether, within the N-N pattern, French also has VNx compounds and, if that were the case, what the relationship between the deverbal head and the non-head would look like. Villoing (2012:50), for instance, argues that a process-argument relation is impossible within French N-N compounds. On the other hand, Noailly (1990:119-121) claims that French VNx N-N compounds are grammatical. She suggests that the argument (N2) in these compounds is generally a proper noun that corresponds to the subject of the verb underlying the deverbal head, as in *administration Reagan* (“Reagan administration”), *gestion Mitterand* (“Mitterand management”) or *solution*

*Lees* (“Lees’s solution”), but she does not even exclude the possibility for the argument to be a direct object, as in *exposition photos* (“photography exhibition”), *prévention incendie* (“fire prevention”), and *achats vêtements* (“clothes shopping”). Moreover, in corpora it is not impossible to find examples of VN<sub>x</sub> N-N compounds in which the argument corresponds to an indirect object, such as *assurance accident* (“accident insurance”). Given that, it seems that a thorough corpus-based study has to be performed in order to show whether verbal-nexus compounding is an available pattern in contemporary French, and if so, what kind of relationship may be found between the deverbal head and its argument.

This paper aims at presenting results of such a study performed using the FrWac corpus (Baroni et al., 2009), a web corpus whose size reaches  $1,6 \times 10^9$  words. In order to collect representative data, the author has extracted complete frequency lists of pairs of words (W1 and W2) preceded by an article or a preposition, and from these lists he has filtered word forms that may correspond to a French noun, using the GLAFF database (Hathout et al., 2014).<sup>1</sup> In this database of potential binominal structures that contains over 400 000 lemmatized types, potential VN<sub>x</sub> compounds have been filtered using various techniques, such as (1) identification of potential deverbal nouns on the leftmost position using the Verbaction database (Tanguy; Hathout, 2002); or (2) identification of binominals N1-N2 (*exposition photos*) that appear in the corpus also as a N1-de-N2 phrase (*exposition de photos*) (for the rationale of this step see Radimský, 2015:204-205). A mutually exclusive application of these filters made it possible to get a good recall, but a very low precision. Therefore, accurate manual filtering has been essential for identification of VN<sub>x</sub> N-N compounds and their subsequent analysis.

The data show that French VN<sub>x</sub> N-N compounds differ from Italian compounds of the same type both in type frequency and in internal typology. As for quantitative data, only 400 VN<sub>x</sub> compounds could be identified in FrWac, while over 1300 VN<sub>x</sub> compounds have been found in ItWac using the same procedure, which means that VN<sub>x</sub> compounds are much rarer in French (notice that a slight difference in the size of these two corpora – 1,6G vs 1,9G – cannot account for such a difference in type frequency). As far as the internal structure of VN<sub>x</sub> compounds is concerned, data show that Italian VN<sub>x</sub> N-N compounds are much more regular than the French ones. In Italian, 84% of types pertain to one regular scheme (in terms of Construction morphology) where the argument (N2) corresponds to the direct object of the verb underlying the deverbal head noun (N1); in French, on the other hand, different schemes compete and display a balanced proportion of type frequency, namely those where the N2 may be interpreted as direct object, indirect object, subject or adjunct of the verb underlying the deverbal head noun (N1).

## References

- ARNAUD Pierre J.L. (2015), Noun-noun compounds in French. In: Peter Muller et al. (eds.), *Word formation. An international handbook of the Languages of Europe*. Vol. 3. Berlin/Boston, Walter de Gruyter, 2015, 673-687.
- ARNAUD Pierre J. L. (2003), *Les composés timbre-poste*. Lyon, Presses universitaires de Lyon.
- BARONI Marco et al. (2009), The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, vol. 43, no. 3, 209-226.

<sup>1</sup> Any reference to POS tagging of the corpus has been avoided on purpose, since the statistical tagging based on HMM algorithm might provide skewed results for N-N sequences, given that these are not very common in French.

- DARDANO Maurizio (2009), *Costruire parole*. Bologna, Il Mulino.
- GAETA Livio, Davide RICCA (2009), Composita solvantur: Compounds as lexical units or morphological objects? *Rivista di Linguistica*, 22/1, 35-70.
- HATHOUT Nabil, Franck SAJOUS and Basilio CALDERONE (2014), GLÀFF, a Large Versatile French Lexicon. *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC'14), Reykjavík, Iceland, 1007-1012.
- NOAILLY Michèle (1990), *Le substantif épithète*. Paris, PUF.
- RADIMSKÝ Jan (2015), *Noun-Noun Compounds in Italian. A corpus-based study*. České Budějovice, Jihočeská univerzita, edice Epistémé, 387 p.
- SCALISE Sergio, BISETTO Antonietta (2009), The classification of compounds. In: Lieber R., Štekauer P., *The Oxford handbook of compounding*. Oxford university press, 2009, 34-53.
- TANGUY Ludovic and HATHOUT Nabil (2002), Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du Web. In: *Actes de la 9e Conférence Annuelle sur le Traitement Automatique des Langues Naturelles* (TALN-2002), Nancy. ATALA.
- VILLOING, Florence (2012), French compounds. *Probus*, 24/1, 2012, 29-60.

## Structuration du lexique de la Langue des Signes Française (LSF). Focus sur la nature et les fonctions des composants infra-unité lexicale

**Yana SENNIKOVA,**

*Université Paris 8-Université Paris Lumière et CNRS*

yana.sennikova@gmail.com

**Brigitte GARCIA,**

*Université Paris 8-Université Paris Lumière et CNRS*

brigitte.garcia@univ-paris8.fr

Notre présentation a trait à la structuration du lexique de la langue des signes française (LSF) appréhendée notamment à travers les procédés de la création lexicale dans cette langue. Nous défendons l'hypothèse que le composant infra-unité lexicale en LSF est porteur de sens en soi et qu'il est particulièrement impliqué dans la nature des liens forme-sens qui structurent le lexique.

Pour alimenter cette hypothèse, nous présenterons les premiers résultats de notre thèse en cours, dont un enjeu est de montrer que la compréhension de la structure interne de l'unité lexicale (ou signe) permet de mieux comprendre comment le lexique lui-même est structuré.

### **Etat de l'art et options théoriques**

La linguistique des langues des signes (LS) est une discipline jeune : la première analyse linguistique d'une LS a été publiée en 1960 (W.C. Stokoe pour la LS américaine, l'ASL). La question de la structuration du lexique d'une LS n'a commencé à être abordée comme telle que dans les années 1990 (e.g. Padden 1983, 1988, Brennan 1990, Johnston & Schembri 1999, Cuxac 2000, 2004, Fernald & Napoli 2000, Bré-Le Corre 2002).

N'ayant pas de forme écrite instituée, les LS ont été pendant longtemps décrites par la médiation de la forme écrite des langues vocales (LV) environnantes. Ainsi la majorité des dictionnaires élaborés depuis la fin du 18<sup>ème</sup> siècle pour les LS ont proposé des entrées par les mots écrits de la LV du pays. Ceci a eu pour conséquence de fausser notre représentation de la structure du lexique des LS, une correspondance unilatérale ayant été établie entre mot de la LV et « signe » de la LS.

L'association mot-signe a pu déboucher par ailleurs sur l'affirmation d'une pauvreté lexicale des LS (voir notamment Pointurier-Pournin 2014). Il est certes incontestable que le nombre d'unités lexématisques (UL) d'une LS est inférieur à celui des LV environnantes : en moyenne 3000 à 4000 UL pour les LS institutionnelles (Johnston & Schembri 1999, Johnston 2012) tandis que l'*Oxford English Dictionary* contient environ 170000 entrées, par exemple. Notre contribution vise toutefois à mettre en question la réduction du lexique d'une LS à son seul inventaire de lexèmes.

En ce qui concerne précisément les composants de formation des UL, les descriptions courantes des LS leur octroient un statut de phonème (e.g. Brentari 2008, Sandler 2003, Perlmutter 1992), même si nombre d'auteurs s'accordent depuis longtemps sur l'observation que ces composants infra-unités lexicales sont aussi, et régulièrement, porteurs de sens (Frishberg & Gough 2000, Brennan 1990, Johnston & Schembri 1999, Cuxac 1996, 2000,

2004, Fernald & Napoli 2000, Garcia et al 2008, Garcia et al 2013). Dans la perspective théorique qui est la nôtre, l'hypothèse est plutôt que ces composants sont bien, et avant tout, des morphèmes.

Ce cadre théorique est celui du *Modèle Sémiologique* (e.g. Cuxac 1996, 2000, Cuxac & Sallandre 2007, Garcia 2010, Garcia & Sallandre 2014, Sallandre 2014). Il met en évidence la possibilité structurelle, propre aux LS, de recourir à deux modes du dire : le *dire en montrant* et le *dire sans montrer*. Le mode du *dire en montrant* repose sur le recours à des structures dites « structures de transfert », dont le mode de production du sens est directement l'iconicité. Cette manière de dire est centrale et spécifique aux LS. Les structures de transfert mises en évidence permettent de rendre compte des constructions hautement iconiques si nombreuses en discours (les « unités de transfert », UT) dont la valeur est spécifique. Le *dire sans montrer* exploite les unités lexématiques, proches des unités lexicales des LV. L'iconicité peut, ou non, y être présente : elle n'en constitue pas le mode de production du sens, ces UL se caractérisent par leur valeur générique et la conventionnalité de leur sens global.

Les LS ont une morphologie très riche. Ceci a été observé pour plusieurs LS (entre autres, Brennan pour la LS britannique, la BSL, Fernald & Napoli pour l'ASL, Aronoff et al pour l'ASL et la LS israélienne, l'ISL, Johnston & Schembri pour la LS australienne, l'Auslan). Néanmoins, il s'avère délicat de décrire la morphologie des LS par le biais des catégories les plus usuelles de la morphologie des LV, la dérivation et la flexion, dont l'existence même dans les LS est questionnée (voir notamment Fenlon et al 2014, Johnston & Schembri 2007, Liddell 2003, Meir 2012).

Des études récentes sur la LSF (Garcia 2010, Garcia et al 2013) ont dégagé les procédés de création lexicale les plus fréquents : (i) modification d'un (ou 2) composants morphémiques d'une UL préexistante ; (ii) assemblage simultané de composants forme-sens d'UL préexistantes et/ou d'UT ; (iii) lexicalisation d'UT (voir également Johnston & Ferrara 2012, Fusellier-Souza 2006, Cuxac 2004). Ces études mettent en évidence que la morphologie des LS opère crucialement par l'exploitation de la structure compositionnelle de ses unités — des UL ainsi que des UT — et plus précisément de leurs composants morphémiques.

Ces composants morphémiques ancrent les regroupements des UL en *familles de signes*, une famille de signes étant comprise comme un ensemble d'UL unies par un ou plusieurs composants forme-sens (Fernald & Napoli 2000). Alors que divers auteurs ont noté une densité remarquable de ces liens de famille dans le lexique des LS (e.g. Frishberg & Gough 2000, Fernald & Napoli 2000, Cuxac 2000, Bré-Le Corre 2002, Garcia 2010, 2013, Meir 2012), ceci reste très peu étudié dans la littérature.

### Questions de recherches et méthodologie

L'objectif principal de notre thèse étant de décrire la structuration du lexique propre des LS, nous supposons que (i) ce sont les liens forme-sens qui structurent le lexique des LS ; (ii) la coexistence des deux types de structures permettant les deux modes du dire rend superflue une réserve large d'UL dans les LS : grâce à leur structure compositionnelle au moins en partie commune, les UL et les UT rendent possibles et fréquents les procédés de délexicalisation (défigement) qui sont au fondement de la créativité lexicale en LS (Garcia et al 2013, Johnston & Ferrara 2012).

Nous avons élaboré un protocole méthodologique en trois phases dans le but de mettre en évidence la productivité des liens de familles de signes et le rôle joué à cet égard par les composants morphémiques infra-UL. Notre objectif est de montrer que l'économie du lexique de la LSF s'articule autour de ces composants.

Phase 1 : analyse du lexique 'in vitro'. Nous avons analysé le corpus lexicographique

le plus substantiel de la LSF — le dictionnaire IVT qui recense environ 4500 UL (Girod et al 1997). Nous poursuivons deux objectifs ici : (i) évaluer le caractère morphémique du composant infra-UL et (ii) dégager des hypothèses quant à l'organisation des UL en familles de signes constituées à partir des composants morphémiques.

Phase 2 : analyse du lexique ‘in vivo’. Nous avons complété la phase 1 par l’analyse fine de 20 dialogues métalinguistiques entre locuteurs sourds sur les UL récemment émergées en LSF (Corpus Creagest, Garcia & L’Huillier 2011). Notre point de focus ici est de voir comment les locuteurs perçoivent et analysent eux-mêmes la structure des UL récentes et, notamment, s’ils établissent des liens forme-sens entre les UL qu’ils invoquent dans leurs analyses.

Phase 3 : nous appuyant sur les données recueillies dans les phases 1 et 2, nous avons constitué une série de tâches que nous présentons à 3 groupes de locuteurs signeurs sourds (1 groupe pilote ( $n=3$ ) et 2 groupes contrôle ( $2 \times n=4$ ),  $\pm 9h$  d’enregistrement). Ces tâches incluent notamment : a) la présentation d’ensemble d’UL de différentes familles dont nous avons fait l’hypothèse à partir des phases 1 et 2 ; l’objectif est de voir si les signeurs sourds relient les UL par des liens forme-sens et corroborent leurs liens de famille ; b) la présentation d’UL d’autres LS afin de voir si les participants font appel aux liens forme-sens pour accéder au sens des UL inconnues ; c) la présentation d’UL récentes dont la structure mobilise bien les liens forme-sens avec des UL déjà existantes mais sont perçues comme problématiques par les locuteurs du corpus Creagest ([ENSEIGNER EN LSF/LA LSF], [FORMATEUR LSF], etc). L’enjeu est de dégager d’éventuelles contraintes dans la nature des liens qui se tissent dans le lexique.

## Résultats

Nos premières analyses corroborent l’hypothèse proposée par Cuxac (2000) de la nature morphémique du composant infra-UL en LSF : les résultats de la phase 1 montrent qu’un grand nombre des UL répertoriées dans le dictionnaire IVT contient au moins un ou deux composants que l’on peut analyser comme des morphèmes, le même composant associé à la même valeur de sens apparaissant dans plusieurs UL : par exemple, les UL [MEDIA], [RESEAU], [INFORMER], [NOUVELLES], [DIFFUSION], [ENVOYER], [DRAGON] sont réalisées avec le même mouvement d’‘ouverture’— effectué par une ou deux mains - et partagent toutes le sens ‘émission’. Ce lien forme-sens unit ces UL dans la même famille de signes.

Les composants morphémiques des UL déjà existantes sont utilisés de manière dense dans la création de nouvelles UL. Les données des phases 2 et 3 confirment la productivité des trois procédés morphologiques dégagés par Garcia et al (2013) : (i) modification d’un (de deux) composant(s) : l’UL [ILLETTRE] résulte du changement de la configuration de l’UL [LIRE], tandis que le changement du mouvement dans l’UL [MAISON] crée une UL [COMMUNAUTÉ] ; (ii) assemblage des composants : dans l’UL [GENOCIDE] la main dominée reprend la configuration des structures de transfert associée à une « forme mince et allongée », régulièrement utilisée pour référer à un être humain, tandis que la main dominante reprend les composants configuration et mouvement de l’UL [NETTOYER]/[SUPPRIMER] ; (iii) lexicalisation : l’UL [CLE USB] s’est lexicalisée à partir d’une UT ‘insérer une forme plate et longue dans une espace devant le signeur’.

Parallèlement, les données des phases 2 et 3 montrent que les signeurs mobilisent les liens forme-sens préexistants afin d’expliquer spontanément le sens des UL récemment émergées : par exemple, l’UL [THÈSE] est analysée par référence à la configuration ‘C’ et à l’emplacement ‘front’ partagés avec les UL [CONNAISSANCE] ou [ÉTUDIANT] et portant le sens ‘connaissance’, ‘activité intellectuelle’, ainsi que par référence à la configuration commune à l’UL [GROUPE] au sens ‘épaisseur’. De plus, les participants insistent sur

certains liens forme-sens qui ne peuvent pas être « rompus » : l’UL [MEMOIRE] n’est possible qu’à l’emplacement ‘front’ parce que « ... *un mémoire c'est un travail du cerveau et le cerveau se trouve dans la tête* » (phase 3, trad. auteur).

### **Conclusions et Perspectives**

Notre analyse ‘in vitro’ corrobore le statut de morphème des composants infra-UL sur ces bases, la prégnance des liens de familles dans la structuration du lexique de la LSF. Les données des phases 2 et 3 ‘in vivo’ appuient les conclusions de Garcia et al (2013) sur les trois procédés de création lexicale les plus récurrents. Le dégagement par des locuteurs sourds de liens forme-sens afin d'accéder au sens de nouvelles UL atteste la compositionnalité morphémique des UL de la LSF. Ceci nous permet d'avancer deux conclusions importantes : (i) la structuration du lexique en LS ne se limite pas aux unités lexématiques ; les unités lexicales en LS se trouvent à plusieurs niveaux : niveau lexématique mais aussi infra-lexématique ; (ii) l'économie lexicale en LS est assurée, d'une part, par la structure compositionnelle en partie commune des UT et des UL, qui permet le va-et-vient entre elles, et, d'une autre part, par le composant morphémique infra-UL.

Nous soulignerons au final l'apport attendu de l'étude du phénomène de signes « peu adaptés » mentionné par les locuteurs de notre corpus. A titre d'exemple, l'UL [PEDIATRE] est un assemblage de l'UL [ENFANT] - par le maintien par la main dominée de la configuration ‘bec de canard’ - et de l'UL [MEDECIN] - par la reprise par la main dominante de la configuration ‘P’ et du mouvement circulaire. Nous avons observé la haute productivité de ce type de combinaison dans nos corpus, néanmoins cette UL n'est pas acceptée par les locuteurs. Nous supposons que ceci peut être expliqué par l'existence d'une hiérarchie entre les composants morphémiques infra-UL et/ou de liens privilégiés entre tel et tel composant. Nous avançons ainsi l'hypothèse que certains liens forme-sens sont tellement productifs qu'ils ‘monopolisent’ l'association de certains composants formels à certaines valeurs de sens. Ceci explique pourquoi le signe [PEDIATRE] n'est pas possible : les mêmes configuration et mouvement sont employés tellement fréquemment dans les UL ayant le sens commun ‘entendant’, qu'ils ont créé un lien forme-sens fort qui apparaît spontanément quand un locuteur les rencontre.

### **Références**

- ARONOFF, M., MEIR, I., & SANDLER, W. (2005). The paradox of sign language morphology. In *Language, Journal of the Linguistic Society of America*, 81 (2), 301-344.
- BRE LE CORRE, G. (2002). *L'organisation structurelle du sens en langue des signes française (LSF) : du statut « figura » du signe standard à l'isotopie « structurelle » du discours signé*. Thèse de doctorat.
- BRENNAN, M. (1990). *Word formation in British Sign Language*. Stockholm. University of Stockholm Press.
- BRENTARI, D. (2008). Sign language phonology : issues of iconicity and universality. In Pizzuto, Pietrandrea & Simone (eds), *Verbal and Signed Languages : Comparing Structures, Constructs and Methodologies*. Berlin: Mouton de Gruyter, 59-80.
- CUXAC, C. (2004). ‘Phonétique’ de la LSF : une formalisation problématique. Silexicales 4 (special issue) *Linguistique de la LSF : recherches actuelles*, 93-113.
- CUXAC, C. (2000). La Langue des Signes Française (LSF). *Les voies de l'iconicité. Faits de langues*, 15-16. Paris : Ophrys.
- CUXAC, C. (1996). *Fonctions et structures de l'iconicité des langues des signes*. Thèse de Doctorat d'Etat, Paris, Université Paris V.

- CUXAC, C. & Sallandre, M-A. (2007). Iconicity and arbitrariness, in French Sign Language: Highly Iconic Structures, degenerated iconicity and diagrammatic iconicity. In Pizzuto, Pietrandrea & Simone (eds), *Verbal and Signed Languages: Comparing Structures, Constructs and Methodologies*. Berlin: Mouton de Gruyter, 13-33.
- FENLON, J., Cormier, K. & Schembri, A. (2014). Building BSL SignBank: The Lemma Dilemma Revisited. *International Journal of Lexicography*, 28(2), 169-206.
- FERNALD, T. & Napoli, D. (2000). Exploitation of morphological possibilities in signed languages. Comparison of American Sign Language with English. *Sign Language Linguistics*, 3, 3-58.
- FRISHBERG, N. & GOUGH, B. (2000). Morphology in American Sign Language. In *Sign Language & Linguistics* 3 (1), 103-131.
- FUSELLIER-SOUZA, I. (2006). Processus de création et de stabilisation lexicale en langues des signes (LS) à partir d'une approche sémiogénétique. In Sabria (éd), *Les langues des signes (LS) : recherches sociolinguistiques et linguistiques*, Glottopol 7, DYALANG, 72- 95
- GARCIA, B. (2013). *Emergence lexicales en langue des signes française (LSF)*. Enseignement et questionnements sur la nature des unités et les niveaux d'organisation en LS. Séminaire de l'UMR SFL, 15 avril 2013, Paris.
- GARCIA, B. (2010). *Sourds, surdité, langue(s) des signes et épistémologie des sciences du langage. Problématiques de la scripturisation et modélisation des bas niveaux en Langue des Signes Française (LSF)*. Mémoire d'Habilitation à Diriger des Recherches, Université Paris 8.
- GARCIA, B. & L'HUILLIER, M-T. (2013). Discursive and morphological processes in the formation and the stabilisation of lexematic units in French Sign Language (LSF). *Theoretical Issues in Sign Language Research Conference (TISLR 11)*, University College London, Deafness, Cognition and Language Research Centre, 10-13 July 2013.
- GARCIA, B. & L'HUILLIER, M-T. (2011). *Corpus de dialogues entre adultes sourds*. Corpus Creagest-Adultes, ANR Corpus.
- GARCIA, B., ROCH, G. & BOUTET, D. (2008). The morpho-phonetic Structuring of LSF (French Sign Language), Actes électroniques de *Theoretical Issues in Sign Language Research 9* (TISLR 9), Florianópolis, Brasil, Université Fédérale de Santa Catarina, 6-9 décembre 2006, Editora Arara Azul.
- GARCIA, B. & SALLANDRE, M-A. (2014). Reference resolution in French Sign Language. In Cabredo Hofherr & Zribi-Hertz (Eds), *Crosslinguistic studies on Noun Phrase structure and reference*. Syntax and semantics series, 39. Leiden: Brill, 316-364.
- GIROD, M. et al (1997). La langue des signes. Dictionnaire bilingue LSF/français, IVT Editions.
- JOHNSTON, T. (2012). Lexical Frequency in Sign Languages. *Journal of Dead Studies and Deaf Education*, 17 (2). 163-193.
- JOHNSTON, T. & Schembri, A. (2007). *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge: Cambridge University Press.
- JOHNSTON, T. & Schembri, A. (1999). On defining lexeme in a sign language. *Sign Language and Linguistics* 2, 115-185.
- LIDDELL, S. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
- MEIR, I. (2012). Word classes and word formation. In Pfau, Steinbach, and Woll (Eds): *Handbook on Sign Language Linguistics*. Mouton De Gruyter, 77-112.
- PADDEN, C. (1983). *Inflection of morphology and syntax in American Sign Language*. Thèse de doctorat en linguistique, Université de Californie, San Diego.

- PADDEN, C. (1988). *Interaction of Morphology and Syntax in American Sign Language*. New York: Garland Publishing.
- PERLMUTTER, D. (1992). Sonority and syllable structure in American Sign Language. *Linguistic Inquiry*, 23, 407-442.
- POINTURIER-POURNIN, S. (2014). *L'interprétation en Langue des Signes Française : contraintes, tactiques, efforts*. Thèse de doctorat de linguistique. Université de la Sorbonne nouvelle, Paris III.
- SALLANDRE, M-A. (2014). Compositionnalité des unités sémantiques en langues des signes. Perspective typologique et développementale. Mémoire d'Habilitation à Diriger des Recherches, Université Paris 8.
- SANDLER, W. (2003). Sign Language Phonology. In Frawley (ed), *The Oxford International Encyclopaedia of Linguistics*. Oxford : Oxford University Press. 4, 57-60.
- STOKOE, W. (1960). Sign Language Structure: an outline of the visual communication systems of the american deaf, *Studies in linguistics: Occasional papers*, 8. Buffalo: Dept. of Anthropology and Linguistics, University of Buffalo.

## Les doubles suffixes revisités : pour un changement radical de l'analyse dérivationnelle

**Gauvain SCHALCHLI,**  
*Université Bordeaux-Montaigne*  
gauvain.schalchli@gmail.com

### **Introduction**

Les « doubles suffixes » sont un ensemble de constructions de régularité variable définies ainsi par Roché (2009) : « les différentes configurations qui voient se succéder, à la fin d'un lexème construit, deux séquences ayant la forme d'un suffixe » (p. 143). Dans l'approche courante de la construction des mots, ces « configurations » posent un problème d'interprétation, dont le titre de l'article témoigne<sup>1</sup>, à savoir est-ce qu'il y a un parallélisme entre les formants apparents dans le mot construit et les opérations réelles qui ont présidées à la construction. Corbin (1987) avait déjà posé cette question et répondu en distinguant des « doubles suffixes » réguliers résultant d'une double suffixation, c'est à dire une chaîne de deux dérivations successives à partir d'une base donnée, et des « doubles suffixes » irréguliers (ou « conventionnels ») résultant de l'aménagement d'une dérivation simple par l'ajout postérieur à la construction régulière d'un « segment parasite » suffixoïde. Cependant, cette réponse était seulement indicative et appelait de ses vœux une étude spéciale qui apporterait une réponse empiriquement fondée. Roché (2009) proposait un cadre d'analyse de ces constructions irrégulières. En partant de données plus nombreuses, il confirmait la distinction entre les deux grandes classes de doubles suffixes esquissées par Corbin (1987) mais d'une part contestait la solution de continuité entre le cas de régularité et le cas d'irrégularité et d'autre part relevait cinq types différents de doubles suffixes ne relevant pas d'une « dérivation en chaîne ».

Blevins (2006) distingue deux types d'approche en morphologie. Les approches « constructives » se basent sur des unités minimales et des règles pour construire les mots. Au contraire, les approches « abstractives » se basent sur les mots pour en abstraire les régularités et les éléments de formation. L'approche abstractive a été argumentée principalement en morphologie flexionnelle mais la morphologie dérivationnelle reste majoritairement « constructive ». La notion même de « double suffixe » et le problème que pose l'interprétation des configurations qu'elle recouvre découlent de l'approche constructive. Dans une approche abstractive, l'analyse de ces configurations ne se distingue pas de l'analyse des mots construits par suffixation en général et ne pose pas d'autre problème que celui de l'émergence des régularités dérivationnelles à partir des mots construits existants. Au contraire, l'irrégularité apparente des « doubles suffixes » montre l'inadéquation de l'approche constructive, dont les analyses de Corbin (1987) et Roché (2009) relèvent, et oblige à poser le problème de l'abstractivité dérivationnelle dans toute sa complexité. Nous proposons de revisiter les doubles suffixes dans une perspective abstractive dans la lignée des travaux de la « morphologie lexicale »<sup>2</sup>.

---

1 « Un ou deux suffixes ? Une ou deux suffixations ? ».

2 Roché (2009b, 2011), Roché et Plénat (2012, 2014)

Roché (2010) a montré la nécessité d'une étape de formation du radical dans la construction des mots. Les travaux de la morphologie lexicale ont montré d'autre part les conséquences analogiques d'un « effet de série ». Dans une approche abstractive, on peut adopter les concepts symétriques de formation du suffixe et d'« effet de réseau ». Nous proposons ici de tester ces hypothèses sous la forme d'un modèle abstractif minimal de la construction des mots par suffixation basé sur les définition suivantes de la construction des mots :

- a. La construction d'un mot est un résultat de l'analogie dérivationnelle.
- b. L'analogie dérivationnelle est la combinaison d'une opération de formation du suffixe et d'une opération de formation du radical.
- c. L'opération de formation du suffixe est l'abstraction de la série dérivationnelle.
- d. L'opération de formation du radical est l'abstraction de la famille dérivationnelle.

De plus, la définition du réalisme des données qu'implique une perspective abstractive<sup>3</sup> conduit à définir un lexique de base psychologiquement compatible avec le point de vue et les connaissances du locuteur, c'est à dire un échantillon du lexique macrolectal sélectionné en fonction de la fréquence.

Nous proposons d'appliquer ces propositions à l'analyse des néologismes en *-ariat*.

## 1. La constitution du lexique

Dans une perspective abstractive, un lexique de base contient les paradigmes dérivationnels qui sont le support de l'analogie, à savoir les séries dérivationnelles et les familles dérivationnelles<sup>4</sup>.

Nous nous sommes basés sur la base de données Lexique3 puisque d'une part les entrées de cette base sont les mots (ou les formes fléchies) et non pas les lexèmes (ou lemmes) et d'autre part parce qu'elle contient des informations de fréquence.

Pour constituer un lexique de base des mots en *-ariat*. On est parti du corpus d'exemples de Roché (2009), qu'on a étendu à partir de Lexique 3. Cette série formelle nous a servi ensuite à capter les familles formelles par soustraction de *-ariat* à la forme des membres de la série formelle et en cherchant les occurrences de cette chaîne radicale dans lexique 3.

Les exemples de mots en *-ariat* de Roché (2009) est la liste suivante : *bonzariat*, *entreprenariat*, *mécénariat*, *mannequinariat*, *vedettariat*, *flicariat*, *intermittariat*, *interprétariat*, *podestariat*, *presbytariat*, *robotariat*, *prophétariat*, *putariat*. Seul le mot *vedettariat* est représenté dans la base, ce qui concorde avec le caractère néologiques des exemples de Roché (2009). Les mots en *-ariat* extraits de Lexique 3 ajoutés sont les suivants : *actionnariat*, *commissariat*, *notariat*, *partenariat*, *prolétariat*, *salariat*, *secrétariat*, *volontariat*.

Pour définir les informations associées aux entrées, nous avons retenu deux catégories d'information : la fréquence et les propriétés « sémantiques » (au sens large, c'est à dire « non phonologiques »). L'information de fréquence a été extraite de Lexique 3 et correspond à la fréquence de lemme issu du corpus Frantext. L'information « sémantique » de notre lexique regroupe à la fois les propriétés grammaticales ou morpho-syntaxiques (genre, nombre, catégorie, ...) et les informations de sémantique lexicale ou référentielles (notion, classes sémantique, ...). Les informations grammaticales ont été extraites de lexique 3. Les informations lexicales sont des éléments de définitions reprenant la notion de base et les diverses classes et fonctions sémantiques impliquées dans la construction du mot. Toutes les informations « sémantiques » sont réduites au minimum pour les besoins de l'exemple, sont

3 Blevins et al. (2017).

4 Cf. Hathout (2011)

rassemblées en une seule catégorie d'information et sont ordonnées de la plus abstraite (la catégorie syntaxique) à la plus concrète (la notion de base).

## 2. La génération abstractive

Le mini-lexique regroupe les « néologismes » (considérés comme mots potentiels) en *-ariat* et leurs familles « phonologiques ». L'échantillon doit servir de « matériel de base » à partir duquel le processus d'abstraction permettra de fournir les formants phonologiques pour la construction des mots en *-ariat*. Nous avons donc défini comme « néologismes » tous les mots en *-ariat* de fréquence nulle (*notariat*, *entreprenariat*, *salariat*, *putariat*, *interprétariat*, *bonzariat*, *mécénariat*, *mannequinariat*, *flicariat*, *intermittariat*, *presbytariat*, *robotariat*, *prophétariat*) et retenu dans le lexique de base tous les mots ayant une fréquence positive.

Le processus d'abstraction morphologique opère sur des paradigmes. La sélection des paradigmes est guidée sémantiquement à partir des propriétés sémantiques du mot à construire. Ce « modèle » sémantique sert d'abord à sélectionner la série sémantique par soustraction de la propriété la plus spécifique jusqu'à constitution d'une série (même réduite à un seul élément). Ensuite, on sélectionne la famille avec le même processus à partir des propriétés éliminées par la sélection de la série, c'est à dire les plus concrètes, et soustraction de la propriété la moins spécifique jusqu'à constitution de la famille (réduite éventuellement à un seul membre).

Une fois qu'une série et une famille sont sélectionnées sémantiquement, on peut simuler l'opération de formation du radical et du suffixe par abstraction des propriétés phonologiques (ou orthographiques) communes de la famille ou de la série. La somme des résultats de ces deux abstraction aboutissant, associée aux propriétés sémantiques visées au départ, le néologisme.

## 3. Exemple : La construction abstractive du nom de personne collectif référant à la notion populaire de la police (*flicariat*)

Propriétés sémantiques

(du mot à construire) : *NOM-m-ensemble-personne-POP-police*

(retenues pour la sélection de la série) : *NOM-m-ensemble*

Série sélectionnée: *partenariat* *NOM-m-ensemble-personne-partenaire*

*prolétariat* *NOM-m-ensemble-personne-proléttaire*

*actionnariat* *NOM-m-ensemble-personne-actionnaire*

Propriétés phonologiques

(abstraites de la série) : *-ariat*

Propriétés sémantiques

(retenues pour la sélection de la famille): *POP-police*

Famille sélectionnée : *flic* *NOM-m-personne-POP-police*

*flicard* *NOM-PEJ-personne-POP-police*

Propriétés phonologiques

(abstraites de la famille) : *flic-*

Combinaison

(des propriétés phonologiques abstraites de la série et de la famille) : flicariat

#### 4. Interprétation des résultats

Certains résultats sont conformes aux attentes (*interprétariat* « fonction d'interprète » ou « ensemble des interprètes » à partir de la série *partenariat/actionnariat/prolétariat* et de la famille *interpréter/interprétation*). Mais d'autres sont le sont moins :

*bonzariat* « ensemble des bonzes » aboutit à *bonzariat* (base : *bonze*)

*notariat* « fonction de notaire » aboutit à *notaariat* (base : *notaire, notarial, ...*)

*salariat* « ensemble des salariés » ou « statut de salarié » aboutit à *salaariat* (base : *salaire, salarié, ...*)

*entreprenariat* « ensemble des entrepreneurs » aboutit à *entreprariat* (base : *entreprendre, entrepreneur, ...*)

*intermittariat* « ensemble des intermittents » ou « statut d'intermittent » aboutit à *intermittentariat* (base : *intermittent*)

*prophétariat* « ensemble des prophètes » ou « statut de prophète » aboutit à *prophèteariat* (base :

*mécénariat* « ensemble des mécènes » qui aboutit à *mécénatariat* (base : *mécénat*)

Par contraste, l'échec relatif des cas plus complexes permet d'envisager comment affiner le modèle et met en lumière sous une forme abstraite les diverses contraintes dérivationnelles (phonologie VS graphie, allomorphie, polysémie, organisation du réseau familial, stratification de la série modèle, dynamicité du processus abstractif, interaction analogiques de la formation du suffixe et de la formation du radical, ...). On améliorerait facilement le modèle en prenant la forme phonologique plutôt que la forme orthographique et en prenant l'ensemble des formes fléchies attestées de chaque lexème ainsi que leurs fréquences propres plutôt que la forme de citation associée à la fréquence globale du lexème.

La polysémie et la définition des propriétés sémantiques des entrées lexicales posent le problème de la constitution de base de données sémantiquement réalistes. L'allomorphie pose le problème de la pertinence en morphologie dérivationnelle d'une représentation phonologique plus concrète (syllabation, accès aux traits phonologiques) que la représentation phonématische. Plus généralement, l'implication des familles des membres de la série-modèle dans la formation du suffixe et celle des séries des membres de la famille-base dans la formation du radical, l'accès à différents niveaux d'abstraction et l'interaction de l'abstraction serielle et de l'abstraction familiale au cours du processus permettrait de capter les conséquences complexes de la dimension paradigmatisques<sup>5</sup> (ou lexicale) de l'analogie dérivationnelle, à savoir les « effets » de série et de réseau. Cependant, malgré les imperfections techniques et les incertitudes théoriques, cette illustration permet de prouver qu'une approche de la dérivation basées sur des données et un processus réaliste est possible et féconde et que la nécessité d'une description surfaciste du lexique construit permettant un recensement des familles et de séries dérivationnelles est nécessaire.

#### Références

BLEVINS, J. P. (2006). “Word-based morphology”. *Journal of Linguistics*, 42, 531–573.

5 Cf. Van Marle (1985). *On the paradigmatic dimension of morphological creativity*, volume 18. Foris Publications, Dordrecht—Holland.

- BLEVINS et al. (2017)
- CORBIN, D. (1987), *Morphologie dérivationnelle et structuration du lexique*, 2 vol. TÜBINGEN, Max Niemeyer Verlag ; 2e éd., Villeneuve d'Ascq, Presses Universitaires de Lille, 1991.
- HATHOUT, N. (2011b) "Une approche topologique de la construction des mots: propositions théoriques et application à la préfixation en anti-", in M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique*, Paris / Londres , Hermès-Lavoisier, 251-318
- ROCHE, M. (2009a), « Un ou deux suffixes ? Une ou deux suffixations ? », in B. Fradin, F. Kerleroux et M. Plénat (dir.), *Aperçus de morphologie du français*, Saint-Denis, Presses Universitaires de Vincennes, 143-173.
- ROCHE, M. (2009b), « Pour une morphologie lexicale », *Mémoires de la Société de Linguistique de Paris*, Nouvelle série n° 17, 65-87.
- ROCHE, M. (2010), « Base, thème, radical », *Recherches Linguistiques de Vincennes* 39, 95-133.
- ROCHE, M. (2011), « Quelle morphologie ? », in M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique*, Paris / Londres , Hermès-Lavoisier, 15-39.
- ROCHÉ, M. & PLÉNAT, M. (2014), « Le jeu des contraintes dans la sélection du thème présuffixal », In F. Neveu et al. (éd.), *4e Congrès Mondial de Linguistique Française* (pp. 1863-1878). Paris : ILF.

## Typological flexibility: Anglo-French morphology & syntax and its sociolinguistic implications

**Yela SCHAUWECKER,**  
*University of Stuttgart*  
yela.schauwecker@ling.uni-stuttgart.de

In this talk, I present first results of an ongoing investigation of the linguistic contact situation in Mediæval England<sup>1</sup>, between Middle English, continental Old French, and Anglo-French. My findings contradict earlier research, suggesting that Anglo-French was showing its “imminent death by exhibiting the standard features of a dying language” (Kibbee 1996, 15, cited in Ingham 2010, 164). In a similar vein, Tanquerey, among others, diagnosed a decline of Anglo-Norman from 1250 (Tanquerey 1916), and Berndt concluded that later Anglo-Norman was “clearly a language learned at school” (Berndt 1972, 354). This suggests a non-native like acquisition scenario of Anglo-French. On the other hand, it has been stated that “the syntactic properties were resiliently transmitted – across a period of about 300 years, from the Conquest to the Black Death.” (Ingham 2012, 159), thus suggesting native-like competence for speakers of Anglo-French for the period in question. My approach allows a finer-grained analysis of this linguistic contact constellation. Moreover, from a methodological perspective, my contribution shows 1) the potential that lies in applying acquisitionist considerations to historical data and 2) that the Talmyan dichotomy can serve as a touchstone in investigating historical contact situations. To my knowledge, this approach that has not previously been mentioned in the literature.

### A remark on Anglo-French:

Anglo-French is a variety of Old French spoken in Mediæval England for almost four hundred years following the Norman Conquest (1066 to the 15th century). As such, it is in constant contact with two typologically different languages (Talmy 1985, 2000): English is a prototypically satellite-framed language, whereas French is clearly verb-framed. This typological distinction, according to Slobin (2006 and subsequently Mora Gutiérrez 2011, Engemann 2013, Stolova 2015, among others), is based on different ways of perceiving and conceptualizing event structures.

For example, with manner-of-motion verbs, speakers of satellite-framed languages can express all components of PATH with elements outside the main verb: *The man ran back down into the cellar*. In contrast to this, speakers of a verb-framed language lexicalize such an event using a PATH-verb, such as *descendre* or *monter*. When manner-of-motion is to be explicated, they typically rely on subordinate structures (durative sub-clauses, gerundive-constructions).

---

<sup>1</sup> This research originates from my work within the DFG-funded project *Borrowing of Argument Structure in Contact Situations* (BASICS) at the Universities of Mannheim and Stuttgart.

### Theoretical consideration:

If the restrictions imposed on syntax by typology emanate from (or reflect) cognition, I would expect them to prove extremely resistant even in a cross-typological contact situation, at least as far as a typical adult L2-scenario is concerned, because spatial cognition is fully developed by the end of early childhood, and because grammar is acquired target-like up to about 6 years of age (Meisel 2006, De Houwer 1990). In turn, these findings suggest that, cross-typological exposure during early childhood should result in a nevertheless target-like acquisition of morphology in *both* languages. From this perspective, absence of typological deviation would lead me to assume natural and balanced exposure to both languages in early childhood. On these grounds I assume that: if the distribution of cross-linguistic deviation in Anglo-Norman is found to vary with time, this would imply varying sociolinguistic contact constellations in Mediæval England over time (Johanson 2002). I essentially argue that, periods with a higher rate of satellite-framed structures in Anglo-French suggest a post-childhood L2 learner-scenario, whereas an extremely low rate reflects a more natural acquisition scenario, with both languages being acquired in early childhood.

### Empirical basis:

Albeit never explicitly defined, the group of verbs expressing *manner-of-motion* readily serves as a touchstone to my hypothesis. In a verb-framed language, manner-of-motion verbs, such as *ride*, *run*, *crawl*, *swim* among many others, do not allow for resultative complements. Thus, a construction of the type: *The bird flies on(to) the roof* is not only not accepted, but even considered ridiculous by French speakers. In terms of the Vendlerian *Aktionsarten*, activity-verbs do not license telicizing constituents and therefore cannot be transformed into accomplishment-verbs in this type of language.

I isolated some 20 manner-of-motion verbs that existed in Mediæval times and that still exist in Modern French. In order to find them, I relied on Dubois und Dubois-Charlier (1997) and on Levin (1993), cross-checking in Tobler/Lommatsch (Tobler u. a. 1925), in the etymological part of the TLFi<sup>2</sup> as well as on the FEW<sup>3</sup>.

Secondly, I traced these verbs independently in insular and continental Old French. For Old French, I used all texts in the *Base du Français Médiéval*-database<sup>4</sup> (BFM) except those from England. For the Anglo-Norman part, I relied on the Anglo-Norman texts included in the BFM as well as on the texts that can be found in the ANHub-database (avoiding double counts for doublets). In sum, my corpus includes roughly three million tokens for Old French, and the same for Anglo-French.

I searched for certain morpho-syntactic patterns that are characteristic for satellite-framed languages, performing each query on both corpora independently.

---

<sup>2</sup> TLFi : *Trésor de la langue Française informatisé*, <http://www.atilf.fr/tlfii>, ATILF - CNRS & Université de Lorraine.

<sup>3</sup> Le dictionnaire étymologique et historique du galloroman (français et dialectes d'oïl, francoprovençal, occitan, gascon) Französisches Etymologisches Wörterbuch (FEW), cf. <https://apps.atilf.fr/lecteurFEW/index.php/>

<sup>4</sup> Barbance-Guillot, Céline; Heiden, Serge; Lavrentiev, Alexei (2007): Typologie des textes et des phénomènes linguistiques pour l'analyse du changementlinguistique avec la Base de Français Médiéval. Linx, No. spécial (2e trimestre), 125-139. <http://txm.ish-lyon.cnrs.fr/bfm/html/CSU.html>

**Observations:**

1. Anglo-Norman texts include s-framed patterns that are incompatible with a verb-type language, namely GOAL-constituents (ex. *chevauchier a Leicestre* “ride to Leicester”), pronominals (ex. *il i chevauche* “he rides there”), and locative adverbs expressing GOAL (ex. *il chevauche la* “he rides there”)
2. These patterns are not found – or considerably less frequent – in continental Old French.
3. The most frequent v-framed telicizing construction from OF, the durative *tant ... que* sub-clause, is underused to the point of being virtually absent from AF sources (e.g. *il chevauche tant qu'il vient a Thebes le matin* “he rides until he comes to Thebes in the morning”)
4. Their distribution varies over times and texts, in a sense that satellite-framed patterns occur more frequently in some authors, and tend to be more frequent in the oldest texts (prior to the beginning of the thirteenth century). Subsequently, they virtually vanish until the middle of the fourteenth century, when their frequency starts again rising.

At this point, and based on manner-of-motion verbs, our findings confirm societal bilingualism from the second third of C13 to the end of the fourteenth century, but adult L2-acquisition of Anglo-French for some individual authors/texts.

**References**

- BERNDT, R. (1972). The period of the final decline of French in Medieval England (fourteenth and early fifteenth centuries)“. *Zeitschrift für Anglistik und Amerikanistik* 20: 341–69.
- DUBOIS, Jean, und Françoise DUBOIS-CHARLIER. (1997). *Les verbes français*. Paris: Larousse.  
<http://rali.iro.umontreal.ca/rali/?q=fr/lvf>.
- ENGEMANN, Helen. (2013). Motion Event Expression in Bilingual First Language Acquisition“. Unpublished PhD thesis, Cambridge: University of Cambridge, UK.
- INGHAM, Richard, Hrsg. (2010). The Transmission of Later Anglo-Norman: Some Syntactic Evidence“. In *The Anglo-Norman language and its contexts*, 164–82. Woodbridge : Woodbride, Suffolk, UK: York Medieval Press ; in association with Boydell & Brewer.
- . (2012). *The transmission of Anglo-Norman: language history and language acquisition*. Language faculty and beyond 9. Amsterdam ; Philadelphia: John Benjamins Pub. Company.
- JOHANSON, Lars. (2002). Contact-induced change in a code-switching framework“. In *Language change: the interplay of internal, external and extra-linguistic factors*, herausgegeben von M. C. Jones und E. Esch. Berlin: Mouton de Gruyter.
- KIBBEE, Douglas A. (1996). Emigrant Languages and Acculturation: The Case of Anglo-French“. In *The Origins and Development of Emigrant Languages: Proceedings from the Second Rasmus Rask Colloquium, Odense University, November 1994*, herausgegeben von Hans F. Nielsen, Rasmus Rask Colloquium, und Institut for Sprog og Kommunikation, 1–20. Rask Supplement 6. Odense: Odense Univ. Press.
- LEVIN, Beth. (1993). *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press.
- MORA GUTIÉRREZ, Juan Pablo. (2011). Directed motion in English and in Spanish“. <http://elies.rediris.es/elies11/>.
- SLOBIN, Dan I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse and cognition“. In *Space in languages: linguistic systems and cognitive categories*, herausgegeben von Maya Hickmann und Stéphane Robert, 59–81. Typological studies in language, v. 66. Amsterdam ; Philadelphia: J. Benjamins Pub. Co.

- STOLOVA, Natalya. (2015). *Cognitive Linguistics and Lexical Change: Motion Verbs from Latin to Romance*. Amsterdam/Philadelphia: John Benjamins.
- TALMY, Leonard. (1985). Lexicalization patterns: Semantic structure in lexical forms“. In *Language typology and syntactic description: Grammatical categories and the lexicon*, herausgegeben von Tomohty Shopen, 3:57–149. Cambridge: Cambridge University Press.
- . (2000). *Toward a Cognitive Semantics. Volume II, Volume II,*. Cambridge, Mass.: MIT Press.
- TANQUEREY, F. 1916. *Lettres anglo-françaises*. Paris.
- TOBLER, Adolf, Erhard LOMMATZSCH, Hans Helmut CHRISTMANN, und Richard BAUM. 1925. *Altfranzösisches Wörterbuch*. Stuttgart: Steiner.

## The naming potential and acquisition of novel compounds and phrases

**Marcel SCHLECHTWEG,**  
*University of Kassel, Germany*  
 marcelschlechtweg@gmail.com

The distinction between compounds and phrases has been hotly debated over the last decades. The current paper aims at investigating whether compounds and phrases differ with respect to how well they fulfill the naming function and to how well they are acquired. **Novel** German adjective-noun (AN) compounds and phrases are examined for this purpose.

Generally speaking, compounds and phrases might be differentiated on structural, semantic-functional, and cognitive grounds. With respect to the structural level, it is assumed that inflectional agreement functions as the decisive factor to define a compound, whose adjective does not carry an inflectional suffix (e.g. *Grünreiher* ‘green heron’ (= specific kind of heron)), and a phrase, which contains an adjective with an inflectional suffix expressing number, gender, case, and/or definiteness (e.g. *grüner Reiher*, ‘green heron’ (= any heron that is green for whatever reason)) (cf., e.g., Booij 2010). Inflectional agreement is considered the so-called primary factor, i.e. the only factor that defines constructions as either compounds or phrases. Other structural factors, so-called secondary factors, do not define but can characterize only. Stress, orthography and types of adjectives belong to the latter category. German AN compounds typically carry initial stress, are written as one unit and show certain restrictions with regard to the types of adjectives that can occur as a first constituent. In contrast, phrases are known to bear non-initial stress, to be spelled with a space between the two elements and to be free of any limitations concerning the adjective types (cf., e.g., Schlücker 2014). All these criteria, however, only characterize German AN compounds and phrases by default and some exceptions exist (e.g. compounds such as *Schlechtwetter* ‘bad weather’ that have non-initial stress).

With respect to semantic-functional aspects, it has been observed that compounds normally *name* complex lexical concepts, i.e., e.g., they refer to kinds (e.g. *Braunbär*, ‘brown bear’ (= specific kind of bear)), while phrases are primarily used for the descriptive function in German (e.g. *brauner Bär*, ‘brown bear’ (= any bear that is brown for whatever reason)) (cf., e.g., Hüning 2010). Bücking (2009, 2010) and Härtl (2015) apply this idea to novel constructions and argue that compounds are inherently semantically non-compositional and, therefore, prone to fulfill the naming function. Barz (1996) and Härtl (2015) ascribe this property to the structural nature of compounds. Since compounds, due to the lack of inflectional adjectival suffixes, differ from phrases, i.e. from the default constructions, on structural grounds, they inherently trigger a deviation in meaning from the standard meaning, which is expressed through a phrase. According to Härtl, this is the reason why compounds are more likely to be lexicalized: They are non-compositional by default and the non-compositional semantics have to be stored. It has to be emphasized that the author focuses on *novel* constructions and claims, in opposition to Schlücker & Hüning (2009), that compounds are non-compositional in meaning right from their creation onwards, i.e. without having been lexicalized. Although phrases can also be non-compositional and represent naming units, they can do so only later in time.

These ideas are the starting point of the experiments described in the present paper. If compounds, as opposed to phrases, are by their nature non-compositional right from the beginning (and not just after lexicalization) due to their non-default structure, there might be a difference in how well compounds and phrases are used as naming units and in how well the two construction types are processed and stored in the mental lexicon. In the literature, these questions have been addressed over the last years. Several contributions have approached the first issue, examined the naming potential of compounds and phrases, and, overall, indicate mixed results (cf., e.g., Härtl 2017; Schlücker & Plag 2011). Recently, several studies have also looked at the second question and asked whether a cognitive contrast between compounds and phrases exists (cf. Kotowski, Böer & Härtl 2014; Schlechtweg 2018; Schlechtweg & Härtl 2016). The current paper aims at expanding the research on these issues and at presenting the results of new experiments that have addressed the two aforementioned questions, i.e., whether compounds and phrases differ on semantic-functional and cognitive grounds. The results of the studies will be interpreted against the background of a structural, semantic-functional, and cognitive distinction between compounds and phrases.

## References

- BARZ, Irmhild. (1996). Komposition und Kollokation. In Knobloch, Clemens & Schaefer, Burkhard (eds.), *Nomination: Fachsprachlich und gemeinsprachlich*, 127-146. Opladen: Westdeutscher Verlag.
- BOOIJ, Geert. (2010). *Construction Morphology*. Oxford: Oxford University Press.
- BÜCKING, Sebastian. 2009. How do phrasal and lexical modification differ? Contrasting adjective-noun combinations in German. *Word Structure* 2(2). 184-204.
- BÜCKING, Sebastian. (2010). German nominal compounds as underspecified names for kinds. In Olsen, Susan (ed.), *New impulses in word-formation* (Linguistische Berichte, Sonderheft 17), 253-281. Hamburg: Buske.
- HÄRTL, Holden. (2015). Semantic non-transparency in the mental lexicon: On the relation between word-formation and naming. In Brinker-von der Heyde, Claudia & Kalwa, Nina & Klug, Nina-Maria & Reszke, Paul (eds.), *Eigentlichkeit: Zum Verhältnis von Sprache, Sprechern und Welt*, 395-416. Berlin: de Gruyter.
- HÄRTL, Holden. (2017). Normality at the boundary between word-formation and syntax. In Franz d'Avis & Horst Lohnstein (eds.), *Normalität in der Sprache* (Linguistische Berichte, Sonderheft 22). 71-98.
- HÜNING, Matthias. (2010). Adjective + Noun constructions between syntax and word formation in Dutch and German. In Onysko, Alexander & Michel, Sascha (eds.), *Cognitive perspectives on word formation* (Trends in Linguistics: Studies and Monographs 221), 195-215. Berlin: De Gruyter Mouton.
- KOTOWSKI, Sven & BÖER, Katja & HÄRTL, Holden. (2014). Compounds vs. phrases: The cognitive status of morphological products. In Rainer, Franz & Gardani, Francesco & Luschützky, Hans Christian & Dressler, Wolfgang U. (eds.), *Morphology and meaning: Selected papers from the 15<sup>th</sup> International Morphology Meeting, Vienna, February 2012* (Current Issues in Linguistic Theory 327), 191-203. Amsterdam: John Benjamins.
- SCHLECHTWEG, Marcel. (2018). *Memorization and the compound-phrase distinction: An investigation of complex constructions in German, French and English* (Studia Grammatica 82). Berlin: De Gruyter.
- SCHLECHTWEG, Marcel & HÄRTL, Holden. (2016). Memorization and the morphology-syntax divide: A cross-linguistic investigation. *SKASE Journal of Theoretical Linguistics* 13(2). 46-68.

- SCHLÜCKER, Barbara. (2014). *Grammatik im Lexikon: Adjektiv-Nomen-Verbindungen im Deutschen und Niederländischen* (Linguistische Arbeiten 553). Berlin: De Gruyter.
- SCHLÜCKER, Barbara & HÜNING, Matthias. (2009). Compounds and phrases: A functional comparison between German A + N compounds and corresponding phrases. *Rivista di Linguistica* 21(1). 209-234.
- SCHLÜCKER, Barbara & Ingo PLAG. (2011). Compound or phrase? Analogy in naming. *Lingua* 121(9): 1539-1551.

## A relational nominal structure in nominal predicates

**Masaharu SHIMADA,**

*University of Tsukuba*

shimada.masaharu.fu@u.tsukuba.ac.jp

**Akiko NAGANO,**

*Tohoku University*

nagano@ling.human.is.tohoku.ac.jp

### 1. Introduction

Japanese predicates which are characterized as ending with the copula *-da* in a conclusive form are divided into two types according to whether it changes into *-na* or *-no* in prenominal modification, as illustrated in (1):

- |        |                                      |   |                               |
|--------|--------------------------------------|---|-------------------------------|
| (1) a. | <i>kono doresu-wa gooka-da.</i>      | → | <i>gooka-na doresu</i>        |
|        | this dress -TOP gorgeous-COP         |   | gorgeous-NA dress             |
|        | ‘This dress is gorgeous.’            |   | ‘a gorgeous dress’            |
| b.     | <i>kono doresu-wa furansufuu-da.</i> | → | <i>furansufuu-no doresu</i>   |
|        | this dress -TOP French -COP          |   | French -NO dress              |
|        | ‘This dress is in the French style.’ |   | ‘a dress in the French style’ |

The predicate *gooka-da* in (1a), taking the form *gooka-na* as a prenominal modifier, is often called a nominal adjective, and the predicate *furansufuu-da* in (1b), taking the form of *furansufuu-no* as a prenominal modifier, a nominal predicate.

Though their difference in surface form observed in this particular environment is well known in Japanese linguistics, their difference in morphosyntactic structure has not been fully understood. Adopting and extending Adger’s (2013) analysis of relational nouns, this paper aims to morphosyntactically differentiate the two types of predicates. Specifically, it is shown that nominal predicates, in contrast to nominal adjectives, consist of functional categories licensing relational meanings and realization of relational nouns.

### 2. Adger’s (2013) analysis of relational nouns

Relational nouns are usually assumed to take arguments as verbs do. One weakness to this idea is that arguments of relational nouns are actually deletable. For example, the relational noun *edge* can omit its PP argument of *the table*, as in (2):

- (2) The edge (of the table) was rough. (Adger (2013: 62))

Adger (2013) addresses the question of why this is the case, and proposes that nouns are just atoms lacking argument structures and do not take a complement structurally.

According to Adger, the relational interpretation in (2), that is, the part-whole relationship between *the edge* and *the table* is established through such a functional head as  $\sqrt{PART}$ . The following is a rough illustration of the structure:

- (3) [ρ the edge [ρ of the table [ρ $\sqrt{\text{PART}}$  ]]]

Both *the edge* and *of the table* are selected by the part-denoting functional head. The functional head is a locus of a conceptual meaning of the part-whole relation. Notice that the edge of the table is the part of the table. So the outer position occupied by *the edge* in (3) is the position for the element concretely instantiating the notion of part, phonologically realizing the functional head. When we focus on other parts, such as a side, the outer position is occupied by *the side*, producing the expression *the side of the table*. Relational nouns such as *edge* and *side* are thus just a concrete or phonological instantiation of the abstract relational notion represented by functional heads. Other relational meanings are established by other functional heads. For example, a representation-denoting functional head  $\sqrt{\text{REP}}$  derives such a phrase as *a photo of Lilly*, whose morphosyntactic structure is [ρ *a photo* [ρ of *Lilly* [ρ $\sqrt{\text{REP}}$  ]]]. Turning to the inner position occupied by *the table*, it is the complement position of the functional head. *The table* is not an argument of the noun *edge* itself. Rather its occurrence is attributed to the functional head  $\sqrt{\text{PART}}$ .

### 3. Application: Nominal predicates as involving relational nouns

Notice first that the stem of nominal predicates in Japanese usually takes compositional structures. For instance, the nominal predicate *furansufuu-da* ‘in the French style’ in (1b) can be analyzed as *furansu-fuu-da* ‘(lit.) France-style-Cop’. The stem *furansufuu* thus consists of the free morpheme *furansu* ‘France’ and the bound morpheme *-fuu* meaning ‘style’. Other examples are given below:

- (4) a. *ee-gumi-da* (lit.) A-class-COP ‘be in group A’  
 b. *cha-iro-da* (lit.) brown-color-COP ‘be brown’  
 c. *roo-rei-da* (lit.) old-age-COP ‘be old’

The examples in (4) are all nominal predicates, since they take *-no* forms in prenominal modification, as shown, for example, in *ee-gumi-no tiimu* ‘(lit.) A-type-NO team’, meaning ‘a team in group A’.

The proposal here is that *-fuu* ‘style’, *-gumi* ‘class’ *-iro* ‘color’ and *-rei* ‘age’ correspond to relational nouns. In Emonds’s (2000) terminology, they are semi-lexical categories denoting abstract notions concerning style, type, color or age. Extending Adger’s analysis, their realization can be attributed to functional categories. For example, the stem *furansu-fuu* ‘France-style’ thus has the following abstract morphosyntactic structure and its surface form results from an independent process of phonological mapping:

- (5) [ρ *-fuu* [ρ *furansu* [ρ $\sqrt{\text{STYLE}}$  ]]]

Note that Japanese has other surface forms for  $\sqrt{\text{STYLE}}$  such as *-ryuu*. If it is utilized as an instantiation of the notion, we get such an expression as *furansu-ryuu-da* ‘in the French style’. Semi-lexical nouns like *-fuu* or *-ryuu* are phonological realizations of a functional head denoting the abstract notion.

Note that while the semi-lexical nouns or the relational nouns *-fuu* and *-ryuu* are bound morphemes, *kumi* ‘group’ is a free morpheme like such an English relational noun as *edge*. Assuming that it expresses the notion of CLASS, (4a) takes the morphosyntactic structure as in:

- (6) [ρ *kumi* [ρ *ee* [ρ $\sqrt{\text{CLASS}}$  ]]]

In the case of the free morpheme *kumi*, it can also be realized preceding the complement, as *edge* is realized preceding *the table*. So (7) is another option of a phonological realization form of (6).

- (7) *kumi A-no tiimu*  
 group A-LINKER team  
 ‘be in a team group A’

This ordering pattern may support the present analysis. Predicate nominals, but not nominal adjectives, have complex structures consisting of functional heads for relational nouns.

#### 4. The difference between nominal predicates and nominal adjectives

Like nominal predicates, nominal adjectives also seem to have complex structures. For example, *gooka-da* ‘be gorgeous’ in (1a) can be analyzed as *goo-ka-da* ‘(lit.) gorgeous-gorgeous-COP.’ This is shown by the fact that *goo* ‘gorgeous’ and *ka* ‘gorgeous’ are separately written with a logogram called *kanji* scripts, representing a lexeme. The nominal adjective *goo-ka* ‘gorgeous-gorgeous’ consists of two lexemes. However, it does not involve functional heads occurring in nominal predicates. As the gloss shows, it is a dvandva compound, having a coordinate interpretation. Similar examples are (8a) and (8b), and both are nominal adjectives:

- (8) a. *koo-dai-da* (lit.) wide-large-COP ‘be wide’  
 b. *yuu-shuu-da* (lit.) good-good-COP ‘be good’

Note that there is a non-dvandva nominal adjective like (9), which also seems to have a complex structure:

- (9) *yuu-noo-da* (lit.) with-ability-COP ‘be competent’

As shown by the gloss, (9) consists of two elements, a suffix-like element, *yuu-*, and a lexemic element, *-noo*. Importantly, this case again lacks a meaning attributed to such a functional head as licensing a nominal predicate.

To sum up, it is impossible to abstract any conceptual meaning from morphemes in (8) and (9) and identify functional heads like  $\sqrt{\text{STYLE}}$  in nominal adjectives. Given that nominal predicates in Japanese always require functional heads like  $\sqrt{\text{STYLE}}$  to combine two elements, it can be said that there is a difference in derivational process between nominal predicates and nominal adjectives. A derivational process of nominal predicates is not a typical process of producing compounds, in which two lexical categories are just combined to derive a new lexical category. A process of deriving nominal adjectives like (8) and (9) corresponds to a typical case of compound formation, if such a Japanese morpheme as *yuu-* ‘with’ in (9) is not a prefix but a kind of lexeme, as often assumed. This difference in word formation process between nominal predicates and nominal adjectives may support Emonds’s (2000) view that there are two types of compounding processes, compounding involving only lexical categories and that involving both lexical and functional categories.

#### References

- ADGER, D. (2013). *A syntax of substance*, Cambridge, Mass.: MIT Press,  
 EMONDS, J. (2000). *Lexicon and Grammar: the English syntacticon*, Berlin: Mouton de Gruyter.

## Construction Morphology applied to the analysis of medical neologisms in the Middle Ages: benefits and challenges

**Céline SZECEL,**  
*KU Leuven*  
celine.szecel@kuleuven.be

The purpose of this paper is to show how Construction Morphology (Booij 2010) can be fruitfully applied to the morphological analysis of medical neologisms, created in the Middle Ages. In our research project, we are investigating why certain French neologisms which emerged in the field of medicine during the Middle Ages managed to survive in modern French, like *fistule* (“fistula”), while others, *festre* (“fistula”) and *afistuler* (“form a fistula”) for example, disappeared after some time. Our study is based on a corpus of medieval medical texts, CHROMed or Historical Corpus Of French MEDical texts (1.363.499 words), containing both translations from Latin and texts directly written in French from the 13<sup>th</sup>, 14<sup>th</sup> and 15<sup>th</sup> centuries.

We consider all terms created during the Middle Ages as neological, even when they were not coined in the texts of our corpus, except “hereditary” words. Since French originates from Latin, plenty of French terms have evolved from Latin words and went through phonetic changes. *Festre*, for example, is a “hereditary” form which can be traced back to the Latin etymon *fistula*. All analyzed neologisms belong to the medical field of medieval pathology. They have been identified on the basis of etymological information found in the most important dictionaries of French (cf. *infra*), and were then classified as loanwords or French creations, *viz* derivatives and compounds, by applying the still valid typology of Deroy (1956: 215-234), next to the “hereditary” words.

To establish when a word has been coined, one can consult etymological dictionaries like the FEW from Von Wartburg (1922-), and other Middle French dictionaries, like the Godefroy (1880-1902) and the DMF 2015. For Old French, we can rely on the Tobler & Lommatzsch (1915). The TLFi also provides us with etymological notices and gives information about the first appearance of words. Nevertheless, there are some general issues with determining when a word was first coined. Dictionaries use certain corpora of texts and thus can't obviously be exhaustive. Moreover, a lot of medieval manuscripts have been lost which means the first attestations of terms must be regarded as provisory, especially since medieval texts contained in manuscripts could be rediscovered at any moment. Another issue is that dictionaries sometimes provide us with wrong definitions of words. At other times, they don't identify the passage in which the term appears, making it impossible for us to know if they interpreted this word correctly. Because we also have to know the meaning of words to judge if they are neological or not, this is very problematic. All these reasons explain why we didn't look for first attestations of the neologisms, but we simply verify whether these medical terms have been coined during the Middle Ages or not.

By analyzing these medieval neologisms found in our corpus, we would like to examine the hypothesis that being embedded in word families with systematic and transparent correlations between form and meaning reinforces the use of these neologisms and helps them survive. Construction Morphology (Booij 2010) considers complex words as “constructions on the word level” and the notion ‘construction’ is understood as “a conventionalized pairing of form and meaning”. Based on the principles of a hierarchical model, we will show how to establish more abstract-level generalizations, leading to the creation of word-families and abstract schemas, forming a morphological network (Booij 2008). Thus, we will analyze the word-family of *quartenaire* (“one who suffers from quartan fever”) and *quartain* (which means “quartan fever”, when combined with *fievre* “fever”), borrowed respectively from the Medieval Latin *quartenarius* (“one who suffers from quartan fever”, DLD) and the Classical Latin *quartanus* (“quartan”, DLD). As *quartana febris* “quartan fever” is attested in Classical Latin (DLD), *fievre quartaine* is in fact a structural loan. We will compare this word-family to that of the “hereditary” forms *quartenier* (“one who suffers from quartan fever”) and *quart* (which means “quartan fever”, when combined with *fievre* “fever”) in order to ascertain which word-family is more likely to survive and for which reasons.

We’ve adapted the format of Booij’s (2008) constructional schemas to investigate if certain affixes are combined with learned – and thus loaned from Latin – stems/roots, rather than native French ones. According to Zwanenburg, the distribution of stems and affixes doesn’t seem arbitrary: native French affixes could be combined with learned or native stems, whereas learned affixes could only be combined with learned stems (Zwanenburg 1985 : 180-182 ; cf. also Zwanenburg 1992). We would like to verify this hypothesis in our corpus – at least for all the analyzed neologisms, which belong to the medical field of medieval pathology –, maybe thus confirming Zwanenburg’s (1985: 180) claim that “the distribution of suffixes is the fundamental criterion to distinguish learned derivation from non-learned derivation”.

Before creating these constructional schemas and studying which affixes are combined with native or non-native stems/roots, it’s very important to define these affixes clearly. While it may seem straightforward, it is not always easy to analyze (medieval) neologisms in stems/roots and affixes. For words like *apostemation*, *-ion* should be considered as the suffix and *apostemat-* as the stem, as confirmed by Bonami, Boyé & Kerleroux 2009 and Huot 2001, but other affixes can be more problematic. For example, it would be possible to analyze *-ence* as the suffix of *epilence* (“epilepsy”), on the one hand, and *-encie* as the suffix of *epilencie* (“epilepsy”), on the other. However, we’ve decided to consider *epilenc-* as the stem/root of both *epilence* and *epilencie*, and, *-e* and *-ie* as their respective suffixes, by analogy with other neologisms, like *alopice* (“baldness, mange”) and *alopicie* (“baldness, mange”). The first term consists of the root *alopic-* and the suffix *-e*, the second, of the root *alopic-* and the suffix *-ie*. Both words have exactly the same meaning and should be regarded as variants, as is the case with *epilence* and *epilencie* (“epilepsy”). However, presenting a paper at ISMo will be a nice opportunity to get valuable feedback from experts in morphology on these problematic cases.

## References

- BONAMI, O., BOYÉ, G. & KERLEROUX, F. (2009). “L’allomorphie radicale et la relation flexion construction”, in B. Fradin, F. Kerleroux & M. Plénat (éds.), *Aperçus de morphologie du français*. Saint-Denis : Presses Universitaires de Vincennes, 103-125.
- BOUJ, G. (2008). “Composition et morphologie des constructions”, in D. Amiot (éd.), *La composition dans une perspective typologique*, Arras : Artois Presses Université, 49-73.
- BOUJ, G. (2010). *Construction Morphology*, Oxford : Oxford University Press.
- CHROMED: *Historical Corpus Of French MEDical texts*, to be published online, under the supervision of Michèle Goyens, in collaboration with Céline Szecel and Ildiko Van Tricht.
- DEROY, L. (1956). *L'emprunt linguistique*, Paris : Belles Lettres.
- DLD : *Database of Latin Dictionaries*. Turnhout : Brepols Publishers. Site internet : <[clt.brepolis.net/dld/](http://clt.brepolis.net/dld/)>.
- DMF 2015 : *Dictionnaire du Moyen Français*, version 2015. ATILF CNRS – Université de Lorraine. Site internet : <[www.atilf.fr/dmf](http://www.atilf.fr/dmf)>.
- GODEFROY, F. (1880-1902). *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX<sup>e</sup> au XV<sup>e</sup> siècle*, Paris : Vieweg.
- GOYENS, M., SZECHEL, C. & VAN GOETHEM, K. (2017). “Une famille qui fait ‘suer’. Problèmes d’analyse des néologismes médiévaux *sudoral*, *sudorable*, *resudation* et *desudation*”, *Le français en diachronie. Dépendances syntaxiques, Morphosyntaxe verbale, Grammaticalisation*, éds. S. Prévost & B. Fagard. Berne : Peter Lang, 371-403.
- HUOT, H. (2001). *Morphologie. Forme et sens des mots du français*. Paris : A. Colin.
- TLFi : *Trésor de la langue française informatisé*. Site internet : <[atilf.atilf.fr](http://atilf.atilf.fr)>.
- TOBLER, A. & LOMMATSCH, E. (1915). *Altfranzösisches Wörterbuch. Adolf Toblers nachgelassene Materialien bearbeitet und herausgegeben von Ehrhard Lommatsch*, Berlin, Wiesbaden : Fr. Steiner.
- VAN TRICHT, I. (2015). *Science en texte et contexte : la terminologie médicale française utilisée dans les Problèmes d’Evrart de Conty à la lumière du discours médical médiéval* (Thèse de Doctorat, KU Leuven).
- VON WARTBURG, W. (1922-). *Französisches etymologisches Wörterbuch: eine Darstellung des galloromanischen Sprachschatzes*, Basel : Zbinden.
- ZWANENBURG, W. (1985). “La naissance de la dérivation savante en français”, Actes du XVII<sup>e</sup> Congrès International de Linguistique et Philologie Romanes, III, Aix-en-Provence : Université de Provence, 177-188.
- ZWANENBURG, W. (1992). “Composition savante et moyen français”, *Vox Romanica*, LI, 169-177.

## Verbes statifs et mots composés Verbe-Nom du français : une préférence discrète

**Florence VILLOING,**

*Université Paris Nanterre & CNRS UMR 7114 Modyco,*

*Florence.Villoing@parisnanterre.fr*

Les schémas morphologiques sont sensibles aux propriétés sémantiques des verbes de base et inscrivent au titre de la règle des propriétés de restriction (cf. Kerleroux & Fradin 2003, Namer & Villoing 2008, Ferret & Villoing 2015 pour le français). La composition verbe-nom (désormais VN), en tant que procédé morphologique de formation de lexèmes, témoigne de ces restrictions sélectionnelles et sélectionne prototypiquement des verbes dynamiques (cf. 1) en rejetant les verbes statifs typiques (cf. 2) (Villoing 2009, Namer & Villoing 2007).

(1) OUVRE-BOITE, ESSUIE-GLACE, TROUBLE-FETE, PERCE-OREILLE, LECHE-VITRINE, SAUTE-MOUTON.

(2) \*CONNAIT-ALLEMAND, \*SAIT-MATHEMATIQUE, \*AIME-ANIMAL, \*DESIRE-CADEAU

Cependant, une étude extensive de corpus révèle que certains mots composés verbe-nom construits sur des verbes dynamiques sélectionnent une autre valeur aspectuelle de ces mêmes verbes : la valeur stative (cf. 3)

(3) PORTE-BONHEUR, CACHE-SEXÉ, COUVRE-LIT, PARE-SOLEIL, PROTEGE-LIT.

En effet, nombre de verbes présentent une polysémie aspectuelle et la composition VN sélectionne l'une ou l'autre de leurs valeurs ou les deux (cf. 4a pour la valeur dynamique de CACHER et (4b) pour sa valeur stative) :

(4) a. CACHE-DOUILLE : (TLF) « *n.m.* Voleur qui dissimule une pièce dans un compte. »  
CACHE-TAMPON : (G.R.) « *n.m.* Jeu d'enfant où l'on cache un mouchoir ou un objet quelconque que l'un des joueurs doit découvrir. »

b. CACHE-COEUR : (G.R.) « *n.m.* Petite pièce de l'habillement fém., portée croisée sur la poitrine. (...). »

CACHE-RADIATEUR : (G.R.) « *n.m.* Revêtement (...) destiné à cacher un radiateur d'appartement. »

L'ambiguïté de ces verbes est aujourd'hui relativement bien identifiée dans les structures syntaxiques (cf. Huyghe & Jugnet 2010 ; Kratzer 2000). Notre communication se propose de montrer comment elle se révèle aussi en morphologie constructionnelle, approche originale dans la mesure où la littérature s'est surtout focalisée sur l'étude des noms déverbaux statifs (Meinschaeffer 2003, Beauseroy 2009, Haas 2009, Huyghe & Jugnet 2010, Alexiadou 2011, Fradin 2011, Barque, Fabregas & Marin 2012 ; voir Knittel 2015 pour un aperçu) ; négligeant les noms déverbaux sur bases statives dénotant des objets (cf. Fradin 2012 sur la lecture « moyen »).

Nous tenterons de déterminer

- (i) à quels types de verbes statifs appartiennent les verbes sélectionnés par la composition VN ;
- (ii) et s'il existe un lien entre la classe aspectuelle du verbe de base et la dénotation du composé.

## Corpus

Nous mènerons notre analyse à partir de l'étude d'un corpus lexicographique de 1448 mots composés VN du français (cf. le corpus de Villoing 2002) issus de cinq grands dictionnaires généraux, dictionnaires de langue et dictionnaires encyclopédiques non spécialisés (le Trésor de la Langue Française, le Grand Robert de la Langue Française, le Grand Larousse Universel, le Dictionnaire de la langue française, le Dictionnaire Général de la Langue Française). L'ensemble des mots composés VN inscrits dans ces dictionnaires, en entrée ou au sein d'une entrée, sont enregistrés dans le corpus, y compris ceux relevant de dénominations spécialisées ou paraissant désuets.

## Méthodologie

L'analyse de la valeur sémantique du verbe s'est fait en ayant recours aux tests d'identification de la stativité traditionnellement proposés dans la littérature (pour une synthèse, cf. Martin 2008, Marin 2013), tels que, par exemple :

Les verbes statifs

- n'admettent pas la forme progressive (*en train de*),
- n'acceptent pas l'interprétation habituelle du présent,
- n'acceptent pas l'impératif,
- ne sont pas compatibles avec les adverbes d'intentionnalité orientés vers le sujet (comme *délibérément*, *volontairement*, *intentionnellement*...).

Cependant, certains de ces tests présentent un certain nombre de problèmes :

- (i) confusion entre la stativité lexicale (l'Aksionsart) et la stativité grammaticale : problème du test du progressif et de l'habitualité (Bertinetto 1994) ;
- (ii) confusion entre la stativité et l'agentivité : problème des tests identifiant l'intentionnalité puisque l'agentivité n'est pas systématiquement synonyme de dynamicité (Huyghe & Jugnet 2010),
- (iii) mauvaise adaptation au français des tests conçus pour l'anglais (Marin 2013, Copley & Roy 2015).

L'application de chacun de ces tests doit donc systématiquement s'accompagner des précautions nécessaires pour éviter les confusions et autres difficultés.

## Résultats

Les premiers résultats nous conduisent à reconnaître les propriétés suivantes :

- les verbes impliqués dans la composition VN sont rarement des verbes à interprétation stative unique (comme peuvent l'être certains verbes psychologiques) mais des verbes à interprétation ambiguë ;
- dans leur valeur stative, ces verbes relèvent de classes sémantiques bien spécifiques :
  - (i) verbes d'obstruction qui impliquent une relation spatiale : CACHER, COUVRIR, GARDER, PARER, PROTEGER (cf. Kratzer 2000 et Alexiadou 2011 pour le grec)
  - (ii) verbes de soutien et de tenue : PORTER, SOUTENIR.

- la composition VN sélectionne soit exclusivement (cf. 5) soit préférentiellement (cf. 6) la valeur stative des verbes ambigus :

(5) a. PROTEGE-BAS : (G.R.) « *n.m.* Petit chausson (...) servant à protéger le bas du contact de la chaussure. »

PROTEGE-CAHIER : (G.R.) « *n.m.* Couverture en matière souple qui sert à protéger un cahier d'écolier. »

b. \*PROTEGE-ENVAISSEMENT (« Il aimait à protéger les citoyens contre les envahissements de l'illégalité » TLFi) ; \*protège-gamin (« les gendarmes qui avec leurs sabres la protègent des gamins dont les huées l'insulteraient » TLFi).

(6) a. ABAT-JOUR : (G.R.) "n.m.inv. (...) 1. Archit. Fenêtre, soupirail percé obliquement pour éclairer une pièce, un sous-sol de haut en bas.

ABAT-SON(S) : (G.R.) "n.m.inv.(...) Ensemble de lames insérées dans les baies d'un clocher pour rabattre le son des cloches vers le sol(...)".

b. ABAT-FEUILLE : (G.R.) " Dispositif d'une presse à imprimer qui abat et maintient la feuille de papier.".

ABAT-FOIN : (G.R.) "n.m.inv. (...)Trappe par laquelle on fait tomber le fourrage du grenier dans une étable, une écurie.(...)".

- aucun lien sémantique ne se dessine clairement entre l'interprétation du composé VN et la valeur aspectuelle du verbe : en effet, les composés VN construits sur des valeurs statives des verbes peuvent aussi bien dénoter des humains (SOUFFRE-DOULEUR, AIME-DIEU) que des objets (CACHE-RADIATEUR, PORTE-DOCUMENTS), les plus largement représentés, au même titre que les instruments pour les composés VN construits sur des valeurs dynamiques des verbes.

## Conclusion

La composition VN sélectionne aussi bien des verbes statifs que dynamiques pour construire des noms d'objet ou d'humain. Elle se rapproche en cela de la suffixation déverbale en *-eur* qui, à côté des bases dynamiques, peut également choisir des bases statives (cf. ADMIRATEUR, AMATEUR, BAFOUILLEUR, CONNAISSEUR, POSSESSEUR, RONFLEUR), propriété qui n'a pas non plus attiré l'attention des derniers travaux (cf. Rappaport & Levin 1992, Fradin & Kerleroux 2003, Fradin 2005, Alexiadou & Schäffer 2010 sur l'anglais, Roy & Soare 2012 et Huyghe & Tribout 2015 sur le français). La compatibilité des verbes statifs avec les composés VN renforce l'hypothèse que ce schéma de composition n'est pas restreint à la construction de noms d'agent ou d'instruments mais nominalise n'importe lequel des participants sémantiques au procès décrit par la base verbale (cf. Villoing 2002, 2009, Fradin 2005), qu'elle soit dynamique (agent, patient,...) ou stative (expérient, siège...)

Il reste néanmoins encore à déterminer :

- pourquoi la composition VN exclut certains verbes statifs, en particulier des verbes considérés comme non ambigüs : cf. POSSEDER, APPARTENIR, CONNAITRE (Huyghes & Jugnet 2010), ou les verbes de mesure, d'existence/ présence/absence (Alexiadou 2011), par exemple ; les contraintes sont-elles d'ordre sémantiques ou autres ;
- quel lien existe-t-il entre la préférence pour certains types de verbes statifs et la façon dont la composition VN construit le sens ?

Par la suite, il serait nécessaire d'étudier les verbes statifs des mots composés VN d'autres langues romanes (en particulier l'italien et l'espagnol), afin d'évaluer la robustesse de cette propriété dans ces langues où ce schéma de composition reste très productif.

## Références

- ALEXIADOU, A. (2011). Statives and nominalization. *Recherches Linguistiques de Vincennes* 40: 25-52.
- ALEXIADOU, A. & SCHÄFER F. (2010). "On the syntax of episodic vs. dispositional -er nominals", in A. Alexiadou & M. Rathert (eds), *The Syntax of Nominalizations across Languages and Frameworks*, Berlin: Mouton de Gruyter, 9-30.
- BARQUE, L., A. FABREGAS & R. MARIN. (2012). "Les noms d'état psychologique et leurs "objets": étude d'une alternance sémantique". *Lexique*, 20. 21-41.
- BEAUSEROY, D. (2009). Syntaxe et sémantique des noms abstraits statifs : des propriétés verbales ou adjectivales aux propriétés nominales. Thèse de doctorat. Université Nancy 2/UMR 7118-ATILF.
- BERTINETTO, P. M. (1994). Statives, progressives and habituals: analogies and differences. *Linguistics* 32, 391-423.
- COPLEY, B. & ROY, I. (2015). "Deriving the readings of French *être en train de.*" *Romance Languages and Linguistic Theory*, selected proceedings of *Going Romance 2013, 27th Symposium on Romance Linguistics*, ACLC Amsterdam, 103-118.
- FERRET, K. & F. VILLOING. (2015) « French derived instrumentals N-age: semantic properties of the base verb." In *Morphology*, Ingo Plag & Sabine Lapp (editors). Netherlands : Springer : 473-496.
- FRADIN, B. (2005). « On a semantically grounded difference between derivation and compounding". In *Morphology and its Demarcations*, Dressler W. U., D. Katovsky & F. Rainer (eds) Amsterdam / Philadelphia: John Benjamins.
- FRADIN, B. (2011). Remarks on State-denoting nominalizations. *Recherches linguistiques de Vincennes* 40:73-99.
- FRADIN, B. (2012). "Les nominalisations et la lecture 'moyen'". *Lexique* 20:125-152.
- FRADIN, B. & F. KERLEROUX (2003). Quelles bases pour les procédés de la morphologie constructionnelle? In B. Fradin, G. Dal, N. Hathout, M. Plénat, & M. Roché (Eds.), *Les unités morphologiques, Vol. 3, Silexicales* (pp. 76–84). Villeneuve d'Ascq, SILEX: CNRS & Université Lille 3.
- HAAS, P. (2009). *Comment l'aspect vient aux noms. Les propriétés aspectuelles des noms à l'épreuve des restrictions de sélection imposées par certaines prépositions*, PHD, Université Lille 3.
- HUYGHE, R. & A. JUGNET (2010). Nominalisations statives et transfert aspectuel : quel héritage sémantique ? In F. Neveu, V. Muni Toke, T. Klingler, J. Durand, L. Mondada, S. Prévost (éds), *Congrès Mondial de Linguistique Française 2010*, Paris : Institut de Linguistique Française, 1713-1726.
- HUYGHE, R. & D. Tribout (2015). « Noms d'agents et noms d'instruments : le cas des déverbaux en -eur », *Langue française* 2015/1 (n° 185), p. 99-112.
- KNITTEL, M. L. (2015). Les nominalisations : un état des lieux. *Le Français Moderne*, numéro thématique Nominalisations, sous la dir. de V. Magri & I. Roy, 3-17.
- KRATZER, A. (2000). « Building statives ». In L.J./ Conathan et al. (Eds), *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* : 385-399.
- MARIN, R. (2013). La stativité dans tous ses états. Mémoire d'habilitation, Université Paris 8.
- MARTIN, F. (2008). *Les Prédicats statifs. Etude sémantique et pragmatique*. Paris/Bruxelles: De Boeck.
- MEINSCHAEFFER, J. (2003). Nominalizations of French psychological verbs. Syntactic complements and semantic participants. In J. Quer, J. Schrotten, M. Scorretti, P. Sleeman & E. Verheugd (eds.) *Selected Papers from 'Going Romance'*, Amsterdam, 6-8 December 2001. Amsterdam: Benjamins. 235-250.

- NAMER F. & F. VILLOING (2007). "Have cutthroats any thing to do with tracheotomes ? Distinctive properties of VN vs NV compounds in French". In: G. Booij, L. Ducceschi, B. Fradin, E. Guevara, A. Ralli & S. Scalise (Ed.), *Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*, Fréjus, 15-18 sept. 2005), Bologne : 105-124.
- NAMER F. & F. VILLOING (2008). « Interpréter les noms déverbaux : quelle relation avec la structure argumentale du verbe de base ? Le cas des noms en *-oir(e)* du français ». In : J. Durand, B. Habert, B. Laks (Ed.), *Actes du 1<sup>er</sup> Congrès Mondial de Linguistique Française (CMLF 2008)*, Paris, 9-12 juillet 2008, Paris : EDP Sciences. (<http://www.linguistiquefrancaise.org/>) : 1551-1569.
- RAPPAPORT HOVAV M. & B. Levin (1992). “-er nominals: implications for a theory of argument structure”, in T. Stowell & A. Wehrli (eds), *Syntax and Semantics 26: Syntax and the Lexicon*, New York: Academic Press, 127-153.
- ROY I. & E. SOARE (2012). « L'enquêteur, le surveillant et le détenu : les noms déverbaux de participants aux événements, lectures événementielles et structure argumentale », *Lexique* 20, 207-231.
- VILLOING, F. (2002). *Les mots composés [VN]<sub>N/A</sub> du français : réflexions épistémologiques et propositions d'analyse*, mémoire de Thèse de doctorat, Université Paris X-Nanterre.
- VILLOING F. (2009). « Les mots composés VN ». In B. Fradin, F. Kerleroux, M. Plénat (Ed.). *Aperçus de morphologie du français*. Saint-Denis : Presses Universitaires de Vincennes : 175-198.

## Connecting lexica in monolingual and bilingual cross-script morphological processing: base and series effects

**Madeleine VOGA,**

*Université Paul-Valéry Montpellier III, LLACS, France*

[madeleine.voga@univ-montp3.fr](mailto:madeleine.voga@univ-montp3.fr)

**Anna ANASTASSIADIS-SYMEONIDIS,**

*Aristotle University of Thessaloniki, Greece*

[ansym@lit.auth.gr](mailto:ansym@lit.auth.gr)

### 1. Introduction

The question of how words are organised, represented and accessed in the bilingual or monolingual lexicon has been the object of extensive research in the domain of cognitive linguistics for 30 years. Consider for instance a bilingual person: even if language switches can be quite frequent in appropriate circumstances, it is a fact that she can function on only one of the two languages, as if the other didn't exist, at least as far as performance is concerned. At the same time, there is evidence that the bilingual cannot suppress the activation of the languages she handles, even if the context of the task is strictly monolingual. This brings up the question of the connection between the two lexica. If we assume the existence of a unified lexicon independent of the language (e.g. Schoonbaert, Duyck, Brysbaert & Hartsuiker, 2009), this question translates the way in which the words from the different languages are coded in the common lexico-semantic architecture.

The issue of the connection within and through languages is obviously related to morphological representation and processing. However, a review of the literature reveals that the role of morphological representation in the bilingual lexicon has been somehow neglected, despite the wealth of experimental data demonstrating morphological effects for a variety of language combinations (e.g., Kirsner, Lalor & Hird, 1993; Sánchez-Casas & García-Albea, 2005; Voga, Anastasiadis-Symeonidis & Giraudo, 2014; Voga, 2015; 2017). A review of the literature also shows that the anglo-centric view of language has somehow prevented the emergence of a morphological account of bilingual competence (i.e. morphology as a factor organising the lexicon), at least as far as processing is concerned.

Given that the organisation of the L1 lexicon is highly likely to be qualitatively similar to that of the L2, as Bybee suggests ('morphology transcends languages', Bybee, 1985), the present study considers not only bilingual speakers (L1: Greek; L2: French), but also (more or less) monolingual speakers of Greek. The experiments reported below aim to compare the two different populations with respect to processing of a special category of words, i.e., among the great variety of loan words borrowed from French and Anglo-American (Anastasiadis-Symeonidis, 1994), prefixed words with a non-existing stem (see below).

Our research aims to fill a gap in psycholinguistics: since the 90s, it has been acknowledged (e.g. Altenberg & Cairns, 1984) that the access to the bilingual lexicon is non-selective. Today, cognitive psychologists working both on production and lexical access, mostly refer to

language co-activation (e.g., van Hell & Tanner, 2012), which reflects the fact that both languages are simultaneously activated. However, one of the problems with ‘lexical co-activation’ in processing/lexical access (e.g. Casaposa & Duñabeitia, 2016) is that very often it refers to purely visual, orthographic, low-level factors, which are not very informative for the organisational principles of the lexicon/lexica under study. Lexical access approaches, using on-line processing protocols are mainly (but not mandatorily, see, e.g. Voga & Giraudo, 2017) *bottom-up* and consequently tend to attribute the primacy to form and visual factors. This kind of approach fails to address important issues about the mental lexicon, such as the role of central factors (semantics, morphology), as several authors underline (Baayen, 2014; Voga & Giraudo, 2017).

On the other hand, *top-down* approaches (e.g. RHM, Revised Hierarchical Model, Kroll & Stewart, 1994; De Groot, Borgwaldt, Bos & van den Eijnden, 2002 ; for the latest version, Kroll, van Hell, Tokowicz & Green, 2010) using production protocols, focus on characteristics of conceptual nature, thus neglecting the level in which co-occurrences between form and meaning (Bybee, 1985; 1988; Sánchez-Casas & García-Albea, 2005) are represented. *Top-down* protocols seem difficult to apply to particular categories of words, e.g. morphologically complex words, as well as to the study of cross-linguistic differences or similarities (e.g. Dijkstra, Miwa, Brummelhuis, Sappelli & Baayen, 2010).

To resume, both approaches, *bottom-up*, more compatible with the idea of a unified lexicon independent of the language (e.g. Schoonbaert et al. 2009), and *top-down*, claiming one lexicon for the L1 and another one for the L2 (RHM), when strictly implemented and interpreted, lead to an overreliance on general and mandatory processes. Consequently, they tend to neglect what we could call the ‘materialities’ of the language or languages taken into account (expression coined by Professor P. Sauzet). The role of these ‘materialities’ is however crucial, in the extent that particular characteristics of the materials studied can inform us on the connections between words, within and through languages, as well as on their strength.

We examine a special category of Greek words, which is formed by loan words from French (mostly) and English, perfectly integrated in the Greek lexicon, such as ρεαλιστής /realistis/, ‘realist’ (Corbin, 1987). These words have a much more important frequency (82.700 occurrences in Google for ρεαλιστής), compared to the closer Greek equivalent, e.g. πραγματιστής /pragmatistis/, ‘pragmatist’ (15.400 occurrences in Google), when it exists. Although these words have a small morphological family, (e.g. ρεαλισμός /realismos/, ‘realism’), their base does not mean anything, in our example, the ‘base’ ρεαλ- real- does not activate any lexeme in Greek, except the name of the famous football team (Real Madrid). There is thus a difference between these non-existent bases and bound-stems, since bound-stems activate the base-lexeme (for French bound stems, Giraudo & Voga, 2016), whereas there is no base such as ρεαλ- real- in Greek. Most of these words are naturally cognates (only cognates were selected for this study among this category), but not exactly in the same way as porte – πόρτα /porta/ ‘door’ or centre – κέντρο ‘center’ which has an existing base at the centre of the morphological family. At the same time, they constitute Greek words, both from a use and a lexicographic (they exist in the dictionary) point of view. These words thus constitute items of an intermediate type, somewhere in-between the Greek and the French/English lexicon. In two masked priming experiments with different groups of participants, we compare these words to two categories of Greek words, cognates (to French words) and non-cognates, in order to study the morphological priming pattern between our three categories of constructed words: cognates without base; cognates with Greek base known to induce robust identity and morphological effects (e.g. Voga & Grainger, 2007; Voga, 2015; 2017); non-cognates (see Table 1).

## 2. The study

### 2.1. Subjects.

- Two different groups are (currently) tested:
- a) In Exp. 1 (cross-language, cross-script priming) a bilingual Greek (L1) – French (L2) group ( $N=30$ ), having spent a certain amount of time in France (usually between 1 and 4 years). They are all students at the University Montpellier III and they generally have an advanced diploma in French as a second language before arriving to France.
  - b) In Exp. 2, a monolingual Greek group ( $N=30$ ), to the greatest extent possible, given that the vast majority of Greeks (below 50) have learned English and/or French (to a lesser extent), not only in school, but also in small private schools called *frontistiria*. These participants are selected because they don't have any relevant diplomas, or, if they do, they obtained them a long time ago and they have never really practiced English or French.

### 2.2. Stimuli.

Both groups are tested in all three categories of stimuli. In the monolingual experiment (Exp. 2) the prime as well as the target are in Greek, and in the bilingual experiment (Exp. 1), the prime is in Greek and the target in French, i.e. cross-language, cross-script priming, in the L1 to L2 direction.

Three different types of target are used:

a) Loan-words without Greek base, such as *ρεαλιστής* /realistis/, ‘realist’; the majority are formed with the suffix *ιστής* –iste ‘-ist’. These words share an important phonological overlap and mean exactly the same thing in Greek and in French, i.e. they are cognates for the bilingual experiment. Note however that, given the alphabetic difference between the two languages operating as a ‘language cue’ (e.g. Gollan, Forster & Frost, 1997; Author 2015; 2016) the processing system will be directly orientated towards the appropriate lexicon. Consequently, our prediction is that the effects will be of important amplitude, possibly of the same amplitude as the cognate condition (b), at least for the translation conditions.

b) Cognates which are Greek words that have been borrowed from English and French, such as *μαγικός* /magikós/ ‘magical’. These words do not necessarily have the same suffix as in (a), and are not always of the same grammatical category. They have a base word at the centre of the morphological family, which was kept as small as possible, in order to avoid differences in the pattern of effects due to the different MF sizes. More importantly, this category of cognates has been shown to induce robust identity and morphological effects (Voga, 2015). It thus represents the category of materials for which maximum facilitation effects are predicted for the bilingual experiment.

For the monolingual experiment, the difference between (a) and (b), should spring from the activation of the base, since words in (b) have an existing base-word while those in (a) do not. The comparison between (a) and (b) will allow us to see if and to what extent the non-existing stem suffixed words lead to morphological facilitation. Do they lead to ‘full’ activation, such as the one (normally) induced by (b)?

c) Non-cognate words with a similar suffix than in (a), referring to a person, e.g. *λαϊκιστής* /laikistis/ ‘populist’, *δανειστής* /danistis/ ‘borrower’. These words, exactly as in (b), have an existing base in Greek and should thus induce effects equivalent to (b) in the monolingual experiment (priming from L1 to L1), but lesser effects in the bilingual experiment, given that they are non-cognates. We must point out however, that, following several previous investigations in cross-script, cross-language priming (e.g. Gollan, Forster & Frost, 1997; Voga & Grainger, 2007), significant effects (both translation and morphological) are expected here. This category aims to test for series effects, given that it is crucial to distinguish the activation springing from the ‘base’ from that springing from the suffix.

Materials for all three categories are balanced for frequency.

The above materials are tested in three priming conditions: translation (or identity for the monolingual experiment), morphological and unrelated. Table 1 gives examples for the materials used in the different conditions (for the bilingual experiment). The complete results are not available yet.

**2.3. Procedure.** The prime duration for both experiments is of 50ms, which is the duration for which robust morphological effects, both monolingual and bilingual are found. The experiment ran with the help of the DMDX program (Forster & Forster, 2003) that presents experimental items and gives reaction times for every answer. The task was lexical decision (Yes/No), in the L1 for the monolingual experiment, and in L2 for the bilingual one (the participants are unaware of the existence of the prime, as always in masked priming protocols).

	Primes		
Target	Translation/identity	Morphological	Unrelated
réaliste	ρεαλιστής /realistik/ ‘realist’	ρεαλισμός /realismos/ ‘realism’	υπόθεση /ypóthesi/ ‘supposition’
magique	μαγικός /magikós/ ‘magical’	μαγεία /magíá/ ‘magic’	έλεγχος /éleghos/ ‘control’
populiste	λαϊκιστής /laikistik/ ‘populist’	populisme /laikismós/ ‘populism’	κύκλωμα /kyklooma/ ‘circuitry’

**Table 1** — Stimuli sample for the bilingual experiment. In the monolingual experiment, all stimuli are in Greek.

### 3. Discussion

The results will be discussed with respect to the nature and the coding of the morphological family in the bilingual and the monolingual lexicon: do non-existing stems such as *real-* induce morphological priming? Does mastery of the L2 (French) induce greater effects for the experimental items of (b)? Does the monolingual speaker exhibit effects for all categories or only for those having a Greek stem? In other words, does the monolingual speaker need to have a base-stem in order to establish a morphological connection between the morphological relatives, e.g. ρεαλιστής /realistik/ ‘realist’ and ρεαλισμός /realismos/ ‘realism’? If there is facilitation between the words sharing an inexisting stem, does it come from the ‘base’, or from the morphological series?

### References

- ALTENBERG E. P. & CAIRNS H. S. (1984). The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behavior* 22. 174-188.
- ANASTASSIADIS-SYMEONIDIS, A. (1994). Emprunt néologique en grec moderne. Analyse morphophonologique des emprunts directs au français et à l'anglo-américain. [Νεολογικός δανεισμός της νεοελληνικής. Άμεσα δάνεια από τη γαλλική και αγγλοαμερικανική. Μορφοφωνολογική ανάλυση]. Thessalonique.
- BAAYEN, H. (2014). Experimental and psycholinguistic approaches to studying derivation. In Lieber Rochelle & Štekauer Pavel (eds.), *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press.

- BYBEE, J. L. (1985). *Morphology: A study of the relation between meaning and form.* Amsterdam: John Benjamins.
- BYBEE, J. (1988). Morphology as lexical organisation. In M. Hammond & M. Noonan (eds.), *Theoretical Morphology. Approaches to modern linguistics*, 119-142. San Diego: Academic Press.
- CASAPONSA, A. & DUÑABEITIA, J. A. (2016). Lexical organization of language-ambiguous and language-specific words in bilinguals. *The Quarterly Journal of Experimental Psychology*, 69(3), 589-604.
- CORBIN, D. (1987). *Morphologie dérivationnelle et structuration du lexique.* Tübingen, Max Niemeyer Verlag. Villeneuve d'Ascq: Presses Universitaires de Lille 1991.
- DE GROOT, A. M. B. BORGWALDT S. BOS M. & VAN DEN EIJNDEN E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language* 47. 91-124.
- DIJKSTRA, T., MIWA, K., BRUMMELHUIS, B., SAPPELLI, M. & BAAYEN, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and language*, 62 (3), 284-301.
- DIJKSTRA, T. MOSCOSO DEL PRADO MARTIN F. SCHULPEN B. SCHREUDER R. & BAAYEN R. H. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes*, 20, 7-41.
- FORSTER, K. & FORSTER J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavioral Research Methods: Instruments & Computers* 35. 116-124.
- GIRAUDET, H. & VOGA, M. (2016). Words matter more than morphemes: An investigation of masked priming effects with complex words and nonwords. *Italian Journal of Linguistics* 28.1: 49–78.
- GOLLAN, T. FORSTER K. I. & FROST R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition* 23. 1122-1139.
- KIRSNER, K. LALOR E. & HIRD K. (1993). The bilingual lexicon: Exercise meaning and morphology. In R. Schreuder & B. Weltens (eds.), *The bilingual lexicon*. Amsterdam: John Benjamins.
- KROLL, J. VAN HELL J. TOKOWICZ N. & Green D. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language & Cognition* 13. 373-381.
- KROLL, J. F. & STEWART E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* 33. 149–174.
- SÁNCHEZ-CASAS, R. & GARCÍA-ALBEA J. E. (2005). The representation of cognate and noncognate words on bilingual memory: Can cognate status be characterized as a special kind of morphological relation? In J. F. Kroll y A. M. B. De Groot (eds.), *Handbook of Bilingualism: Psycholinguistic Approaches*. 226-250. New York: Oxford University Press.
- SCHOONBAERT, S., DUYCK, W., BRYSBERT, M. & HARTSUIKER, R. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings, *Memory and Cognition* 37(5), 569-586.
- VAN HELL, J. & TANNER, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, 62(2), 148-171.
- VOGA, M. & ANASTASSIADIS-SYMEONIDIS, A. & Giraudo, H. (2014). Does morphology play a role in L2 processing? Two masked priming experiments with Greek speakers of ESL. *Linguisticae Investigationes*, 37 (2). 338-352.
- VOGA, M. & GRAINGER J. (2007). Cognate Status and Cross-script Translation Priming. *Memory and Cognition*, 35, 938-952.

- VOGA, M. (2015). Towards a supra-lexical representation of morphology in the bilingual lexicon: evidence from Greek-French cognates. [Προς μια υπερ-λεξική αναπαράσταση της μορφολογίας στο δίγλωσσο νοητικό λεξικό: δεδομένα από ελληνο-γαλλικά διαγλωσσικά ομότυπα]. *Studies in Greek Linguistics*, 35, 106-130.
- VOGA, M. (2017). Form, morphological structure and representation in the bilingual lexicon: evidence from masked priming with Greek-French cognates and non-cognates [Μορφή, μορφολογική δομή και αναπαράσταση στο δίγλωσσο νοητικό λεξικό: Μεταφραστική προτεραιοποίηση με ελληνογαλλικά μεταφραστικά αντίστοιχα]. *Studies in Greek Linguistics*, 37, 129-152.
- VOGA, M. & GIRAUDO, H. (2017). Word and beyond-word issues in morphological processing. *Word Structure*, 10.2 (2017): 235–254.