International Symposium of Morphology (ISMo) 2019

Université de Paris

Campus Rive Gauche, Amphithéâtre Buffon
15 rue Hélène Brion, 75013 Paris

Programme and Abstracts

Berthold Crysmann and Florence Villoing

25–27 September 2019

Contents

4

Poster session A:

Wed 24 September 15:00–16:00   **121**

Part I

Programme

Schedule

Wednesday, 25 September 2019

12:30 *Registration*

14:00 Keynote lecture 1 (chair: Florence Villoing)

Rochelle Lieber *The semantics of -ing: eventivity, quantification, aspect*

15:00 *Poster session A*

16:00 *Coffee*

16:30 Session 1 (chair: Nabil Hathout)

16:30 Gergana Popova and Andrew Spencer *Verbal periphrasis in Bulgarian*

17:00 Natalia Bobkova and Fabio Montermini *The interplay of phonological constraints in the construction of Russian denominal adjectives*

17:30 Sascha Gaglia *The direction of analogical extensions in the verbal roots of Old French and Old Florentine Italian: a corpus study*

18:00 *Cocktail*

Thursday, 26 September 2019

09:00 *Registration*

09:30 *Keynote lecture 2* (chair: Berthold Crysmann)

Vito Pirrelli *Investigating inflection as a complex system*

10:30 *Coffee*

11:00 *Session 2* (chair: Delphine Tribout)

11:00 Stéphanie Lignon, Fiammetta Namer, Nabil Hathout and Mathilde Huguin *When sarkozysation leads to the hollandade, or the rejection of phonological well-formedness constraints by anthroponym-based derived words*

11:30 Bernard Fradin *The lexicon beyond lexemes*

12:00 Alice Missud and Florence Villoing *French -age suffixation versus verb to noun conversion: quantitative approaches on surface and underlying properties*

12:30 *Lunch*

14:00 *Session 3* (chair: Sebastian Fedden)

14:00 Olivier Bonami, Matías Guzmán Naranjo and Delphine Tribout *The role of morphology in gender assignment in French*

14:30 Dimitra Melissaropoulou *Accounting for morphological complexity vs. simplification in situations of language contact: evidence from Cappadocian Greek*

15:00 *Poster session B*

16:00 *Coffee*

16:30 *Session 4* (chair: Louise Esher)

16:30 Jean-Pierre Koenig and Karin Michelson *Conversion, structured inflection, and the ontological/semantic organization of the lexicon in Oneida*

17:00 Dimitri Leveque and Thomas Pellard *Description of verbal morphology of Asama: a realizational and implemented approach*

17:30 Berthold Crysmann *Morphotactic dependencies in Yimas: a constructional approach*

Friday, 27 September 2019

09:00 *Session 5* (chair: Enrique Palancár)

>**09:00** Jeremy Pasquereau and Patricia Cabredo Hofherr *Multiple event marking in the Seri verbal paradigm*

>**09:30** Stefan Hartmann *The curious case of wandering case morphemes*

>**10:00** Selena Rorberi and Claudia Marzi *Modelling the interaction of regularity and morphological structure: the case of Russian verb inflection*

10:30 *Coffee*

11:00 *Session 6* (chair: Fabio Montermini)

>**11:00** Jan Radimsky *Are French NNs variants of N-PREP-N constructions?*

>**11:30** Daniel Gutzmann and Katharina Turgay *In defense of the "phrasal compounds as quotations" thesis*

>**12:00** Masaharu Shimada and Akiko Nagano *Word formation with loanwords: A case of "Japanese English"*

12:30 *Lunch*

14:00 *Session 7* (chair: Patricia Cabredo-Hofherr)

>**14:00** Louise Esher and Jean Léo Léonard *A Paradigm Function Morphology approach to Moksha objective conjugation*

>**14:30** Angelo Costanzo *"Prestigious plurals" and conjugational class variation*

>**15:00** Gilles Boyé and Gauvain Schalchli *A quantitative comparison between word-formation & inflection: A look at paradigms in French*

15:30 *Coffee*

16:00 *Session 8* (chair: Fiammetta Namer)

>**16:00** Serena Dal Maso and Sabrina Piccinin *Formal and semantic transparency in L1 and L2 processing*

>**16:30** Chiara Melloni and Maria Vender *Playing with nonwords: morphological skills in dyslexia*

>**17:00** Madeleine Voga *Competition in the bilingual lexicon and cross-language priming asymmetries: A morphological connection?*

17:30 *Farewell*

Poster sessions

Session A: Wednesday, 25 September, 15:00 – 16:00

- Grigory Agabalian *Deux propositions pour la description sémantique des noms de systèmes d'idées en -ISME*

- Anna Anastassiadis-Symeonidis and Maria Mitsiaki *Revisiting inflectional morphology: Towards a new paradigm for teaching nominal inflection in Modern Greek as a second language*

- Xavier Bach and Pavel Štichauer *Auxiliary selection in Romance and inflectional classes*

- Matthew Baerman, Jeremy Pasquereau and Carolyn O'Meara *Incremental realization*

- Gladys Camacho Rios *A re-analysis of verb morphology in South Bolivian Quechua A case study of the Uma Piwra rural variety*

- Gasparde Coutanson *Postverbal liaisons in traditional songs: a morphological reanalysis ?*

- Magdalena Derwojedowa *Integration of comparative degree into the adjective paradigm*

- Maximilien Guérin, Louise Esher, Jean-Léo Léonard and Sylvain Loiseau *Modelling diasystemic inflexion: Verb morphology in the Croissant linguistique*

- Petr Kos and Jana Kozubíková Šandová *Predicting cells in word-formation paradigms – a case study*

- Julie Marsault *The prefixal template of Umoⁿhoⁿ*

- Franz Rainer and Sara Matrisciano *Romance compounding and language contact: Origin and spread of the pattern vert bouteille 'bottle-green'*

- Neige Rochant *When a causative could hide a plural marker: A quest for the origins of the causative in Andi (Nakh-Daghestanian)*

- Pavol Stekauer, Livia Kortvelyessy and Pavol Kacmar *On the influence of creativity upon the formation of complex words*

- Yoko Sugioka *Event/entity polysemy and head identification in deverbal compounds*

Session B: Thursday, 26 September, 15:00 – 16:00

- Alexandra Bagasheva, Livia Kortvelyessy, Pavol Stekauer, Salvador Valera and Jan Genci *Cross-linguistic research into derivational networks*

- Gilles Boyé *Stem spaces in abstractive morphology: A look at defectiveness in French conjugation*

- Bien Dobui *Derivational morphemes in Xochistlahuaca Amuzgo*

- Edwige Dugas, Pauline Haas and Rafael Marín *French Denominal Verbs: from countability to aspect*

- Noam Faust *Paradigm migration in the QoTeT verbs of Modern Hebrew*

- Sebastian Fedden and Greville Corbett *The continuing challenge of the German gender system*

- Rusudan Gersamia and Alexander Rostovtsev-Popiel *Morphology, Semantics, and Pragmatics of Negative Rhetorical Questions in Megrelian*

- Matías Guzmán Naranjo *Entropy, analogy and paradigm structure*

- Lior Laks *Competing vowels in feminine formation: Evidence from Hebrew and Jordanian Arabic*

- Nicola Lampitelli, Paolo Roseano and Francesc Torres-Tamarit *Non-linear morphology in Romance: the case of vowel length in Friulian verbs*

- Ryohei Naya *The Status of Affixes and the New Words by -ment in Present-Day English*

- Andrew Spencer *Uninflecting and uninflectable lexemes: implications for paradigm structure*

- Malgorzata Sulich-Cowley *The influence of 'absence' on Sanskrit morphology – the case of negative compounds*

- Elena Voskovskaia *Composés N-N et N-A dans la littérature française du 17e au 20e siècle : la productivité morphologique*

Committees

Programme committee

We would like to express our deepest gratitude to the members of the programme committee, as well as six anonymous reviewers, for their expertise on the papers submitted to this conference. Without you, we would not have been able to put this programme together.

Berthold Crysmann and Florence Villoing (ISMo 2019 co-chairs)

Paolo Acquaviva
Dany Amiot
Anna Anastassiadis-Symeonidis
Giorgio Francesco Arcodia
Peter Arkadiev
Mark Aronoff
Jenny Audring
Sacha Beniamine
Olivier Bonami
Geert Booij
Gilles Boyé
Dunstan Brown
Basilio Calderone
Bruno Cartoni
Berthold Crysmann (chair)
Georgette Dal
Serena Dal Maso
Walter De Mulder
Laura Downing
Louise Esher
Sebastian Fedden
Bernard Fradin
Zoe Gavriilidou
Nicola Grandi
Stefan Hartmann
Nabil Hathout
Claudio Iacobini
Marianne Kilani-Schoch
Jean-Pierre Koenig

Stéphanie Lignon
Maria-Rosa Lloret
Rafael Marin
Francesca Masini
Louise McNally
Chiara Melloni
Fabio Montermini
Fiammetta Namer
Muriel Norde
Albert Ortmann
Enrique Palancar
Vito Pirrelli
Ingo Plag
Franz Rainer
Angela Ralli
Davide Ricca
Andrew Spencer
Pavel Stichauer
Dejan Stosic
Gregory Stump
Anna Maria Thornton
Delphine Tribout
Jesse Tseng
Kristel Van Goethem
Florence Villoing (chair)
Madeleine Voga
Géraldine Walther

Local organisers

Gabriel Aguila-Multner
Jeff Alexandre
Aixiu An
Olivier Bonami (local chair)
Maria Copot
Berthold Crysmann
Angélique Jaber Suzanne Lesage
Saïda Loucif
Alice Missud
Alexandre Roulois (webmaster)
Vincent Segonne

ISMo standing committee

Dany Amiot
Olivier Bonami
Gilles Boyé
Georgette Dal
Bernard Fradin
Hélène Giraudo
Nabil Hathout
Stéphanie Lignon
Fabio Montermini
Fiammetta Namer
Delphine Tribout
Florence Villoing

Acknowledgements

Part II

Abstracts

Keynote lectures

The semantics of *-ing*: eventivity, quantification, aspect

Rochelle Lieber

University of New Hampshire

In this paper I consider the lexical semantic contribution of the affix *-ing* (as in *the doctor's examining of the patient*) to deverbal nominalizations. I will first present data from a corpus-based study (Andreou & Lieber, 2019) showing – contrary to claims in the literature (Grimshaw 1990, Biese 1941, Langacker 1991, Brinton 1995, Alexiadou 2001, among others) – that the suffix *-ing* can convey both referential and eventive readings, can exhibit both mass and count quantification, and that when eventive, can convey both bounded and unbounded aspect. I will show that *-ing* in fact does not differ from deverbal nouns formed by conversion in the breadth of readings that it can support. Deverbal nouns formed by conversion also allow both referential and eventive, mass and count, bounded and unbounded readings. Indeed, I will show that the formal means of nominalization does not determine eventivity, quantification, or aspect, but rather than these facets of meaning are largely determined by context. The question then arises of how we can model the semantics of nominalizers like *-ing* (or conversion) in syntactic or morphological theory. My argument will be that nominalizers cannot be analyzed as "rigid designators," as assumed in Borer's (2013) exoskeletal framework or in most versions of Distributed Morphology that I am aware of, but rather that a theoretical treatment must allow for substantial under-specification of affixal semantics as well as for mechanisms that allow context to determine key aspects of the nominalization's reading, as is possible within the Lexical Semantic Framework (LSF) of Lieber (2004, 2016).

References

Alexiadou, Artemis 2001. *Functional structure in nominals*. Amsterdam: John Benjamins.

Andreou, Marios & Rochelle Lieber (2019). Aspectual and quantificational porperties of deverbal conversion and *-ing* nominalizations: The power of context. *English Language and Linguistics,* in press.

Biese, Y 1941. Origin and development of conversions in English. *Annales Academiae Scientiarum Fennicae B XLV-2*: 1-495.

Borer, Hagit 2013. *Taking form: Structuring Sense vol 3*. Oxford: Oxford University Press.

Brinton, Laurel 1995. The aktionsart of deverbal nouns in English. In Pier Marco Bertinetto, Valentina Bianchi, James Higginbotham, and Mario Squartini (eds.), *Temporal Reference, Aspect, and Actionality*, *vol. I, Semantic and Syntactic Perspectives*. Torino: Rosenberg and Sellier, 27-42.

Grimshaw, Jane 1990. *Argument structure*. MIT Press.

Langacker, Ronald 1991. *Foundations of Cognitive Grammar. Vol. 2 Descriptive Applications.* Stanford: Stanford University Press.

Lieber, Rochelle 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

Lieber, Rochelle 2016. *English Nouns: The Ecology of Nominalization*. Cambridge: Cambridge University Press.

Investigating inflection as a complex system

Vito Pirrelli

Istituto di Linguistica Computazionale, Consiglio Nazionale delle Ricerche, Pisa Italy

From a cross-linguistic perspective, different inflection systems appear to apportion word processing costs differently, depending on when and where, in the full form, morpho-lexical and morpho-syntactic information is encoded. The resulting balance is the outcome of an interaction between form frequency and morphological productivity, responding to basic communicative requirements. Big families of stem-sharing inflected forms constitute the productive core of an inflection system. This core is easy to learn, as it requires memorization of one stem only, with all inflected forms being redundantly built upon it. Unsurprisingly, generalizable paradigms are less sensitive to token frequency effects, and tend to be located in the long, low-frequency tail of the Zipfian distribution of word forms. In contrast, the head of the Zipfian distribution mostly contains small families of alternating and possibly suppletive stems, which, however shorter, morpho-phonologically simpler and easier to process, require high token frequency to be learned and resist pressure towards regularization.

Of late, considerable converging computational evidence has accrued to support this picture (Ackerman and Malouf 2013, Balling & Baayen 2012, Blevins et al. 2017, Bonami & Beniamine 2016, Marzi et al. 2018, Marzi et al. in press, Pirrelli 2018). In the talk, I will show how the dynamic tension between ease of learning and ease of processing can shape and structure the inflection systems of typologically different languages. In the end, each language (and arguably each individual learner) is likely to strike a different balance, which nonetheless falls within a reasonably tight range of variation, bounded by a few learnability and processability constraints. This suggests that full investigation of morphological systems will likely benefit from the use of basic concepts from the toolkit of complexity theory in biological networks, such as emergence, non-linearity and self-organization.

References

Ackerman, F. and Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language* 89, 429–464.

Balling, L. and Baayen, R.H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition* 125, 80–106.

Blevins, J. P., Milin, P. and Ramscar, M. (2017). The zipfian paradigm cell filling problem. In F. Kiefer, J. P. Blevins and H. Bartos (Eds.) *Morphological Paradigms and Functions* (Leiden, Germany: Brill), 141-158.

Bonami, O. and Beniamine, S. (2016). Joint predictiveness in inflectional paradigms. *Word Structure* 9, 156–182.

Marzi, C., Ferro, M. and Pirrelli V. (in press). A Processing-oriented Investigation of Inflectional Complexity. *Frontiers in Communication*.

Marzi, C., Ferro, M., Nahli, O., Belik, P., Bompolas, S. and Pirrelli V. (2018). Evaluating Inflectional Complexity Crosslinguistically: a Processing Perspective, In: *LREC 2018 - Eleventh International Conference on Language Resources and Evaluation* (Miyazaki, Japan, 7-12/05 2018). Proceedings, article 745. European Language Resources Association (ELRA), 2018.

Pirrelli, V. (2018). Morphological theory and computational linguistics. In J. Audring and F. Masini (Eds.) *The Oxford Handbook of Morphological Theory* (Oxford, UK: Oxford University Press), 573-593.

Oral presentations

The interplay of phonological constraints in the construction of Russian denominal adjectives

Natalia Bobkova, Fabio Montermini

CLLE-ERSS, CNRS & Université de Toulouse Jean Jaurès

1 Introduction

The main focus of our paper are denominal adjectives in Russian, that are mainly considered from the phonological point of view. In particular, the paper deals with such issues as the selection of the stem of the base lexeme (which can undergo phonological modifications), affixal allomorphy and the competition between different formal strategies for the same syntactico-semantic construction. As far as the last issue is concerned, in particular, our research is carried on within a Construction Morphology framework (Booij 2010 and subsequent; Audring & Masini 2019), which we complement with a thematic morphology approach in the line of what has been proposed by Roché & Plénat (2014) and Plénat & Roché (2014) for French, and with a constraint-based approach, proposed by the same authors and Hathout (2009), among others. We propose to call the approach we adopt for the analysis of our data Constraint-Based Construction Morphology.

2 Constraint Based Construction Morphology

The approach in question aims at modelling the phonological properties of morphological constructions. Its main tenets are the following:

- morphological constructions operate on stems, which correspond to the phonological representation of a lexeme; more precisely they link a stem (or a collection of stems) with another stem (or collection of stems);
- phonologically, the exponent of a morphological construction (be it affixal or not) is viewed as a set of constraints on the phonological shape its outputs should have. These constraints are hierarchically ordered, construction-specific, and interact with other, more general, constraints.

In the lines of what has been proposed by Roché & Plénat (2014) and Plénat & Roché (2014), we consider that, phonologically speaking, morphological derivation mainly consists in an operation (e.g. an affixation) performed on a radical, a phonological object which is obtained from the stem of the base lexeme (cf. Roché 2010 on the distinction between stem and radical). In the most basic case, the stem of a lexeme and the radical of the derivative simply coincide (cf. Eng. DANCE ↔ DANCER). In other cases, however, the stem may undergo some manipulations in order to better fit the phonological constraints imposed by a derivational construction. These manipulations may correspond, on the one side, to the selection of a specific stem, either one which appears in other derivatives (e.g. Fr. PRIVATISME built on the stem /pʁivat/ of the adjective PRIVÉ 'private', like PRIVATISER), or which belongs to another lexeme within the same morphological family (e.g. Fr. PERSONNALISME, semantically built on the noun PERSONNE 'person', but formally built on the adjective PERSONNEL 'personal', cf. Roché 2009: 159 for details). On the other side, they may correspond to the creation of a specific radical through a phonological manipulation of the base stem. Plénat & Roché (2014), for instance, cite the case of Fr. GEEKARIAT ('the status of being a geek'), built on the

noun GEEK, for which the emergence of the sequence /aʁ/ as an interfix is due to the high number of lexemes in *-ariat* within the series of lexemes constructed with the suffix *-at*. The goal of all the operations described above is to better satisfy the output constraints a derivative is submitted to, which include purely phonological or prosodic constraints, but also lexical constraints that guarantee the maximal phonological homogeneity within a morphological family or a morphological series, as in the case of GEEKARIAT.

The radical thus constructed constitutes then the phonological object on which a morphological construction operates. In the line of what has been proposed by Montermini (2018), we consider that this operation corresponds, in fact, to a set of constraints on the form of the outputs of the construction. For instance, we may consider that the exponent of *-at* suffixation in French includes a constraint of the form *-ariat* > *-at*, which states that, all other things being equal, a derivative ending in *-ariat* is preferred over a derivative where *-at* is preceded by another sequence.

3 Russian denominal adjectives

Russian denominal adjectives, we claim, constitute a good testing ground for the model described above. In fact, Russian possesses various suffixes in order to derive adjectives from nouns, most of which possess more or less phonologically motivated variants, and which interact in various ways with other derivational affixes (including other denominal adjectival suffixes). Although affixal competition (in the sense of Aronoff 2013) is not the focus of the present work, we acknowledge that it could be an important parameter to be considered. It is likely, in fact, that the selection of exponents has strict interactions with the mechanisms described above. To state it simply, a suffix may be chosen over another because it is more compatible with a specific set of phonological constraints. We reserve this issue for further investigation (cf. however Hénault & Sakhno 2016 on affixal competition in adjectival derivation in Russian).

Russian denominal adjectivization constitutes a rich system, with a number of different exponents available, some of which display a high productivity rate. Dictionaries and grammars (for instance Švedova 1980) list a minimum of eight different suffixes, and this number increases to some tens if we include all phonological and morphological variants. Currently, only three suffixes appear to be productive for the construction of adjectives from nouns in Russian: *-ov*, *-sk* and *-n*,[1] for each of which we give an example in (1).

(1)   SLON   'elephant'   ↔   SLONOV(YJ)
      SOSED   'neighbour'   ↔   SOSEDSK(IJ)
      KOŽ(A)   'skin'   ↔   KOŽN(YJ)

Apart from *-ov*, whose only variation is due to phonological factors (namely the presence / absence of a softened – palatalized – consonant in the base stem), the other suffixes may appear in several variants, and in particular be preceded by various sequences, which have different statuses and origins, but all contribute to the definition of the exponent as a set of constraints, as described above. Figure 1 presents an attempt to formalize the different variants encountered with *-sk* and *-n* which are considered productive in dictionaries and grammars.

---

[1] Suffixes are conventionally represented under their 'bare' form, without the corresponding inflectional endings; full lexemes are represented in their citation form, with inflectional endings in brackets.
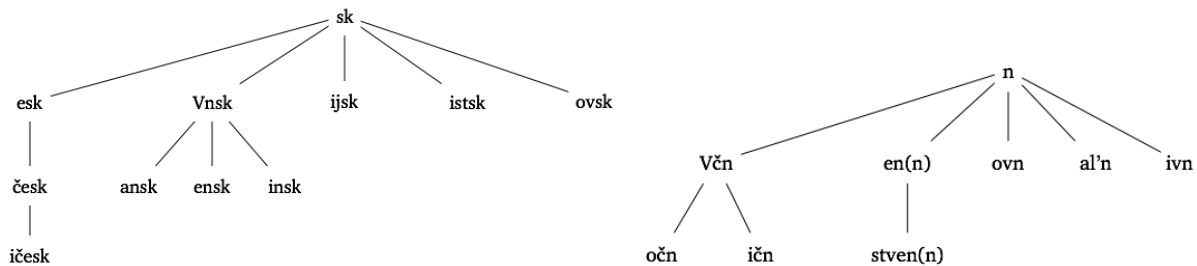
Figure 1: attested variants of *-sk* and *-n* suffixes

What we call a 'variant' corresponds to a surface form under which the corresponding affix appears in a set of derivatives. In (2) we give some examples of *-sk* derivatives displaying different variants of the suffix:

(2)   DRUG   'friend'   ↔   DRUŽESK(IJ)
       OBORON(A) 'defense'   ↔   OBORONČESK(IJ)
       OKEAN   'ocean'   ↔   OKEANIČESK(IJ)

According to the examples above, we might well consider that the three variants in question constitute a hierarchically ordered set of constraints on the form of the output having the following form: *-iček* > *-ček* > *-esk* > *-sk*. In its turn, this hierarchy interacts with other hierarchies connected with the same suffix. Crucially, the emergence of one of the variants in the hierarchy allows the derivative to fit some constraints, which may be either purely phonological or lexico-morphological: i) *-esk* appears in particular with palatal stems of lexemes which display 'hard' (non-palatalized) stems in their inflectional forms (like in DRUG vs. DRUŽESK(IJ)); ii) *-ček* appears in particular with bases containing specific suffixes or combining forms like *-ec, -čik, -ved*, etc. (like in PEREVODČIK 'translator' ↔ PEREVODČESK(IJ)); iii) *-iček* appears in particular in lexemes of foreign origin, and consequently also with lexemes containing specific suffixes / combining forms (e.g. *-ija, -izm, -ik*, etc.). [2]

   The same dynamics combining a hierarchy of preferred forms for a suffixal exponent and an interaction between phonological and morphological factors, including the form of the base, holds, we claim, in all other cases. The examples in (3) illustrate another such case:

(3)   GOSUDARSTV(O) 'state'   ↔   GOSUDARSTVENN(YJ)
       BOG   'god'   ↔   BOŽESTVENN(YJ)
       UM   'mind'   ↔   UMSTVENN(YJ)

These cases show the potential autonomy of *-stven(n)* as a variant of *-n*: in the case of GOSU-DARSTVENN(YJ), the sequence *-stv* (corresponding to a nominalizing suffix) already appears in

---

[2] Of course, the sequence *-ič-* has been historically (and is still) used for the adaptation of foreign lexemes in Russian, as are other sequences included in Figure 1 (*-istsk, -al'n*, etc.). However, we chose here to give all these forms the same status, regardless of their origin. The data in (2) show precisely the impossibility to distinguish them from 'native' Russian forms. Note moreover that some of the sequences in question (*-ič-, -ov-*) may precede either *-sk* or *-n*, which suggests that they should probably be attributed an interfix status. The precise status they have has no impact on the analysis we propose, however.

the base. In the case of BOŽESTVENN(YJ), it appears in a lexeme belonging to the same morphological family as the base (BOŽESTV(O) 'divinity'); here, we observe the same form / meaning mismatch than between PERSONNE, PERSONNEL and PERSONNALISME (see above). Finally, in the case of UMSTVENN(YJ), *-stven(n)* is added to the base as a full variant of the affix (a fact that is probably favoured by the lexicalisation of the 'simpler' form UMN(YJ) as a qualifying adjective meaning 'smart').

Our analysis will be realized by comparing two datasets extracted from a large corpus of Russian (http://www.ruscorpora.ru), one containing the 1,000 most frequent lexemes containing the suffixes mentioned, and a sample of 1,000 lexemes containing the same suffixes and having frequency 1 in the corpus. The comparison between the two datasets will allow to identify the constraints that are active in the productive use of morphological constructions by speakers, and to set them apart to the properties of derived lexemes that should be attributed to lexicalization factors.

References

Aronoff, Mark. 2016. Competition and the lexicon. In Annibale Elia, Claudio Iacobini, & Miriam Voghera (eds.), *Livelli di analisi e fenomeni di interfaccia. Atti del XLVII Congresso internazionale della Società di linguistica italiana*, 39–52. Roma: Bulzoni.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Hathout, Nabil. 2009. *Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie*. Habilitation à diriger des recherches. Toulouse: Université de Toulouse le Mirail.

Hénault, Christine & Sergueï Sakhno. 2016. Čem *supermarket-n-yj* lučše *supermarket-sk-ogo*? Slovoobrazovatel'naja sinonimija v russkix ad"ektivnyx neologizmax po dannym Interneta. In Branko Tošović & Arno Wonisch (eds.), *Wortbildung und Internet*. Graz: Institut für Slawistik.

Masini, Francesca & Jenny Audring. 2019. Construction Morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 365–389. Oxford: Oxford University Press.

Montermini, Fabio. 2018. Les affixes dérivationnels ont-ils des allomorphes? Pour une modélisation de la variation des exposants dans une morphologie à contraintes. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The Lexeme in Descriptive and Theoretical Morphology*, 423–465. Berlin: Language Science Press.

Plénat, Marc & Michel Roché. 2014. La suffixation dénominale en *-at* et la loi des (sous-)séries. In Sophie David, Sarah Leroy & Florence Villoing (eds.), *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux*, 47–74. Nanterre: Presses Universitaires de Paris Ouest.

Roché, Michel. 2009. Un ou deux suffixes ? Une ou deux suffixations ?. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 143–173. Saint-Denis: Presses Universitaires de Vincennes.

Roché, Michel. 2010. Base, thème, radical. *Recherches linguistiques de Vincennes* 39. 95–134.

Roché, Michel & Marc Plénat. 2014. Le jeu des contraintes dans la sélection du thème présuffixal. In Franck Neveu, Gabriel Bergounioux, Marie-Hélène Côté, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost (eds.), *Actes du 4ᵉ Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, 1863–1878. Paris: Institut de linguistique française.

Švedova, Natal'ja Ju. 1980. *Russkaja grammatika*. Moskva: Nauka.

The role of morphology in gender assignment in French

Olivier Bonami *Matías Guzman Naranjo* *Delphine Tribout*
U. de Paris   U. Düsseldorf   U. de Lille

Corbett (1991) played a central role in establishing that grammatical gender assignment is far from being arbitrary. Although there are usually few categorical generalizations, semantic, morphological, and phonological properties of a noun typically contribute to predicting its grammatical gender. While this is clearly established, much remains to be explored on the exact nature of the predictors and the way they interact. As a case in point, consider the situation in French. At least since Tucker et al. (1977), it is firmly established that all three types of predictions have a role to play, as exemplified in (1).

(1) a. Semantic prediction: Nouns referring to properties are overwhelmingly feminine.
  b. Morphological prediction: VN compounds denoting inanimates are all masculine.
  c. Phonological prediction: nouns ending in /jɔ̃/ are overwhelmingly feminine

While such generalizations are clearly correct, there are often correlations between them that make it nontrivial to establish their exact status. For instance, property nouns are also overwhelmingly derived from adjectives by suffixes uniformly outputting feminine nouns (*-ité*, *-eur*, etc.), and nouns ending in the sequence /jɔ̃/ are overwhelmingly formed using the derivational suffix *-ion*.

This abstract presents a quantitative study of gender assignment in French that aims at disentangling the role of morphology and phonology in gender assignment. We start from the observation that about one third of French nouns with a unique gender end in a derivational suffix (see below), and that the relevant suffixes are for the most part compatible with only one gender (Bonami & Boyé, 2019).[1] This suggests that the predictability attributed to phonology since Tucker et al. (1977) could to a large extent be attributable to morphology. To investigate this issue, we annotated by hand a sample lexicon of 3,683 nouns for their phonological and morphological properties. We then expanded on Sokolik & Smith (1992) and Matthews (2005) by training neural networks to learn gender assignment on the basis of phonological and/or morphological predictors. We conclude that morphology plays a subtle role in gender assignment: while phonology is a very good predictor of gender on its own, this is to a large extent due to the way derivational morphology shapes the phonotactic properties of the lexicon.

1   Data collection and annotation

The sample of nouns we used in this study was randomly selected among the nouns contained in the *Lexique 3* database (New et al., 2007), limiting attention to wordforms found in only one gender and lemmas with a frequency above 0.3 per million words. The morphological annotation proceeded as follows. In a first step, we relied on previously published manually curated lexica: 760 nouns found in the lexicon of simplex nouns presented in (Tribout et al., 2014) were tagged as simplex, and 1,019 nouns were tagged as instances of conversion, either because they were listed in Tribout's (2010) database, or because they were homographic to a verb or adjective. We then proceeded to annotate manually the remaining 1,904 nouns. There

---

[1] According to Bonami & Boyé (2019), at least 25% of all French nouns are common gender, i.e. come in pairs of homophonous masculine and feminine nouns, and this proportion is rising steadily in recent years. We disregard such cases in the present study.

was a double morphological annotation. On the one hand, we noted the type of outermost word-formation process (prefixation suffixation, compounding, etc.), and, in the case of affixation processes, the identity of the affix. On the other hand, we noted the outermost suffix, if any was present (non-suffixed nouns were annotated as '0'). This was motivated by the presumption that a suffix might be relevant to gender assignment even where suffixation is not the last operation to have applied: e.g. the feminine gender of *contre-proposition* 'counter-proposal' can be tracked down to the presence of the suffix *-ion*, despite the fact that prefixation of *contre* is the outermost process. In the end, our dataset contains 1,222 nouns ending in a suffix (henceforth 'suffixed nouns') and 2,461 nouns not ending in a suffix (henceforth 'unsuffixed nouns'). Finally, we added to the database phonological transcriptions and syllable boundaries as documented in the GLÀFF lexicon (Hathout et al., 2014).

2   Modelling

We want to explore two main questions: (i) to what degree is gender predictable for French nouns, and (ii) what are the roles of phonological and morphological factors in gender assignment. To answer these questions we train several multilayer perceptrons to predict gender based on morphological and phonological predictors.[2] For the phonological predictors we extracted the last three segments of the noun, the number of syllables and number of segments. As morphological predictor we used the annotated suffix, or 0 for unsuffixed nouns.

We trained the models using `caret` (Kuhn, 2008) with `MxNet` (Chen et al., 2015). The perceptrons had 3 hidden layers with 128, 4, 2 neurons, respectively.[3] The results reported for each model are the aggregated accuracy and kappa scores[4] of 10-fold cross-validation. Our model choice obeyed two main reasons. First, we wanted to keep models consistent across datasets. Second, multilayer perceptrons have been shown to perform well in similar gender/class assignment tasks (Matthews, 2005).

First we consider the whole dataset. Table 1 shows the result of three models: one model with only morphological predictors, one with only phonological predictors, and one model with both phonological and morphological predictors. These initial results clearly show that gender is highly predictable in French nouns. The model trained on morphological predictors only shows that a large portion of the variation is due to morphology. On the other hand, the model trained only on phonological predictors reached the same accuracy as the model using both morphological and phonological predictors. This effect is likely due to the fact that phonological predictors are a good proxy for morphological markers.

Since morphological gender assignment mainly happens on suffixed nouns, we now fit similar models focusing exclusive on these. Table 2 shows the results for this set of models. For this set of nouns, the suffix overwhelmingly determines the gender of the noun, and adding phonological to morphological information does not lead to any increase in accuracy. Nonetheless, because the phonological predictors are a good proxy for the morphological predictors, the model with only phonological predictors reaches a similarly high accuracy.

Next we turn to nouns without a suffix. Table 3 shows the results for this group of nouns.[5]

---

[2]We also tried to use animacy information as an extra predictor, based on the intuition that this might help discriminate e.g. inanimate abstract feminine nouns in *-eur* such as *blancheur* 'whiteness' from animate masculine agent nouns such as *menteur* 'liar'. As it turns out, in none of the conditions described below does the addition of animacy lead to a discernible improvement of model accuracy. Hence we report results for the simpler models without animacy.

[3]We tweaked the momentum and learning rate of each network to ensure convergence. All layers had relu activation. The models were trained using an Nvidia Titan Xp (this card was donated to us by the NVIDIA Corporation).

[4]This metric measures how much better than random chance are the results of the model.

[5]For this dataset we did not train models with morphological predictors.

Predictors	Morphology		Phonology		Both	
	Reference		*Reference*		*Reference*	
Prediction	F	M	F	M	F	M
F	629	32	1092	264	1139	278
M	879	2143	416	1911	369	1897
Accuracy	0.75		0.82		0.82	
95% Accuracy's CI	(0.74, 0.77)		(0.80, 0.83)		(0.81, 0.84)	
Kappa	0.44		0.61		0.63	

Table 1: Three models for the whole dataset.

Predictors	Morphology		Phonology		Both	
	Reference		*Reference*		*Reference*	
Prediction	F	M	F	M	F	M
F	631	18	609	47	639	20
M	22	551	44	522	14	549
Accuracy	0.97		0.93		0.97	
95% Accuracy's CI	(0.96, 0.98)		(0.91, 0.94)		(0.96, 0.98)	
Kappa	0.93		0.85		0.94	

Table 2: Three models for suffixed nouns.

The results in this table show that gender is highly predictable for nouns without a suffix, but the error rate is, as expected, much higher than the error rate for nouns with a suffix.

	Reference	
Prediction	F	M
F	500	269
M	355	1337
Accuracy	0.75	
95% Accuracy's CI	(0.73, 0.76)	
Kappa	0.43	

Table 3: Phonological prediction of unsuffixed nouns.

3   Discussion

The results of our modelling experiments paint a subtle picture of the role of morphology in gender assignment. On the one hand, explicit use of morphological information plays a minor role in accurate prediction of gender: on the whole dataset, a model relying on both morphology and phonology does not outperform a model relying on phonology only; and even on the subset of suffixed nouns, the gain in accuracy of taking explicit morphological information into account is quite limited. This leads to the speculation that speakers may not need to attend to morphological information to correctly assign gender. Only psycholinguistic experimentation will be able to tell whether they do.

On the other hand, morphology plays a crucial role in shaping the phonotactic makeup of the lexicon, in such a fashion that phonological prediction of gender is much more accurate for suffixed than for unsuffixed nouns. Examination of conditional probability distributions between various variables estimated from our dataset help understand the causes of this situation.

Here we use conditional entropy as a rough indication of interpredictibility between variables. First, while the ultimate cause of predictability of gender in suffixed nouns is the fact that suffixes assign gender almost categorically ($H(\texttt{gender}|\texttt{suffix}) = 0.04$), the suffix itself is quite well predicted by the three last segments of the words ($H(\texttt{suffix}|\texttt{last\_3\_segments}) = 0.18$); hence, knowledge of the suffix adds little to knowledge of the last three segments when predicting gender ($H(\texttt{gender}|\texttt{last\_3\_segments}) = 0.06$; $H(\texttt{gender}|\texttt{last\_3\_segments}, \texttt{suffix}) = 0.02$), since phonology alone already approaches categorical prediction. Second, the final substrings of suffixed words and unsuffixed words are different enough on average that final substrings are a decent predictor of whether a word is suffixed ($H(\texttt{suffixed}|\texttt{last\_3\_segments}) = 0.30$). This indicates that unsuffixed nouns which happen to end in a sequence that could be a suffix are not frequent enough to strongly impair prediction of gender by phonology, and helps explain the absence of a contribution of morphological information to accuracy of prediction on the whole dataset.

One conclusion of this study is that explicit morphological knowledge plays a distinct but limited predictive role in the case of suffixed nouns. The limited amplitude of that role may be due to the fact that our models do not take into account any semantic information: it may be the case that semantic and phonological information are jointly sufficient to reach optimal accuracy. We will investigate that issue in the near future.

References

Bonami, Olivier & Gilles Boyé. 2019. Paradigm uniformity and the French gender system. In Matthew Baerman, Oliver Bond & Andrew Hippisley (eds.), *Perspectives on morphology: Papers in honour of Greville G. Corbett*, Edinburgh: Edinburgh University Press.

Chen, Tianqi, Mu Li, Yutian Li, Min Lin, Naiyan Wang, Minjie Wang, Tianjun Xiao, Bing Xu, Chiyuan Zhang & Zheng Zhang. 2015. Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv preprint arXiv:1512.01274* .

Corbett. 1991. *Gender.* Cambridge: Cambridge University Press.

Hathout, Nabil, Franck Sajous & Basilio Calderone. 2014. GLÀFF, a large versatile French lexicon. In *Proceedings of LREC 2014,* .

Kuhn, Max. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software, Articles* 28(5). 1–26.

Matthews, Clive A. 2005. French Gender Attribution on the Basis of Similarity: A Comparison Between AM and Connectionist Models. *Journal of Quantitative Linguistics* 12. 262–296.

New, Boris, Marc Brysbaert, Jean Veronis & Christophe Pallier. 2007. The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28. 661–677.

Sokolik, M. E. & Michael E. Smith. 1992. Assignment of gender to french nouns in primary and secondary language: a connectionist model. *Second Language Research* 8(1). 39–58.

Tribout, Delphine. 2010. *Les conversions de nom à verbe et de verbe à nom en français*: Université Paris Diderot - Paris 7 dissertation.

Tribout, Delphine, Lucie Barque, Pauline Haas & Richard Huyghe. 2014. De la simplicité en morphologie. In *Actes de CMLF 2014,* 1879–1890.

Tucker, G Richard, Wallace E Lambert & André Rigault. 1977. *The French speaker's skill with grammatical gender: An example of rule-governed behavior.* Berlin: Mouton De Gruyter.

A quantitative comparison between word-formation & inflection: A look at paradigms in French

Gilles Boyé
Université Bordeaux-Montaigne & UMR5263

Gauvain Schalchli
Université Bordeaux-Montaigne & UMR5263

1   Introduction

The frontier between inflection and word-formation has been described in many ways from the unified approach of morpheme-based morphology (Hockett 1954; Lieber 1982; Halle 1973) to the total split of lexeme-based morphology (e.g. Beard 1995; Anderson 1982) and some intermediary positions against the total split (Booij 1994). In this paper, we propose to reevaluate the situation in a realistic word-based approach with a discussion about paradigmaticity in word-formation and inflection based on a quantitative comparison on French data.

2   Word-formation vs inflection in theory

The separation of inflection and word-formation made by lexeme-based theories is usually linked with their typical orientations:

- inflection expresses syntactic properties and its exponents sit at the periphery of words

- word-formation aims at producing lexical items and does not depend on syntactic contexts

In this type of frameworks, inflection describes the relation between words attached to the same lexeme and word-formation the relation between lexemes morphologically related. Booij (1994) focuses on a type of inflection that does not realise properties directly linked to contextual agreement (e.g. number in nouns, tense in verbs) and argues that this a sort of intermediary morphology between agreement inflection (contextual inflection) and word-formation (arbitrary coinage) where the speaker can pick the feature he wants but inside a set of choices limited by inflection (arbitrary inflection[1]).

Other arguments about split morphology revolve around three axes that are more characteristic of inflection:

(1) a. productivity: general availability of inflectional forms
     b. semantic regularity: relations between the content of inflectional forms are generally robust
     c. paradigmatic organisation: many to many relations are typical of inflectional systems

The preceding points in (1) have not received the same level of scrutiny when comparing inflection and word-formation. 1a has received a lot of attention from many authors (cf. Dal, 2003, for an overview). Recently Bonami & Paperno (2018) evaluated the relative semantic regularity of word-formation and inflectional semantic relations (1b). In this paper, we propose to compare the paradigmatic organisation of word-formation and inflection with a quantitative study on French.

---

[1] In Booij's terms: *inherent inflection.*

3 Word-formation vs inflection in French

In a realistic approach, sampling data is essential to evaluate the difference and the similarities between word-formation and inflection.

To study the quantitative difference between word-formation and inflection, we used Lexique3, a French lexicon compiled from corpora by New (2006). Our dataset of derivation chains was compiled by using Dérif (Namer, 2003) on the lemmas of Lexique3 and recouping bases for all derivation trees. In Lexique3, Dérif found 17564 derived words associated with 8435 bases and 986 derivations. For inflection, we collected all forms of verbs with at least two co-forms[2] in Lexique3. Our inflectional data set contains 16194 forms for 5835 verbs.

Following Bonami & Boyé (2014), we consider that paradigmatic generalizations originate with series of analogies between forms associated with the same pair of cells. As defended by Bonami & Beniamine (2016), we believe that joint-predictiveness is a central force in paradigm organization and that the number of co-forms for a given pair of cells is an indication of how reliable the generalizations made from this series will be.

Table 1 presents the number of co-forms for the 51 most common word-formation on the left and the number of co-forms for the 51 cells of the verbal paradigm on the right. The shades give an approximation of the number of co-forms (dark orange $\geq$ 2000, orange $\geq$ 300, light orange $\geq$ 30 and gray orange $<$ 30).[3]

Word-formations co-forms   Verb forms co-forms
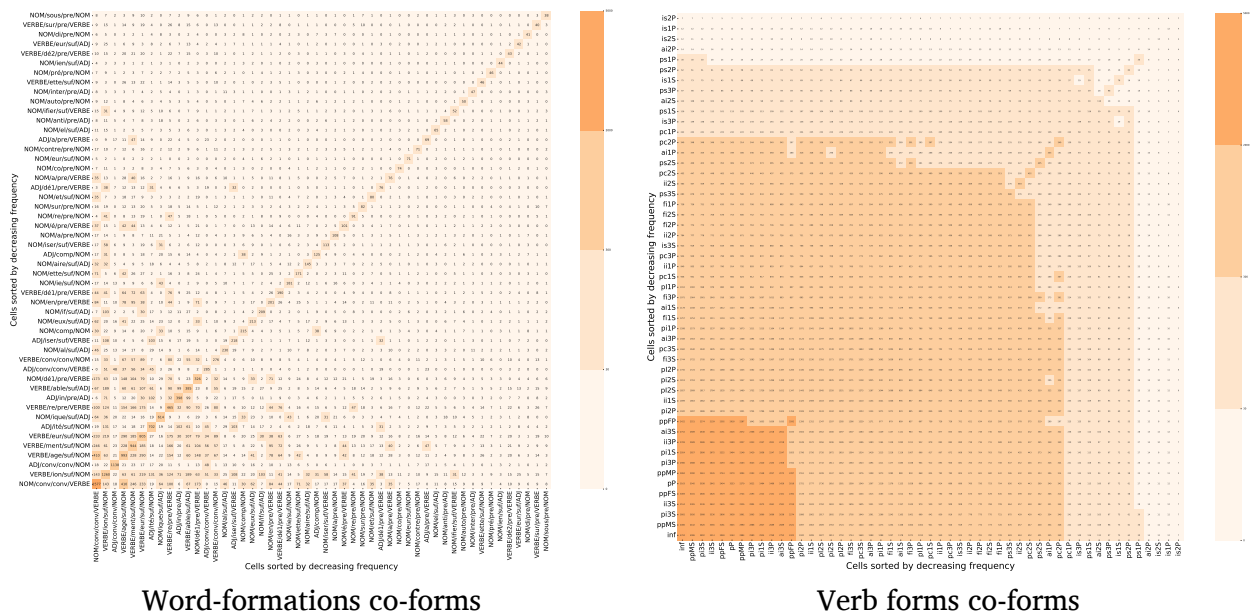
Figure 1: Co-occurence based on Lexique3

As expected, there are huge differences between the two maps in Figure 1:

- the inflection data has large zones of densely populated areas (dark orange, orange) while the word-formation is mostly empty (gray orange).

- inflection zones are concentric but the word-formation zones are patchy

- the dark orange shade does not appear in word-formation

---

[2] We call co-forms a set of forms occupying different paradigm cells for the same lexeme.

[3] In the word-formation figure, the diagonal represents the number derived words in the base series: base⇒derived word.

- the series effect is marked in word-formation by the darker diagonal counting the relation between a derived word and its base while there is no such difference in shade for inflection

However the observations above must not distract us from the similarities between parts of inflection and part of word-formation:

- the peripheral part of inflection is as empty as the majority of derivation (gray orange)

- the orange of word-formation corresponds to the outer of the orange part of inflection

4   Conclusion

From these observations, we propose that the comparison between word-formation and inflection should not be done at a global level but rather on specific sets of data. Some word-formation processes seem to form paradigms in a way similar to core inflection and some part of inflection behave on a par with typical word-formation.

In this paper, we compared word-formation and verbal inflection in French and found some intersection between the two. We think that word-formation paradigmatic phenomena could be closer to that of inflectional systems with much smaller paradigms, for example, in French, nouns and adjectives inflectional paradigms have respectively 2 and 4 cells, bringing them much closer to the size of "derivational paradigms". In this perspective, French deadjectival adverbs in *-ment* or pairs of masculine/feminine animate nouns would be particularly suited for a paradigmatic analysis in word-formation, preserving the intuitions of proponents of inflection and derivation in one analysis.

References

Anderson, Stephen R. 1982. Where's Morphology? *Linguistic Inquiry* 13(4). 571–612.

Beard, Robert. 1995. *Lexeme morpheme-based morphology*. State University of New York Press.

Bonami, Olivier & Sarah Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.

Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux* 17–45.

Bonami, Olivier & Denis Paperno. 2018. Inflection vs. derivation in a distributional vector space. *Lingue e linguaggio* 17(2). 173–196.

Booij, Geert. 1994. Against split morphology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1993*, 27–49. Kluwer Academic Publishers.

Dal, Georgette. 2003. Productivité morphologique: définitions et notions connexes. *Langue française* 3–23.

Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4(1). 3–16. http://www.jstor.org/stable/4177749.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–234.

Lieber, Rochelle. 1982. *On the organization of the lexicon*. Indiana University Linguistics Club.

Namer, Fiammetta. 2003. Automatiser l'analyse morpho-sémantique non affixale : le système dérif. *Cahiers de Grammaire* 28. 31–48.

New, Boris. 2006. Lexique 3: Une nouvelle base de données lexicales. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN)*, 892–900.

"Prestigious plurals" and conjugational class variation

Angelo Costanzo

Bloomsburg University of Pennsylvania

1 Preliminaries

While the past 20 years have seen a number of innovative ways of modelling the organization of conjugational classes (Boyé 2000, Boyé & Bonami 2003, Ackerman & Malouf 2013, Stump 2016, Sims & Parker 2016, etc.), the role that variation plays in conjugational class organization is far less understood. In order to better understand the role that linguistic variation can play in the organization and composition of conjugational classes, I examine the case of plural formation in English and Persian/Tajik. While these phenomena are simple on the surface (e.g., English plurals are generally (though certainly not always) formed by the suffixation of [-s/z/əz], etc.), the actual details (especially those that are the result of language contact) show a much more fragmented situation. The details suggest the invocation of a "constellational" (Joseph & Janda 1986) model, which can then be visualized under a "network" model (Langacker 1987, Bybee 1995).

2 Pluralization in English and Persian/Tajik

In both English and Persian/Tajik noun pluralization, there is an overriding rule that can be applied to (nearly) all nouns in the language. While the vast majority of English nouns can be pluralized with the suffix */-s/*, absolutely all Persian/Tajik nouns **can** be pluralized by the suffix *-hâ* (Perry 2005:63). While both languages have other native means of marking plurality (e.g., ablaut in English, the Persian/Tajik suffix *–ân* used mainly for animate nouns in the literary language, Lazard 2006:39), much of the complication of the plural marking system is the result of language contact. Both languages have a large foreign element in their lexicon; some foreign words maintain a plural form that follows the plural formation patterns of the (typically prestigious) source language. These "prestigious plurals" exist alongside variants with regularized plural morphology. For example many English speakers use *curricula* as a plural form of *curriculum* (< Latin), many speakers alternatively use *curriculums*, and many speakers use both, where the choice of one or the other depends of formality and related considerations. Similar cases are seen in Persian/Tajik, where many nouns of Arabic origin (e.g., *ketâb* 'book') have multiple plural forms (*kotob* < Arabic, *ketâb-hâ* < Persian 'books'), where the choice of one or the other is roughly equated with formality (Wei 1963:20, Perry 2015:65).

3 Theoretical concerns

The existence of "prestigious plurals" raises four key questions regarding the structure of the noun inflectional systems in question.

- What is the status of these cases in the system?
- What are the implications of these cases on organizational models?
- What are the implications of these cases on models of morphological change?
- How can traditional models of conjugational classes deal with variation?

When it comes to describing the morphological process(es) of pluralization in these languages, such instances tend to be set aside as "irregular", and are treated as somehow external to the system being accounted for. However, cases of morphological change suggest this is not the case, as there are indeed instances in which a foreign pluralization pattern has been extended to other nouns. For example, some English speakers pluralize *process* and *bias* as *process*[iz] and *bias*[iz], respectively, following (though not entirely) a more "foreign" pattern (cf. *thesis~theses, crisis~crises*). Again, there are similar cases in Persian/Tajik where native words take Arabic plural morphology, e.g., the plural form of *mive* 'fruit' is frequently *mivejât*, which exists alongside *mive-hâ*). There are even cases where native Persian nouns follow an Arabic non-concatenative (i.e., "broken") plural pattern (e.g., *ostâd > asâtid* 'professors'). This demonstrates that the Arabic plurals are integrated sufficiently enough into the Persian/Tajik system where these patterns have been analogically extended to native words.

4 Constellational/network models of conjugational classes

Taking into account their core similarities, the diversity in the patterns that English and Persian/Tajik nouns show, along with the variation that exists, suggests that the pluralization systems of the two languages be analysed under a constellational model, which has no problems in dealing with the existence of linguistic variation. Taking inspiration from Joseph & Janda (1986) and Langacker (1987), I pursue the fine-grained details and the observed patterns of variation that occur to better understand the range of pluralization patterns in English and Persian/Tajik and the principles by which they are organized. I consider data from a variety of sources, including the following corpora.

- English corpora:
    - Global Web-Based English (GloWbE) (1.9b words)
    - Corpus of Contemporary American English (COCA) (560m words)
- Persian/Tajik corpora:
    - Uppsala Persian Corpus (2.6m words) (Seraji 2015)
    - Talkbank Persian Corpus (474m words)
    - Tajik Web Corpus (93m words)

I visualize the results using a network model to better understand the full complexity of a phenomenon that grammars too readily gloss over, particularly the role of variation in the system. See Figure 1 below for a preliminary network diagram of 25 Persian nouns, with frequencies taken from the Talkbank Persian Corpus.

English and Persian/Tajik pluralization join a growing list of phenomena, e.g., Sanskrit reduplication, Romance conjugational classes, etc., that can be understood under a constellational/network model, which can have implications for what we know about the scope of linguistic generalizations, and consequently, conjugational class organization and structure as well.
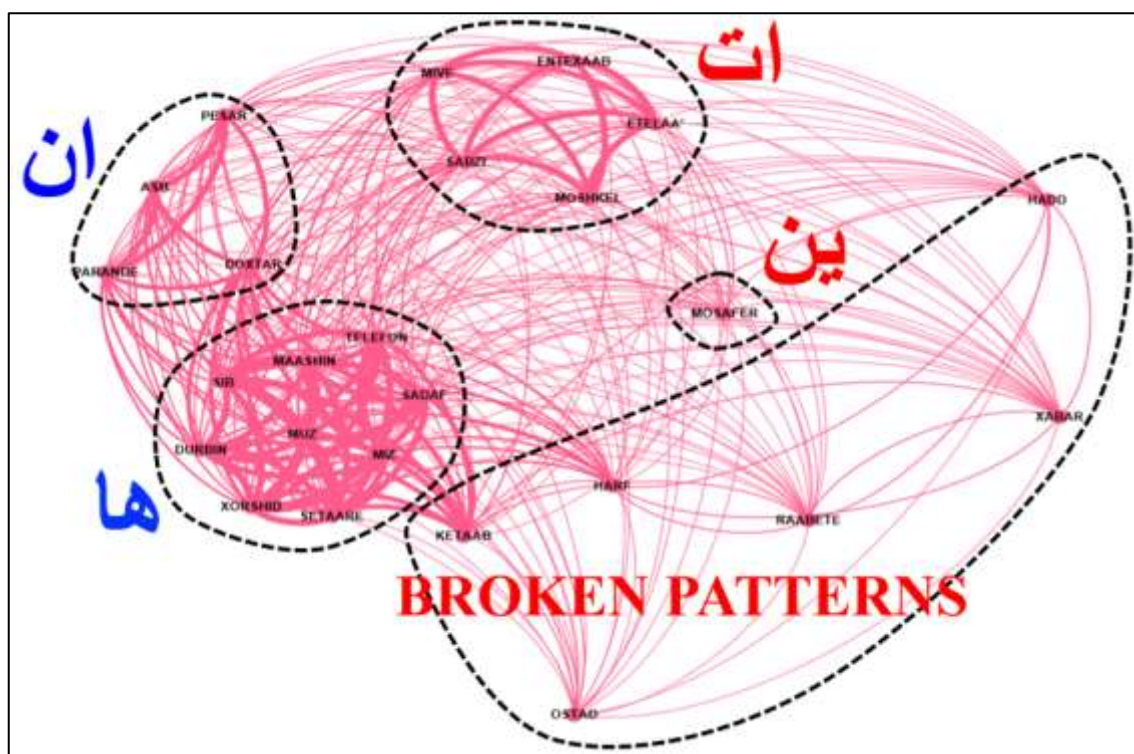
Figure 1: 25 Persian nouns, network model

References

Ackerman, F. & Malouf, R. (2013). Morphological organization : The low conditional entropy conjecture. *Language* 89(3) : 429-464.

Bonami, O. & Boyé , G. (2003). Supplétion et classes flexionnelles. *Langages* 152: 102-126.

Boyé, G. (2000). *Problèmes de morpho-phonologie verbale en français, en espagnol et en italien,* Paris: Université Paris VII dissertation.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and cognitive processes* 10(5):425-455.

Janda, R. D. & Joseph, B. D. (1986). One rule or many? Sanskrit reduplication as fragmented affixation. *Proceedings of the Second Eastern States Conference on Linguistics [ESCOL '85],* 103-119. Columbus: The Ohio State University Department of Linguistics.

Langacker, R. (1987). *Foundations of cognitive grammar: Vol. I: Theoretical prerequisites*. Palo Alto: Stanford University Press.

Lazard, G. (2006). *Grammaire du person contemporain*. Tehran: Institut français de recherché en Iran.

Perry: J. R. (2005). *A Tajik Persian reference grammar.* Leiden: Brill.

Seraji, M. (2015). *Morphosyntactic corpora and tools for Persian*. Uppsala: Acta Universitatis Upsaliensis.

Sims, A. D. & Parker, J. (2016). How inflection class systems work: On the informativity of implicative structure. *Word Structure* 9:2: 215-239.

Stump, G. (2001). *Default inheritance hierarchies and the evolution of inflectional classes.* In L.J. Brinton (Ed.), *Historical linguistics 1999,* (293-307). Amsterdam: John Benjamins.

Stump, G. (2016). *Inflectional paradigms: Content and form at the syntax-morphology interface.* Cambridge: Cambridge University Press.

Morphotactic dependencies in Yimas: a constructional approach

Berthold Crysmann

CNRS, Laboratoire de linguistique formelle

I discuss the system of pronominal affixation in Yimas (Foley, 1991)[1] and its interaction with the marking of modality (negation/potential). I show in particular that the choice of exponence displays a morphotactic dependency between discontinuous markers and suggest that this property is best captured in a constructional approach to inflectional morphology.

1 Pronominal affixes in Yimas

In Yimas, core participants are typically cross-referenced by prefixes on the verb. As shown in Table 1, animates distinguish person and number (SG,DU,PL).[2]

PER	NUM	A	O	S
	SG	ka-	ŋa-	ama-
1	DU	ŋkra-	ŋkra-	kapa -
	PL	kay-	kra-	ipa-
	SG	n-	nan-	ma-
2	DU	ŋkran-	ŋkul-	kapwa-
	PL	nan-	kul-	ipwa-
	SG	n-	na-	na-
3	DU	mpi-	impa-	impa-
	PL	mpu-	pu-	pu-

Table 1: Paradigm of Yimas pronominal affixes (Foley, 1991, 200)

With animates, prefixes come in different shapes for S, A, and O arguments, although syncretism between S and O can be observed for third person markers. A slightly complicating factor is contributed by the fact that there is an alternation between A and S forms, depending on whether realisation is word-initial or not. E.g. initial 2nd and 3rd person markers for the A-argument assume the same shape of the corresponding S marker, whereas non-initial 1st person S markers assume the shape of the corresponding A marker. Yimas has an elaborate noun class system: thus, for inanimates, prefixes distinguish number and gender.

Relative order of the markers is governed by the person hierarchy and, amongst third person markers, by thematic obliqueness, with higher ranked participants expressed closer to the stem. As illustrated in (1–2), with two 3rd person core participants, the O argument is realised peripherally, followed by the marker for the A argument.

(1) pu-   n-   tay
     3.PL.O 3.SG.A see
     'He saw them.'

(2) na-   mpu- tay
     3.SG.O 3.PL.A see
     'They saw him.'

Combination of a 3rd person participant (A or O) with a 1st or 2nd person participant (A or O) always has the 3rd person marker preceding the 1st or 2nd person marker, as shown in

---

[1]The data in this abstract are all taken from Foley's monograph.

[2]For reasons of space, I shall ignore the paucal marker, which is by-and-large orthogonal to the issues discussed here.

(3–4) for 1st person. Thus, depending on the inner marker, the outer marker's affiliation will change (note the neutralisation of A markers (to S) in initial position).

(3) pu- ka- tay   (4) pu- ŋa- tay
    3.PL.O 1.SG.A see   3.PL.A 1.SG.O see
    'I saw them.'   'They saw me.'

Inanimates, i.e. nouns in classes other than I and II appear to pattern alike, with 1st and 2nd person markers surfacing closer to the stem, irrespective of grammatical function. Still parallel, combinations of two 3rd person core arguments have the object marker precede the subject marker, as above.

Finally, I turn to combinations of 1st and 2nd person: if the hearer acts on the speaker, both participants are realised as discrete markers, with 1st person appearing closer to the stem, in accordance with the person hierarchy.

(5) ma- ŋa- tay
    2.SG.A 1.SG.O see
    'You (SG) see me.

If, however, the speaker acts on the hearer, we either find a portmanteau (1.A/2.SG.O), or else 1st person is realised by an independent pronoun.

1.1 Negative/potential

Exponents of negative/potential, which appear left-adjacent to the inner marker, block realisation of the outer prefix. What we find instead is realisation by means of suffixal number markers, distinct in shape from prefixal person or class markers. Most importantly, use of these suffixal markers is illicit outside the context of the negative and the potential.[3]

	SG	DU	PL
A/S	∅	-rm	-ump
O	-ak	-rm	-ump

Table 2: Suffixal number markers

For illustration, I present the positional template for Yimas, as distilled from Fowley's description:

	-3	-2	-1	0	1	2	3
(6)	{2.A,3}	{1,2,3}	IO	stem	TNS	PC	
	NEG/POT	{1,2,3}	IO	stem	TNS	PC	NUM/GEND

Which function the suffixal markers encode depends largely on how competition for the inner slot was resolved according to the person and thematic hierarchies: To start with, combination of 1st and 3rd person only witness a 1st person prefix (slot -2), with 3rd person being expressed by a number/gender suffix in slot 3, cf. (7–8).

(7) ta- ŋkra- tpul ∅   (8) ta- ŋkra- tpul -c -ak
    NEG 1.DU.O hit   PRF SG.A   NEG 1.DU.A hit   PRF SG.O
    'He didn't hit us two.'   'We two didn't hit him.'

---

[3]For reasons of space, I shall only document the negative. According to Foley (1991), however, the morphological distribution of the potential is fully identical.

Next, with two 3rd person participants, the A argument is realised prefixally, and the O argument suffixally.

(9) ta- mpu- tpul -c -rm
    NEG 3.PL.A hit PRF DU
    'They didn't hit those two.'

Combinations of 1st and 2nd person fall into two categories: with 1st person A arguments, there is only ever one exponent in the positive, and so the negative/potential do not differ from the general case in this respect. With 2nd person A and 1st person O arguments, however, the peripheral 2nd person marker is suppressed, giving rise to suffixal realisation. Since the suffixes do not distinguish person, we observe neutralisation between $3 > 1$ and $2 > 1$, cf. (10).

(10) ta- kra- tpul -c um
     NEG 1.PL.O hit PRF PL
     'You all/they didn't hit us.'

Finally, there are (at least) another two complicating factors in the negative/potential: with intransitives, number of the S argument is redundantly encoded for first person, as shown in (11).

(11) ta- kay- wa -r -um
     NEG 1.PL.S go PRF PL
     'We didn't go.'

Furthermore, with third person S arguments, number is equally expressed suffixally, yet the standard person/number prefix is *pu* in all numbers, a marker which otherwise encodes 3PL (cf. Table 1).

(12) ta- pu- wa -nan   (13) ta- pu- wa -na -rm
     NEG 3PL.S go NR PST   NEG 3PL.S go NR PST DU
     'He didn't go yesterday.'   'Those two didn't go yesterday.'

To summarise, Yimas confronts us with a four-fold dependency between exponents: preemption of outer prefixal person markers, dependence of suffixal number markers on prefixal modality markers, modality-based choice of inner person/number exponents, and assignment of suffix grammatical function based on prefixal competition.

1.2 Discussion

The Yimas data pose some clear challenges to most theories of inflection, however for quite different reasons: for morphemic theories the biggest obstacle is discontinuous extended exponence, as observed with number marking of S arguments in the negative or potential. For rule-based approaches, such as A-morphous Morphology (Anderson, 1992) or PFM (Stump, 2001), cascaded rule blocks do not provide for an easy expression of syntagmatic dependencies. The problem here is that combination of forms implies introduction by separate rule blocks, yet separation into rule blocks makes interaction difficult to state. Word-based approaches, such as Blevins (2005), finally, may not have problems with the syntagmatic dependence between exponents (cf. Harris, 2009), but they clearly have difficulty capturing reuse of exponents across different positions or functions.

Instead, I shall propose an analysis in terms of discontinuous morphotactic constructions that directly take into account the dependent nature of the suffixal markers, as well as competition for initial position, and will derive the affiliation with subject or object function on the basis of competition for prefixal realisation.

2   An analysis in terms of morphotactic constructions

The formal analysis I propose is couched within the framework of Information-based Morphology (Crysmann & Bonami, 2016), a constructional theory of inflectional morphology that organises rules associating m functions with n morphs into cross-cutting inheritance hierarchies. Most importantly for our purposes here, exponents introduced by any rule may be discontinuous, owing to the fact that position class information is a first class property of exponence, alongside shape.

For Yimas, I propose three dimensions: one for the prefixal exponents of core grammatical functions, one for the prefixal markers of negative and potential, and a third one for the suffixal number markers. What is crucial is that the partial rule descriptions in these three dimension are underspecified with respect to the maximal number of morphs, in order to permit combination of descriptions by means of unification.

Essentially, the rule descriptions for prefixal pronominal affixes only assert the shape and the prefixal position of the exponent, as a member of the list of morphs, without restricting how many morphs there will be in total. The description for suffixal markers defines the shape and position of the number suffixes, but furthermore requires the presence of two prefixal markers, in positions -3 and -2. Likewise, the constraints introducing the negative and potential in slot -3 will require a morph in slot -2, yet be open to combination with a suffixal marker in slot 3. Intersection of descriptions from the three domains will derive the patterns in the negative/potential. In order to account for the use of pronominal affixes outside the negative/potential construction, it is sufficient to provide a monomorphic template in the dimension for negative/potential which will unify with the pronominal prefixes. Finally, idiosyncratic markers, such as the number-neutral version of *pu-* in the negative can easily be accounted for by making them morphotactically dependent. Moreover, by integrating suffixal and prefixal marking into a morphotactic construction, it becomes straighforward to account for multiple exponence in the case of S arguments. To conclude, cross-classification of underspecified morphotactic rule descriptions can derive rules for pronominal affixation for both constructional and general variants alike that not only capture the independent and dependent uses, but at the same time can account for constructionally induced idiosyncrasies.

References

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Blevins, James P. 2005. Word-based declensions in Estonian. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of morphology 2005*, 1–25. Springer.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in information-based morphology. *Journal of Linguistics* 52(2). 311–374.

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

Harris, Alice C. 2009. Exuberant exponence in Batsbi. *Natural Language and Linguistic Theory* 27. 267–303.

Stump, Gregory T. 2001. *Inflectional morphology. a theory of paradigm structure*. Cambridge: Cambridge University Press.

Formal and Semantic Transparency
in L1 and L2 Morphological Processing

Serena Dal Maso
University of Verona

Sabrina Piccinin
University of Verona

1 Background

While there is a general agreement on the organization of L1 mental lexicon according to morphological parameters (Diependaele et al. 2012 for a review), the debate on the role of morphological factors on L2 lexical representation and processing is still open and lively, as data are still scarce (especially for less commonly learned L2s) and interpretations are controversial. Limiting our attention to derived words (rather than inflected ones), results coming from the masked priming experimental paradigm seem to indicate that proficient L2 speakers adopt processing mechanisms which are (qualitatively) similar to those used by L1 speakers, although L2 processing still remains cognitively more demanding than L1 processing. A growing body of literature indicates that when the prime and the target are morphologically related and when this relationship is both formally and semantically transparent (e.g., *hunter/hunt*), L2 speakers recognize target words significantly faster compared to the condition where the prime is unrelated (e.g., *flower/hunt*). Although different interpretations have been proposed, morphological facilitations have emerged quite consistently across the different L2s and the derivational patterns investigated (Giraudo & Dal Maso 2018 for a review).

The results of L2 studies indicate that proficient learners can in fact efficiently rely on words' structure during processing. However, when it comes to formally or semantically opaque complex words, L1 and L2 processing mechanisms seem to diverge. Specifically, orthographic/formal variation, or actual formal disruption affects more deeply L2 than L1 processing. This emerges when allomorphy is considered both in verbal inflectional systems (Basnight-Brown et al. 2007; Feldman et al. 2010 on L2 English; Jacob et al. 2013 on L2 German) and in the case of derivationally related words (Piccinin 2018 on Italian derived words, e.g., *fiore/floreale*). In fact, starting from the acknowledgment of this major reliance on formal aspects, some scholars went so far as to propose that L2 processing relies uniquely on formal features, rather than on truly morphological ones (Heyer & Clahsen 2015).

The effect of semantic transparency in L2 processing is however far less clear as only one experimental study dealt with this issue (Diependaele et al. 2011). Moreover, the picture is complicated by the fact that there is still very much controversy about this issue even in the L1 processing literature. Basically, there is a lack of general consensus on the actual emergence of morphological effects at the very early stages of word processing and recognition. As a matter of fact, Longtin et a. 2003 and Rastle et al. 2004 (among others) found no significant statistical difference between the facilitation yielded in transparent prime-target pairs (*cleaner/clean*) with respect to what they call opaque prime-target pairs (*department/depart* or *corner/corn*), which were both significantly faster than orthographic controls (*brothel/broth*). This kind of result has been taken as evidence of a 'semantically blind' initial processing stage which is characterized by a morpho-orthographic affix-stripping procedure. In this view, all words with apparent morphological structure would be initially parsed into stem and affix, independent of their actual morphological structure (simple words like corner or genuinely complex words like cleaner). As Baayen (2014) has observed, however, these results raise a

series of concerns with respect to the materials used in the opaque condition, which comprised both pseudo-suffixed (i.e., simple words whose superficial form might be 'decomposed' into possible morphological components, but are in fact monomorphemic, like *corner*) and highly lexicalized words which however display a fully functional suffix (such as *fruitless* or *archer*). Moreover, the semantics of some of the stimuli used such as *fruitful* 'successful' is made less opaque by the fact that *fruit* has also a related figurative meaning as in 'the fruits of one's labors' besides its literal meaning and cannot therefore be considered opaque to the same extent that *corner* is (Baayen 2014). The emergence of a priming effect in the opaque set, therefore, might have been due to the inclusion of this kind of prime-target pairs, which cannot be equated to pseudo-suffixed words. As a matter of fact, when pseudo-suffixed words and opaque derived words are kept apart, they seem to induce different priming patterns. This is the case in Feldman et al. 2009 who found out that morphological facilitation was significantly greater with semantically transparent morphologically related pairs (*coolant/cool*) than with opaque ones (*rampant/ramp*). Although such findings are not always replicated (for instance Beyersmann et al. 2015.), a growing body of research seems to indicate (Feldman et al. 2015, Basnight-Brown et al. 2007) that semantic properties do affect the early stages of complex words processing.

2 Our study

Given these premises, our research aims at verifying whether semantic transparency (vs. opacity) affects L2 processing of complex words or whether, given the major role played by formal aspects on L2 processing, semantics' influence is reduced with respect to native speakers.

Looking at L2 speakers' performance seems particularly interesting since the impact of morphological parameters on lexical organization and processing (with respect to purely formal ones) has been found out to be significantly related to vocabulary size even in the case of native speakers. Interestingly, Andrews & Lo 2013, investigating 92 university students, native speakers of English, found that morphological facilitation was significantly modulated by individual differences. Specifically, speakers with higher vocabulary knowledge obtained robust priming for transparent pairs (*worker/work*) but, crucially little priming for opaque or form pairs (*corn/corner* and *turnip/turn*). In contrast, individuals with lower vocabulary skills showed sustained priming for opaque pairs that was at least as strong as for transparently related pairs. In the same vein, Quémart et al. 2018 investigated the strength of morphological representations in French young readers (third, fifth, and seventh graders) and adults by using a lexical decision task associated with the masked priming paradigm in which targets were preceded by morphological (e.g., *tablette/table*, 'little table/table'), pseudoderived (e.g., *baguette/bague*, 'little stick/ring'), orthographic control (e.g., *abricot/abri*, 'apricot/shelter'), and semantic control (e.g., *tulipe/fleur*, 'tulip/flower') primes. Different patterns of priming were observed across the groups: in developing readers, the processing of written morphology was triggered by the formal properties, while semantic properties were activated later in the time course of word recognition. In adults, patterns of priming were similar, except that the activation of the formal properties decreased earlier in the time course of word recognition. Therefore, French developing readers seem to process both formal and semantic properties of words when reading but show a progressive quantitative change in the development of morphological processing over the course of reading development.

In the light of these results, our study will allow to verify whether and to what extent the reliance on morphological properties during L2 processing is modulated by semantics or whether their reduced and less automatized lexical competence results in a priority for formal

aspects (as observed in the case of developing L1 readers). If the first case is verified, we expect opaque primes to induce weaker morphological facilitation with respect to transparent ones; in the second case, in contrast, we should observe the same amount of facilitation for opaque and transparent prime-target pairs.

2.1 Materials and Procedure

A masked priming experiment associated with a lexical decision task was performed with two groups of subjects: L1 adult native speakers of Italian and L2 upper intermediate and advanced learners of Italian L2. Three priming conditions were included: identity, morphological, and unrelated. 30 Italian transparent and 30 opaque evaluative derived words were selected as primes to be used in the morphological condition. For the transparent set, we selected words derived by means of evaluative suffixes having an augmentative, diminutive, or pejorative semantics. For the opaque set, we selected highly lexicalized derived words (matching the same range of suffixes used in the opaque condition) which display semantic shift or semantic bleaching (e.g. *fumo/fumetto,* 'smoke/comics, comic strip'). Materials were matched for frequency and for length.

	Transparent set		Opaque set	
Conditions	**prime**	**target**	**prime**	**target**
Identity	Borsa 'bag'	Borsa 'bag'	Fumo 'smoke'	Fumo 'smoke'
Morphological	Borsetta 'handbag'	Borsa 'bag'	Fumetto 'comics'	Fumo 'smoke'
Unrelated	Suocera 'mother in law'	Borsa 'bag'	Verdura 'vegetables'	Fumo 'smoke'

Table 1: Experimental design

3 Preliminary Results

Preliminary results for the L1 group reveal a robust priming effect in the morphological condition for the transparent set: RTs yielded after the presentation of a morphologically related prime (*borsetta/borsa*) are significantly faster than those obtained in the Unrelated condition (*suocera/borsa*) and statistically equivalent to those obtained in the Identity condition (*borsa/borsa*). On the contrary, for the opaque set, the morphological facilitation was still significant with respect to the Unrelated condition, but interestingly, in this case, RTs in the morphological condition were significantly different also from those observed in the Identity condition. Therefore, while in the transparent set we observed, as expected according to the literature, a full priming effect, in the opaque set, only a partial priming effect emerged.

As for the L2 group, preliminary results indicate that a similar morphological effect emerges in the two sets, independently of the degree of semantic transparency of the prime-target relationship. If these results were confirmed, they would on the one hand confirm the efficacy of morphological parameters in L2 processing when transparent prime-target pairs are used. However, these findings would, on the other hand, indicate that L2 processing mechanisms, similarly to what observed with developing readers, are firstly driven by formal properties and that major reliance on formal properties reduces the effect of semantic opacity.

References

Andrews, Sally, & Steson Lo. S. 2013. Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language* 68(3). 279–296.

Baayen, Harald R. 2014. Experimental and psycholinguistic approaches to studying derivation. Handbook of Derivational Morphology, 95–117.

Basnight-Brown, Dana M., Lang Chen, Shu Hua, Aleksandar Kostić & Larie Beth Feldman. 2007. Monolingual and bilingual recognition of regular and irregular English verbs: Sensitivity to form similarity varies with first language experience. *Journal of Memory and Language* 57(1). 65–80.

Beyersmann, Elisabeth, Grainger Jonathan, Séverine Casalis, Johannes C. Ziegler. 2015. Effects of reading proficiency on embedded stem priming in primary school children. Journal of Experimental Child Psychology 139. 115-126.

Diependaele, Kevin, Jonathan Grainger & Dominiek Sandra. 2012. Derivational morphology and skilled reading: An empirical overview. In M. Spivey, K. McRae & M. Joanisse (Eds.), The Cambridge Handbook of Psycholinguistics (pp. 311–332). Cambridge: Cambridge University Press.

Diependaele, Kevin, Jon Adoni Duñabeitia, Joanna Morris & Emmanuel Keuleers. 2011. Fast morphological effects in first and second language word recognition. *Journal of Memory and Language* 64(4). 344–358.

Feldman, Laurie Beth, Patrick A. O'Connor & Fermín Moscoso del Prado Martín. 2009. Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review* 16(4). 684–691.

Feldman, Laurie Beth, Petar Milin, Kit W. Cho, Fermín Moscoso del Prado Martín & Patrick O'Connor. 2015. Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience* 9, 1–19.

Feldman, Laurie Beth, Aleksandar Kostić, Dana M. Basnight-Brown, Dusica F. Durđević & Matthew J. Pastizzo. 2010. Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. Bilingualism: Language and Cognition, 13(02), 119–135.

Giraudo, Hélène & Serena Dal Maso. 2018. Towards a constructional approach of L2 processing. In Geert E. Booij (ed.), *The Construction of Words. Advances in Construction Morphology*, 603-622. Dordrecht: Springer.

Heyer, Vera & Harald Clahsen. 2015. Late bilinguals see a scan in scanner AND in scandal: dissecting formal overlap from morphological priming in the processing of derived words. *Bilingualism: Language and Cognition* 18(03). 543–550.

Jacob, Gunnar, Elisabeth Fleischhauer & Harald Clahsen. 2013. Allomorphy and affixation in morphological processing: a cross-modal priming study with late bilinguals. *Bilingualism: Language and Cognition* 16(4). 924–933.

Longtin, Catherine-Mary, Juan Segui & Pierre A. Hallé. 2003. Morphological priming without morphological relationship. *Language and Cognitive Processes* 18(3). 313–334.

Piccinin, Sabrina. 2018. *Native and non-native processing of morphologically complex words in Italian*. PhD dissertation, University of Verona, University of Toulouse.

Quémart, Pauline, Gonnerman Laura, Jennifer Dowing & Hélène S. Deacon. 2018. The development of morphological representations in young readers: a cross-modal priming study. *Developmental science* 18. 1-10.

Rastle, Kathleen, Matthew H. Davis & Boris New. 2004. The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review* 11(6). 1090–1098.

A Paradigm Function Morphology approach to Moksha objective conjugation

Louise Esher

CNRS – CLLE-ERSS (UMR 5263)

Jean Léo Léonard

Université Montpellier 3 & Dipralang (EA 739)

1 Introduction

This study seeks to apply the Paradigm Function Morphology (PFM) framework developed by Stump (e.g. 2001, 2016) to the analysis of verb inflection in Moksha (Mordvin, Uralic). The study focuses on the phenomenon traditionally termed OBJECTIVE CONJUGATION, in which the inflection of transitive verbs is sensitive to person and number values of the definite object as well as the subject (intransitive verbs, and transitive verbs with indefinite objects, present a separate series of inflectional exponents, the SUBJECTIVE CONJUGATION, sensitive only to the person and number values of the subject). An example from the Šokša variety (Djordjevic Léonard & Léonard 2006:277) is given below, illustrating the contrast between objective conjugation (1a) and subjective conjugation (1b).

(1) a. *ćoreng-iś*　　　*ker-i-źe*　　　　　　　*pŕe-ď*
　　　boy.NOM-SG.DEF　　cut_off-PRT-SBJ3SG.OBJ3SG　head-ACC.SG.DEF
　　　'The boy cut off the head.'

　　b. *kere-ś*　　　　*kevejke*　*bandit-eń*　*pŕe*
　　　cut_off-PRT-SBJ.3SG　eleven　　bandit-GEN.PL　head-ACC.SG
　　　'He cut off eleven bandits' heads.'

As established by synchronic and diachronic studies (e.g. Keresztes 1999, Fournet 2004, Léonard 2008, Samvelian 2008), the relationship between morphosyntactic values and inflectional exponents in Mordvin objective conjugation is no longer an entirely transparent one. In contemporary varieties, there is extensive and highly systematic syncretism between inflectional exponents: some patterns are functionally arbitrary, others less so. A further point of interest is that, across Mordvin speech varieties, patterns of distribution are generally stable, while variation in the form of exponents is concentrated in certain areas of the paradigm.

This study proposes an analysis of objective inflection in Moksha in accordance with Stump's theory of PARADIGM LINKAGE, illustrating the suitability of this approach for developing an empirically plausible account of the study data.

2 Overview of objective conjugation in Moksha

Moksha presents a rich system of verb inflection, exhaustive analysis of which is beyond the scope of the present study. Following Kereztes (1999), the study focuses on objective conjugation for a subset of paradigm forms which are traditionally considered basic – the present (here glossed NPST), preterite (PST) and imperative (IMPV). Exponents for these categories in standard Moksha are shown in Table 1; note that the suffixes do not vary across inflectional classes, and that stem forms do not vary with reference to the person or number values of subject or object arguments.

The data in Table 1 illustrate several key structural principles of the system of objective conjugation (subjective forms are provided for comparison). No synthetic forms are available for the combinations {SUBJ.1, OBJ.1} or {SUBJ.2, OBJ.2}; these meanings are instead expressed analytically with a reflexive pronoun *eś* 'self'. There is widespread neutralisation of number contrasts (object number is neutralised where the subject is plural, subject number is

neutralised where the object is 1PL or 2PL). Exponents primarily associated with object agreement values occur nearer to the stem than exponents primarily associated with subject agreement values (and exponents primarily associated with subject agreement values frequently differ in form from the corresponding exponents in the subjective series). First-person objects are consistently associated with an exponent *-ma-*, while second-person objects typically present exponents *-tä-, -dä-, -t'äd'ä-* ; in the non-past, a default exponent *-sa-* is found for non-second-person objects [note that *-si-* in SUBJ.3SG forms is a reflex of older *-saj-*].

| mood | tense | SBJ | *Sbjve* | *Objve*, where DEF.OBJ person/number values are : | | | | | |
				1SG	2SG	3SG	1PL	2PL	3PL
IND	NPST	1SG	-(a)n	—	-t'ä	-sa	—	-t'äd'äź	-sajńə
		2SG	-(a)t	-samak	—	-sak	-samaśt'	—	-sajt'
		3SG	-j, -i	-samań	-tanza	-si	-samaź	-t'äd'äź	-sińə
		1PL	-tamə	—	-t'äd'äź	-saśk	—	-t'äd'äź	-saśk
		2PL	-tadə	-samaśt'	—	-saśt'	-samaśt'	—	-saśt'
		3PL	-jt'	-samaź	-t'äd'äź	-saź	-samaź	-t'äd'äź	-saź
	PST	1SG	-ń	—	-jt'äń	-jńə	—	-d'äź	-jńə
		2SG	-t'	-majt'	—	-jt'	-maśt'	—	-jt'
		3SG	-ś	-mań	-nźə	-źə	-maź	-d'äź	-źəń
		1PL	-mə	—	-d'äź	-śk	—	-d'äź	-śk
		2PL	-d'ə	-maśt'	—	-śt'	-maśt'	—	-śt'
		3PL	-śt'	-maź	-d'äź	-ź	-maź	-d'äź	-ź
IMP		2SG	-k	-mak	—	-k	-maśt'	—	-jt', -śt'
		2PL	-də	-maśt'	—	-śt'	-maśt'	—	-śt'

Table 1. 'Subjective' (Sbjve) and 'objective' (Objve) suffixes (data from Keresztes 1999:67–68).

3   PFM analysis of the standard Moksha system

The PFM framework and the theory of paradigm linkage are founded on two key principles: the notion that 'paradigms are the interfaces of inflectional morphology with syntax and semantics' (Stump 2016:23) and that 'some morphological regularities are, irreducibly, regularities in paradigm structure' (Stump 2016:26).

Three types of inflectional paradigm are assumed to exist: a CONTENT PARADIGM which provides information on available combinations of morphosyntactic/semantic feature values, a FORM PARADIGM which provides information on distributional patterns of exponence, and a REALISED PARADIGM which provides information on the form of exponents. The three paradigms are linked by functions which describe the formal relationships between cells and their exponents. Because this architecture explicitly dissociates syntactic and morphological processes (cf. Zwicky 1992) and provides for autonomously morphological structure (cf. Aronoff 1994), it is well equipped to describe systemic regularities in the distribution of inflectional exponents, whether these regularities map onto morphosyntactic/semantic features or not.

Thus, in the Moksha case illustrated here, one can assume that within the content paradigm there are individual cells corresponding to the sets of feature values {IND NPST SUBJ.1PL OBJ.2SG}, {IND NPST SUBJ.3PL OBJ.2SG}, {IND NPST SUBJ.1SG OBJ.2PL}, {IND NPST SUBJ.3SG OBJ.2PL}, {IND NPST SUBJ.1PL OBJ.2PL} and {IND NPST SUBJ.3PL OBJ.2PL}, because each of these combinations is required by a distinct syntactic context; but that all six of these content paradigm cells map to the same form paradigm cell, because for any given lexeme their

corresponding wordforms display systematic syncretism (e.g. for *kundams* 'take', all six sets of feature values correspond to the single realised form *kundat'äd'äź*)

The distribution and phonological form of inflectional exponents are described using ordered blocks of REALISATIONAL RULES; the rules within each block are in competition and apply from most to least specific (i.e. default). For the data in Table 1, two blocks are required: the first introduces a formative which may be broadly characterised as an object marker, and the second an exponent which may be broadly characterised as a subject marker. Note that the series of exponents found in the present is also shared with the conditional, while the series of exponents found in the preterite is common to the subjunctive, desiderative and conditional-subjunctive (Bartens 1999:139), the optative has a unique series, and the imperfect is only found with subjective inflection (Bartens 1999:131). For this reason, it is desirable to consider the two realisational rule blocks as describing exponents of MORPHOMIC SUFFIX SERIES, rather than exponents of particular TAM categories (cf. Stump 2016:120-126).

series 1		series 2	
SUBJ.2SG OBJ.1 -sama-	SUBJ.2PL OBJ.1 -sama-	SUBJ.2SG OBJ.1 -ma-	SUBJ.2PL OBJ.1 -ma-
SUBJ.3SG OBJ.2SG -ta-	SUBJ.3SG OBJ.2 -tädä	SUBJ.3SG OBJ.2SG —	SUBJ.3SG OBJ.2 -dä-
SUBJ.3SG OBJ.3SG -si-	SUBJ.3SG OBJ.3PL -si-	SUBJ.3SG OBJ.3SG -zä-	SUBJ.3SG OBJ.3PL -si-

Table 2. Exponents of object person and number features for the two morphomic series.

The exponents described by the first rule block are shown in Table 2, highlighting contrasts between the more substantial, robustly object-iconic exponents found in series 1, and the more diverse, idiosyncratically distributed exponents found in series 2. Also of note is the particularly wide variety of exponents associated with second-person objects (uniquely distinguishing the 'contrastive' speech act participant, that which renders the speech act possible).

4   Adapting to dialectal variation

series 1	northern	central	south-western	south-eastern	southern
SUBJ.1SG, OBJ.3SG	*-sa*				
SUBJ.1SG, OBJ.3PL	*-sań* *-sajńë*	*-sajńë* *-sajnä* *-sajëń*	*-sajn* *-sajńë*	*-sajn*	*-sajä* *-sajńë*
SUBJ.2SG, OBJ.1SG	*-samak*				
SUBJ.1PL, OBJ.2SG SUBJ.3PL, OBJ.2SG SUBJ.1SG, OBJ.2PL SUBJ.3SG, OBJ.2PL SUBJ.1PL, OBJ.2PL SUBJ.3PL, OBJ.2PL	*-tädäź*		*-tedeź* *-t'ëd'ëź*	*-tädäź* *-tedeź*	*-tädäź*

Table 3. Comparison of an illustrative selection of inflectional exponents across varieties of Moksha (data from Keresztes 1999:206, 208, 238, 240).

In a framework with three distinct levels of paradigmatic structure, diatopic variation may involve any of three dimensions: from one speech variety to another, the content paradigm

may contain a different array of cells, the form paradigm may consist of different groupings of cells, and the realised paradigm may involve different exponents. Consideration of the data presented by Keresztes (1999) for Moksha varieties shows that, overwhelmingly, diatopic variation only affects the realised paradigm, while the shape of (and mapping between) the form and content paradigms remains constant; furthermore, variation within the realised paradigm is concentrated in certain cells. Illustrative examples of 'present' or 'series 1' suffixes are shown in Table 3 for five varieties of Moksha: some exponents are uniform across the dialect area, while others show much greater variability; notably, the exponent realising combinations of a second-person object and a plural object or subject varies in form, but its paradigmatic distribution remains robustly stable across the dialect area.

5 Conclusions and perspectives

This study proposes a formal PFM analysis of a fragment of Moksha verb inflection, namely the inflectional exponents which occur for transitive verbs with a definite object. The analysis, which provides a starting point for a fuller description of Moksha verb inflection, is comparable to the PFM analysis of equivalent data for Erzya (Mordvin) conducted by Samvelian (2008), but integrates the more recent notion of paradigm linkage, and demonstrates the advantage of this notion in providing an empirically and theoretically satisfying account of the data. Within a paradigm linkage analysis, the syntactic context (presence/absence, definiteness/indefiniteness of the object) does not directly condition inflectional realisations, but instead determines which content cell is to be selected; the mapping between this cell and inflectional forms is mediated by the form paradigm, which groups together cells which share inflectional realisations. The separation of content and form paradigms preserves the principles of morphology-free syntax and syntax-free morphology, while the mapping between the two explicitly builds recurrent patterns of relationship between cells into the structure of the inflectional paradigm. PFM is shown to be an effective tool for capturing the organisational principles of Moksha objective conjugation, and thus revealing the underlying simplicity of this apparently complex inflectional system.

References

Aronoff, Mark. 1994. *Morphology by itself*. Cambridge, MA & London: MIT Press.

Bartens, Raija. 1999. *Mordvalaisten kielten rakenne ja kehitys*. Helsinki: Suomalais-Ugrilainen Seura.

Djordjevic Léonard, Ksenija & Jean Léo Leonard. 2006. *Parlons mordve.* Paris: L'Harmattan.

Fournet, Arnaud. 2014. Anaphore et conjugaison objective en moksha. Conclusion et perspectives. *Études finno-ougriennes* 46. http://journals.openedition.org/efo/4175.

Keresztes, László. 1999. *Development of Mordvin Definite Conjugation*. Helsinki: Suomalais-Ugrilainen Seura.

Léonard, Jean Léo. 2008. Simplicité de la flexion mordve ? *Bulletin de la Société de linguistique de Paris* 103(1). 363–400.

Samvelian, Pollet. 2008. La conjugaison objective en erzya (mordve). Une approche réalisationnelle. *Bulletin de la Société de linguistique de Paris* 103(1). 333–361.

Stump, Gregory T. 2001. *Inflectional Morphology*. Cambridge: CUP.

Stump, Gregory T. 2016. *Inflectional Paradigms*. Cambridge: CUP.

Zwicky, Arnold. 1992. Some choices in the theory of morphology. In Robert Levine (ed.), *Formal grammar: Theory and implementations*, 327–371. Vancouver: University of British Columbia.

The lexicon beyond lexemes

Bernard Fradin

Université de Paris, Laboratoire de Linguistique Formelle, CNRS

1 Introduction

This submission aims to clarify the relationship between the lexicon and the concept of lexical entry on the one hand, and the latter and lexeme-formation morphology (i.e. derivation or compounding) on the other. It is argued that (i) the notion of lexical item/entry must be distinguished from that of lexeme; (ii) lexical entries are not entities relevant for morphology; (iii) the fact that lexical entries include both lexemes and polylexematic units does not undermine the argument that derivational morphology is lexeme-based.

2 Statement 1

Two units which share the same inflectional paradigm but denote distinct meanings must be be considered distinct lexemes. An example is provided by the French nouns FILLE$_1$ 'girl' and FILLE$_2$ 'daughter'. The fact that their meaning is different can be ascertained on the basis of the contrasts shown in examples (1)-(3). Bonami & Crysmann (2018) capture the differences in question, in an HPSG framework, using features LID and PID. The feature LID (LEXEME IDENTIFIER) identifies the lexical unit that is the head at the phrase level, and its value corresponds to the main predicate associated with the lexeme (Sag, 2012). The feature PID (PARADIGM IDENTIFIER), on the other hand, specifies the inflectional paradigm that the lexeme in question instantiates. In the present case, both FILLE$_1$ and FILLE$_2$ follow the same pattern of inflection (same PID) but carry two distinct meanings and, thereby, constitute two distinct lexemes.

(1) *garçons et (filles$_1$ | *filles$_2$)*
    'boys and (girls | *daughters)'

(2) *vêtements de (filles$_1$ | *filles$_2$)* 'girls' clothing'
    vêtements de (garçons | *fils) 'boys' clothing'

(3) *telle mère, telle (filles$_2$ | *filles$_1$)* 'like mother, like daughter'
    tel père, tel (fils | *garçon) 'like father, like son'

As is well-known (Goddard, 2000, 133), only FILLE$_1$ can be correlated with the diminutive FILLETTE 'little girl' (cf. Bonami & Crysmann (2018, 186) for an account). This shows that derivational morphology is selective: derivational processes apply to certain lexemes and not to others in function of their semantics. This conclusion is long-etablished (Kerleroux, 2004) and is even more pronounced with verbs, insofar as verbs frequently head several distinct constructions. This is illustrated with FONDRE, where constructions (4)-(7) are given with the derived lexemes allowed for each verb. Except for FONTE$_1$, the ranges of derived lexemes clearly do not overlap.

(4) X[PAT] FONDRE$_1$ *fonte$_1$* 'melting'
    Dehors, la glace fond. 'Outside, ice is melting'

(5) X[ACT] FONDRE$_2$ Y[PAT]   *fusion$_2$* 'fusion'
    On fond le mélange à une température élevée.
    'The mixture is melted at high temperature'

(6) X[AGT] FONDRE$_3$ Y[PAT] (en Z)   *fonte$_1$* , *fonte$_3$* 'cast iron', *fondeur$_3$* 'foundryman', *fonderie$_3$* 'foundry, smelter'
Ils fondaient le bronze en lingots. 'They smelted bronze into ingots'

(7) X[PAT, FIG] FONDRE$_4$ dans Y[GRND]   *fonte$_4$* 'dissolving'
Le sucre fond dans l'eau. 'Sugar dissolves in water'

3 Statement 2

If we compare the various items FONDRE mentioned in §2 with true homonyms such as the English *bank$_1$* 'the land alongside a river' and *bank$_2$* 'financial establishment (...)' , we can hypothetize that they are somehow correlated. How can we account for this correlation? A classical answer is to consider them as instances of 'lexical readings' subsumed by a unique lexical entry (Katz, 1972, 70), a view which reflects the way dictionaries deal with this issue. This view is still endorsed in recent works, which consider the various constructions listed under (4)-(7) as "lexical variants" of a polysemous lexical item (Gamerschlag et al., 2014). But if we contend that the domain of application of derivation is the lexeme (Aronoff, 1994), then it is more coherent to say that the various instances of FILLE or FONDRE constitute distinct lexemes.

As for the correlations themselves, the approaches that exist do not address exactly the same issues. Among them, we find Lakoff's Idealized Conceptual Models (Lakoff, 1987)(Jurafsky, 1996); the approaches dealing with lexical alternations in different frameworks (Rappaport Hovav & Levin (2010); Davis (2001); Ackerman & Moore (2001)); and frame semantics (Gamerschlag et al., 2014), (Plag et al., 2018). Describing the correlations observed between lexemes that form configurations such as those illustrated by FILLE and FONDRE is an issue which concerns lexical theory but not morphology, and all the more so as lexical entries are not a relevant object to lexeme formation morphology. Only lexemes are. However, in lexical frameworks where lexeme formation rules are a direct extension of the lexicon's hierarchic organization (Koenig, 1999), the way the relations between lexemes are established becomes crucial and has a bearing upon the capacities of derivational morphology. Moreover, the inheritance relations based on the hierarchic organization of the lexicon involves lexemes, not lexical entries.

4 Statement 3

In §1, the idea that lexeme formation patterns selectively apply to lexemes even when the latter (arguably) belong to the same lexical configuration has been put forward. However, data exist that show that derivational rules may apply unselectively in precisely this situation. Fradin & Kerleroux (2009, 87) already pointed this out about the French adjective RARE 'rare'. The verb OUVRIR 'open' is another case. It exhibits more than thirty constructions if reflexive constructions are taken into account (TLFi). Nevertheless, OUVERTURE 'opening' is the only eventuality denoting noun correlated with the various lexemes OUVRIR. This example is an exception however because OUVERTURE was never derived from OUVRIR: it corresponds to a form inherited from Late Latin (*opertura* < Classical Latin *apertura*), which has been correlated with the major meanings of the verb from the Old French. The high frequency of the word may have prevented people from coining derived lexemes with another exponent as new verbal meanings arose.

If we choose a less suspicious case, we see that non-selectivity generally limits itself to a subpart of the lexemes included in a lexical entry. For instance, the verb ENFILER displays three

main constructions, given in (8)-(10)(base-N = *fil* 'thread').

(8) X[AGT] ENFILER$_1$ Y[PAT, FIG, thread] (dans Z[GRND])
     'X thread Y (onto Z)
     enfiler une corde dans une poulie 'thread a rope onto a pulley'

(9) X[AGT] ENFILER$_2$ Y[PAT, FIG] (sur Z[GRND, thread])
     'X thread Y (on Z)
     enfiler des pièces d'or sur un fil 'thread gold coins on a thread'

(10) X[AGT, FIG] ENFILER$_3$ Y[street, GRND]
     Ils enfilaient des rues sombres. 'They took dark streets'

ENFILER$_1$ and ENFILER$_2$ are conceptually very close to each other. What makes them distinct is the spatial argument that denotes the thread: the figure or the ground. ENFILER$_3$ on the other hand, involves a metaphorical extension of the schema involved by ENFILER$_2$: there is no thread but a path and the figure is the agent. The derived nominals that are attested reflect this conceptual shift, as shown in examples (11)-(13).

(11) ENFILER$_1$
     l'enfilage d'une barre sur toute sa longueur (TLFi)
     l'enfilement du fil dans le chas de l'aiguille (Web, 3.2019)
     les enfileurs de soie dentaire (Web, 3.2019) = INS

(12) ENFILER$_2$
     l'enfilage des chaussettes (Web, 4.2019)
     l'enfilement d'un vêtement sur un bébé rétif (Web, 3.2019)
     les enfileurs de perles, de mots (Web, 4.2019) = AGT

(13) ENFILER$_3$
     Ø

No nominalization corresponds to ENFILER$_3$; *enfilade* 'row' is correlated with the stative use of ENFILER$_2$, since this N denotes 'une suite de choses enfilées' 'a sequence of things forming a string' e.g. *une enfilade de quais* 'a succession of docks'. What this example suggests is that lexemes heading constructions which share an identical conceptual setting (here: (i) thread-like object, element with a hole, agent) tend to have the same set of derived nominals. The emerging hypothesis is that selective derivations, such as those in §2, are more likely to take place when the lexically related lexemes in question do not share the same conceptual structure or when the latter is parametrized differently. For instance, ENFILER$_3$ does not share conceptual setting (i); as for FONDRE, the initial setting involves a change of state, which is parametrized along the agentivity dimension (patient, agent / actor), the spatial dimension (FONDRE$_4$ vs. others), and the nature of the change (solid > liquid FONDRE$_1$, FONDRE$_2$, FONDRE$_4$; solid > liquid > solid FONDRE$_3$). Changing the parameters changes the nature of the verb: FONDRE$_3$ is a creation verb and, as such, introduces a new set of arguments (agent, result, instrument, location).

This abstract investigates the way lexemes that belong to a lexical entry can be lexically related on the basis of small set of French verbs, the description of which is reliable (*fondre, doubler, hausser, enfiler,* etc.). It attempts to see whether several types of correlation exist. In particular, it aims to shed light on the role of conceptual settings that are idiosyncratic and inherent compared with those that result from the instantiation of general parameters that appear all over the grammar (agentivity, change of state). I propose that the derivational capacities of each lexeme included in a lexical unit and their distribution will provide us with clues about the links these lexemes have with one another and, consequently, about the organization of the lexicon above the lexematic level.

5 Statement 4

The lexicon also includes polylexematic units of category N, V or predicative PP. Since these almost never have competing units with the same PID value, situations discussed in §§1-3 never occur for them. Some are lexemes and constitute separate lexical entries. They are dealt with as such by derivation e.g. N: *fil-de-fer* 'wire' → *fil-de-fériste* 'tightrope walker', *long cours* 'long range' → *long-courrier* 'long-haul'. Those that are lexicalized phrases appear under their head constituent in the lexicon. Crucially, whereas the derived meaning is built on the meaning of the whole phrase, the derivational exponent is put on the more discriminant lexeme of the phrase, which is not always the head e.g. PP *en vrac* 'in bulk' → *vraquier* 'bulk carrier', V *mettre en scène* 'to stage' → *mise en scène* 'staging', N *camp de concentration* 'concentration camp' → *concentrationaire* 'concentration N'.

References

Ackerman, Farrell & John Moore. 2001. *Proto-properties and grammatical encoding. a correspondence theory of argument selection.* Stanford: CSLI.

Aronoff, Mark. 1994. *Morphology by itself.* Cambridge: MIT Press.

Bonami, Olivier & Berthold Crysmann. 2018. Lexeme and flexeme in a formal theory of grammar. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 203–235. Berlin: LSP.

Davis, Anthony R. 2001. *Linking by types in the hierarchical lexicon.* Stanford: CSLI.

Fradin, Bernard & Françoise Kerleroux. 2009. L'identité lexémique. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 85–104. Saint-Denis: Presses Universitaires de Vincennes.

Gamerschlag, Thomas, Wilhelm Geuder & Wiebke Petersen. 2014. *Glück auf, der Steiger kommt* — a frame account of extensional and intensional *steigen.* In Doris Gerland, Christian Horn, Anja Latrouite & Albert Ortmann (eds.), *Meaning and grammar of nouns and verbs*, 115–144. Düsseldorf: DUP.

Goddard, Cliff. 2000. Polysemy: A problem of definition. In Yael Ravin & Claudia Leacock (eds.), *Polysemy. theoretical and computational approaches*, 129–151. Oxford: OUP.

Jurafsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72. 533–578.

Katz, Jerrold J. 1972. *Semantic theory.* New York: Harper & Row, Publishers.

Kerleroux, Françoise. 2004. Sur quels objets portent les opérations morphologiques de construction. *Lexique* (16). 85–123.

Koenig, Jean-Pierre. 1999. *Lexical relations.* Stanford: CSLI.

Lakoff, George. 1987. *Women, fire, and dangerous things.* Chicago: University of Chicago Press.

Plag, Ingo, Marios Andreou & Lea Kawaletz. 2018. A frame-semantic approach to polysemy in affixation. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 467–486. Berlin: LSP.

Rappaport Hovav, Malka & Beth Levin. 2010. Reflections on manner/result complementarity. In Malka Rappaport Hovav, Edit Doron & Ivy Sichel (eds.), *Syntax, lexical semantics, and event structure*, 21–38. Oxford: Oxford University Press.

Sag, Ivan. 2012. Sign-based construction grammar. an informal synopsis. In Hans Boas & Ivan Sag (eds.), *Sign-based construction grammar*, 69–202. Stanford: CSLI Publications.

The direction of analogical extensions in the verbal roots of Old French and Old Florentine Italian: a corpus study

Sascha Gaglia

Georg-August-Universität Göttingen

1 Introduction

Old French shows a great amount of analogical extensions of verbal root allomorphs (Fouché 1967; Walker, 1987; among others). Some of the involved paradigms where levelled in later stages, for example O. Fr. *trover* (Mod. Fr. *trouver*) 'find', *paroler* (*parler*) 'talk', *amer* (*aimer*) 'love', *plorer* (*pleurer*) 'cry' and *lever* 'lift/elevate'. Other verbs like *venir* 'come' and *tenir* 'have/hold' still show some significant allomorphic patterns in Modern French. Not all of these paradigms were extended in the same way. Where some paradigms generalized the unstressed root (e.g. *trover, paroler*), others generalized a stressed one (e.g. *amer, plorer, lever*). Hence, we know the starting point for analogical extensions and the situation in Modern French as well as the diachronic developments in general. However, a thorough quantitative analysis of analogical extensions based on Old French corpus data is still missing, which is the reason why it is still not possible to unveil the 'path' of each extension in the sense of paradigmatic patterns, subparadigms and paradigm cells. The same also holds true for Old Italian. Having this in mind, I will present some important findings from my ongoing project on the paradigmatic extension of verbal roots in French and Italian funded by the German Research Council (DFG). My first goal here is proposing a quantitative analysis based on corpus data of the direction of the main analogical extensions for the morphomic patterns N, L and U (Maiden 2018). This will be conducted for each century of the Old French period and paradigm cell by paradigm cell, if possible. For a better understanding and to strengthen our observations, I shall include also the Old Florentine Italian verbs *venire* 'come' and *tenere* 'have/hold'. On the basis of these data I will then test some of the factors discussed in the literature. They comprise phonological markedness, morphomic attraction and stability as well as frequency. Moreover, I will discuss one additional factor, namely semantic-syntactic feature specification. I shall show that our data indicate that this factor is relevant to the direction of analogical extensions.

2 Influencing factors (previous research)

The following three factors have been often mentioned while trying to account for the direction of the analogical extension of Romance verbal roots, especially with respect to Old French:

Phonological markedness: In most case, it was the phonologically unmarked root of a paradigm which was extended, e.g. Mod. Fr. *trouver* exhibits the Old French unstressed roots *trouv-* throughout the paradigm, while the segmentally more marked stressed roots got lost (e.g. *truev-/treuv-/truis*(*s*)-/*troiss-*). Mod. Fr. *aimer* 'love' and *pleurer* 'cry' are exceptions due to possible interparadigmatic analogies on the grounds of **esmer* 'esteem' and *pleuvoir* 'rain' (Esher, 2016). In the case of *aimer* and *pleurer*, the stressed roots served as basis for analogical extension.

Frequency: Verbs like O. Fr. *amer* lost their N-pattern distribution of root allomorphs (i.e. 1SG, 2SG, 3SG/3PL vs. 1PL/2PL present indicative/subjunctive) since this pattern had a low type frequency in the first conjugation class of Old French (Esher, 2016). I shall discuss the possible effects of type frequency and of token frequency, which have been proposed in more than one occasion to serve as factor for analogical extension or for being responsible for the resistance of a form against analogical extension (see Bybee 1985, among others).

Morphomic stability is a major factor constraining the direction of the extension of a root in the sense of determining the paradigm cells in which an extension may or may not take place (see 'coherence' in Maiden 2018).

3   Research questions

In line with the aims of the present paper and building upon the factors addressed in previous research, there are two research question which shall be discussed here:

1) Which of the above-mentioned factors can be shown to influence the direction of analogical extensions in Old French and Old Florentine Italian verbal roots by means of a quantitative corpus analysis?
2) Does semantic-syntactic feature specification matter?

4   Methodology

4.1 Data

With these goals in mind, all variants of verbal root allomorphs of the aforementioned Old French verbs *trover, amer, plorer, venir* and *tenir* where extracted from the *Noveau Corpus d'Amsterdam* (= NCA), with a total number of 24,702 tokens. (NCA covers the entire Old French period, i.e. from 1100 to 1350, and includes 3,184,834 words from 299 texts.) The Old Florentine Italian data was taken from the *Corpus OVI dell'italiano antico* (= OVI), including the period from 1100 to Giovanni Boccaccio's death in the year 1375, which conventionally marks the end of the Old Italian period (23,176,174 words from 2,335 texts). Due to practical reasons, only two verbal paradigms were investigated with respect to Old Florentine Italian, i.e. *venire* and *tenere* (5,758 tokens). While data from the NCA could be easily extracted for each paradigm cell together with relevant manuscript metadata via the tool TigerSearch, this was not the case for our Old Florentine Italian data, which I transferred token by token to a Filemaker database and then annotated each token with the relevant information. Hence, the amount of our Old Florentine Italian is significantly smaller than compared to our investigated data from Old French.

To compare similar phenomena for the verbs in question, I divided the lexemes in three groups (1, 2, 3), with respect to the main morphomic patterns they adhere to (i.e. N, L, U; L' corresponds to the DARK-L pattern in Esher, 2016). Moreover, the groups 1 and 2 contain two subgroups (A, B) regarding the direction of analogical extension in prosodic terms (i.e. extension of the stressed or unstressed root). The differences between groups 1 and 2 on the one hand and 3 on the other hand are to be traced back to the overall differences between Old French and Old Florentine Italian. Table 1 (see on next page) represents mainly patterns that concern the present indicative and the present subjunctive and which are the domain of the patterns in question. Analogical extensions that go beyond the present tenses are not included in the table and are instead discussed below. (I use the following abbreviations in Table 1: R_M = root including a monophthong, e.g. Fr. *trouv-* [truv-]; R_D = root including a

diphthong, e.g. Fr. *vien-* [vjɛ̃(n)-]; R_{VEL} = root including a final velar consonant, e.g. It. *veng-* [veŋg-]; $R_{3PL\,PRS.IND}$: extension of a 3PL PRS.IND root, e.g. denasalized Fr. [vjɛn-].)

4.2 Statistical analysis

The ortography of medieval texts is far from being standardised. Moreover, these can include diatopic and stylistic variation. Hence, each cell of the paradigm of a given lexeme can display so-called 'overabundance', i.e. it can have more than one surface representation; see Thornton (2012), who reasonably uses the notion 'cell-mate' instead of 'doublet' for each overabundant form.

Table 1. *Analysed verbal groups*

Group	Patterns	Extended Root	Paradigms (Tokens)	Variety
1A	L >> levelled	R_M (unstressed)	*trover* (4,250)	Old French
1B	L >> L'	R_D (stressed)	*venir* (9,728) *tenir* (2,968)	
		$R_{3PL\,PRS.IND}$ (stressed)		
2	N >> levelled	R_D (stressed)	*amer* (913) *plorer* (1,085)	Old French
3	U >> U	R_{VEL} (stressed)	*venire* (3,289) *tenere* (2,469)	Old Fl. Italian

I, therefore, summarized all cell-mates of a paradigm cell that show the same morpho-phonological phenomenon as in Table 1; e.g. *truev-, troev-, trueu-* are variants of a phonologically diphthongized root (R_D), while *truv-* and *trouv-* phonologically display a monophthong (R_M) in the case of *trover*. Our statistical analysis takes into consideration (a) percentages representing the proportion of analogically extended roots vs. non-analogically extended roots in each relevant cell of the paradigm for a given verb (e.g. R_M vs. R_D in the case of *trover*; our method is comparable to the count of cell-mate ratios in Thornton, 2012), and (b) a normalized count (or *normalized frequency*) of each analogically extended root vs. non-analogically extended roots in relation to corpus size. Every count was limited to time intervals corresponding to the 12th, 13th and 14th century, respectively.

5 Main Results

The following results shall be discussed in order to answer our research questions (see § 2 above):

First, each of the three assumptions found in the literature on the direction of analogical extension can be confirmed by our data: By comparing the verbs of groups 1A and 1B, I will show that in the case of 1SG PRS.IND the root [trɥis] (paradigm of O. Fr. *trover*) behaves more stably towards its extinction than [vɛ̃ɲ] (sometimes [vĩɲ]) and [tɛ̃ɲ], the 1SG forms of O. Fr. *venir* and *tenir*, respectively, which means that the analogical extension of stressed [vjɛ̃(n)-] and [tjɛ̃(n)-] was faster than the extension of unstressed [truv-] (*trover*) towards this cell of the paradigm, as the normalized frequencies in Table 2 show:

Table 2. *Analogical extension to 1SG PRS.IND (normalized frequencies)*

Century	[truv-]	[vjɛ̃(n)-]	[tjɛ̃(n)-]
12th	0.00	1.23	0.41
13th	0.00	1.68	1.97
14th	0.08	2.00	2.88

In line with Bybee's assumption of morpho-phonemic irregularity (Bybee 1985), I claim that this has to do with the fact that [trɥis] was phonologically more marked than [vẽɲ] and [tẽɲ]. Moreover, I claim that the morphomic stability of the L-pattern contributed simultaneously to the relative stability of [trɥis], although due to their low type frequency in Old French, most of the morphomic patterns got lost very early from first conjugation class verbs like *trover* (see also Esher, 2016 with respect to the N-pattern). In the paper I will also argue against token frequency as being responsible for the stability of [trɥis] and, in general, discuss it with respect to 1SG PRS.IND (against Bybee 1985; see more evidence below in this abstract). Morphomic stability is also at issue when comparing the verbs of 1B and 2. It can be shown that wherever a subparadigm was not shaped by a morphomic pattern (i.e. future, conditional, weak preterite forms etc.), analogical extension occurred in these subparadigms first. From there, analogical extension spread to the morphomically shaped present indicative and present subjunctive of *amer* and *plorer*. This confirms, in principle, the findings in Fennell (1975) with respect to the analogical extension of verbal roots in the future tense and conditional. However, our data show that this extension occurred much earlier than assumed in Fennell's corpus analysis of Middle French.

Second, a comparison of the developments affecting the verbs of group 1B and 3 reveals that the Old French denasalised root [vjɛn-] extended first from 3PL PRS.IND to 3PRS.SBJ, which also holds true for O. Fl. It. [veŋg-] (paradigm of *venire*), as shown by its including a final velar consonant (the predecessor was [veɲ-] as a result from sound change). O. Fl. It. *tenere* appears to deviate slightly from this 'path' in that the 3SG PRS.SBJ seems to undergo analogical extension slightly prior to 3PL PRS.SBJ with respect to [teŋg-]. I claim that these results serve as a further piece of evidence against the assumption that token frequency determines the direction of extension, which has already been questioned by Fertig (2013). If token frequency would be a determinant, we would first expect 3SG PRS.SBJ to undergo analogical extension due its higher frequency compared to 3PL PRS.SBJ. (See also the results for token frequency in spoken and written modern Spanish in Bybee 1985 with respect to Person and Number, even if her data concern only the present indicative and preterite.) I propose in line with our resarch question 2 that the path 3PL PRS.IND to 3PL PRS.SBJ can only be explained by taking into consideration the semantic-syntactic generalization of a form, which means that O. Fr. [vjɛn-] and O. Fl. It. [veŋg-] were extended from 3PL PRS.IND to 3PL PRS.SBJ due to their syntactic feature specification being PERS = 3, NUM = PL (or semantically *reference = bystander, cardinality = x > 1*) and TENSE = PRS. In the case of *tenere*, the generalization could have worked for PERS = 3 and TENSE = PRS. All analysed groups (see Table 1) exhibit a third person cell being the first affected by analogy. This can especially be shown for group 1A and 1B: These are the groups in our sample where analogical extension could also affect 3SG PRS.IND, being the cell which has the highest frequency.

Bibliography

Bybee, Joan L. 1985. *Morphology. A study of the relation between meaning and form.* Amsterdam: Benjamins.

Esher, Louise. 2016. Morphome death and transfiguration in the history of French. Journal of *Linguistics* 53. 51-84.

Fennell, T. G. 1975. *Morphologie du futur en moyen français.* Genève: Droz.

Fertig, David. 2013. *Analogy and morphological change.* Edinburgh: Edinburgh University Press.

Fouché, Pierre. 1967. *Morphologie historique du français. Le verbe.* Paris: Klincksieck.

Maiden, Martin. 2018. *The Romance verb.* Oxford: Oxford University Press.

NCA = Stein, Achim. et al. 2011. *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français (ca 1150-1350), établi par Anthonij Dees (Amsterdam 1987), remanié par Achim Stein, Pierre Kunstmann et Martin-D. Gleßgen.* Stuttgart: Institut für Romanistik/Linguistik.

OVI = Larson, Pär et al. 2005. *Corpus OVI dell'Italiano antico.* Istituto Opera del Vocabolario Italiano/CNR.
(http://gattoweb.ovi.cnr.it/(S(gpccezaigip4ix55l2qo4lqx))/CatForm01.aspx)

Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5/2. 183-207.

Walker, Douglas C. 1987. Patterns of analogy in the Old French verb system. *Lingua* 72. 109–131.

In defense of the "phrasal compounds involve quotations" thesis

Daniel Gutzmann   *Kathatrina Turgay*
University of Cologne   University of Landau

Phrasal compounds as a problem for the architecture of grammar   Phrasal compounds (PCs) can be defined as compounds in which one of the immediate constituents (the first one, in most cases) is a phrase. Even though not completely unrestricted, the left constituent can be of almost any category. The head most often is a noun, but it does not have to be.

(1) a. das [$_{CP}$ Für-den-Mathespick-von-heute-lade-ich-dich-ein]-Angebot
        'The I-invite-you-because-of-math-cheat-of-today offer'

  b. ein [$_{PP}$ für-umsonst]-Monat
        'the for-free moth'

  c. die [$_{VP}$ Schnell-leben-und-jung-sterben]-Einstellung
        'the live-fast-and-die-young attitude'

This very definition – words that contain phrases – runs counter some core assumptions of many traditional models of grammar. Many theories of the architecture of grammar are characterized by being *linear* and *modular*. According to the concept of **modularity**, the language system consists of various, rather independent subsystems, or modules. For instance, there is morphology (the "word system") and there is syntax (the "phrase system"), both of which exhibit module-specific rules and work on module-specific objects (i.e. morphemes vs. phrases). According to the concept of **linearity**, the different subsystems are not just separated from each other but are subject to a linear order: One module is active first and generates as its output the input for the next module.

   More concretely, with respect to the the relation between morphology and syntax, these two assumptions lead to the following two constraints:

(M) **Morphologhy ≠ syntax**
      They are different modules and work on different objects. (modularity)

(L) **First morphology, than syntax**
      Morphology provides the input for syntax. (linearity)

Taken together, this rules out PCs. Yet they do exist. Even though the reactions to this problem have been plentiful and diverse, they can be grouped into three classes of responses, depending on which of the two concepts they restrict or even reject.

(RM) Giving up modularity: morphology = syntax (Lieber, 1992; Sproat, 1985)

(RL) Giving up linearity: syntax can provide input to morphology (Ackema & Neeleman, 2004; Meibauer, 2003, 2007)

(RB) Giving up both: holistic, construction-based approaches (Hein, 2015)

Because phrasal compounds lead to such far-reaching consequences for the architecture of grammar, they have been called a "touchstone" for theories of the morphology-syntax interface (Hein, 2015, 56). This idea is precisely what we argue against in this talk. Even if there may be other reasons to reject modularity or linearity (we do not discuss those here), we argue that the existence of phrasal compounds is not a problem for linearity and modularity. We do

this by resurrecting the so-called quotation analysis of phrasal compounds argued for by Wiese (1996) and according to which the phrasal constituent is a quotation. We aim to show that the arguments put forward in the literature against this approach are not convincing, even if they have been quite successful at stopping the adoption of the quotation analysis. Moreover, we also want to show that the quotation analysis can account for some of the particular properties of phrasal compounds for which the other, more radical approaches cannot easily offer an explanation; at least not without further assumptions. This will provide arguments for why the quotation analysis also has arguments in favor of it.

The quotation analysis of phrasal compounds   In light of the challenges raised by PCs, Wiese (1996) proposed an attractive way out of the conundrum: the phrasal part of a PC is a quotation. For that reason, as Wiese argues, it acts as a single unit for the purposes of morphology as its internal structure is inaccessible. Hence, PCs do not pose a problem for linearity, since the phrasal component is not actually part of the compound. Wiese illustrates this as follows:

(2) $[_{X^0} [_{Y^0} [_{``CP"}$ Für den Mathespick von heute lade ich dich ein $]]$ Angebot $]$

The syntactic category of the first conjunct cannot be seen by morphology, which is here symbolized by the quotation marks around the syntactic node. The entire quotation then functions as a word for the purposes of morphology (the category of which does not matter according to Wiese). Wiese presents various arguments for why the quotation hypothesis is warranted. First, many PCs contain material from foreign languages as in (3) or even non-linguistic material as in (4) in the phrasal slot. And in these cases, nobody seriously would argue that these should become part of morphology. So why assume this for phrasal compounds from the same language?

(3)   Zur **"laissez faire"** Haltung in der Erziehung passt dann konsequenter Weise auch die **'C'est la vie'** Haltung.
      'The 'C'est la vie' attitude consequently matches the 'laissez faire' Haltung in education.'

(4)   a.   In the next round, the ♥-team plays against the ♥-team.
      b.   The [*speaker performs a sound*]-sound kept me awake all night.

Nobody would conclude from cases like these that the quoted material must be part of morphology, one shouldn't draw such a conclusion for PCs. And arguably these cases all involve quotations of some sort.

Arguments against the quotation analysis   The quotation analysis involves two assumptions: i) the phrasal component is a quotation ("ZP" in (2)), and ii) the quotation behaves like a word inside the compounds (Y^0 in (2)). Both assumptions were attacked in the literature. First, the phrasal part can be completely new material (as in (1)) and hence assuming that it is a quotation is not warranted (Meibauer, 2007, 240). Secondly, the first part does not behave like words usually do. For instance, compounds are assumed to be "anaphoric isles", which means that one cannot refer to just the first constituent. But with PCs, this is possible.

(5)   a.   * Jedes **Mutter$_i$**söhnchen möchte am liebsten für immer bei **ihr$_i$** wohnen bleiben.
            'Every mother boy wants to live with her forever.'
      b.   Jeder Meine-**Mama$_i$**-ist-die-Beste-Sohn möchte am liebsten für immer bei **ihr$_i$** wohnen bleiben.
            'Every ''Mommy is the best'-boy wants to live with her forever.'

The two arguments were taken to be knock-down arguments against Wiese's quotation analysis and hence that approach did not play a major rule in the following discussion regarding the status of PCs in grammar.

What are quotations? We think that the arguments against quotation analysis are not as forceful as presented in the literature; at least not if we adopt a more sophisticated approach to what quotations are. We follow the work championed by (Recanati, 2001), who in turn bases his ideas on the work by Clark & Gerrig (1990). Accordingly, quotations are basically linguistic demonstrations: the speaker produces linguistics material not to use it the ordinary way, but to demonstrate something with it. Just like that is possible with non-linguistic material.

(6)   And then I showed them how I dance [*speaker demonstrates dancing moves*].

(7)   And then I told him my opinion. "That's bullshit, Hans!"

Of course, we can demonstrate something nobody said before to illustrate a point. Hence, we should not understand the notion of quotation in the narrow sense of "repeating something already said".[1] Crucially, the "target" of the demonstration must be inferred by the hearer: what are the relevant aspects of the demonstration?

   Beside assuming that quotations are demonstrations, Recanati distinguishes between *open* and *closed* quotation. A quotation is closed if the linguistic material is "linguistically recruited" and then takes up a nominal slot in the sentence frame. If it doesn't, a quotation is open (Recanati, 2001, 649)

(8)   Stop that John! 'Nobody likes me', 'I am miserable' … Don't you think you exaggerate a bit? (*open quotation*)

(9)   John keeps crying and saying 'Nobody likes me'. (*closed quotation*)

In closed quotations, the syntactic category of the quoted material does not matter, it always takes up the same kind of slot. As this is also the case for the phrasal part in PCs, we assume that PCs involve closed quotations. For Recanati, there are three levels of meaning active in closed quotation: *a*) the linguistic meaning of the displayed material, *b*) the meaning of the demonstration, and *c*) the referential value of recruited demonstration. The idea is that by using a quotation, the speaker produces linguistic material which of itself has some meaning (*a*). However, the demonstration itself also has some meaning; namely the very target of the demonstration (*b*). Thirdly – and this is the linguistic recruitment – in closed quotation the quotation fills in a nominal slot and refers to target of the demonstration (*c*). So what is crucial here is that the quoted material is not really part of the sentence itself, it is a demonstration. However, one can use a demonstration at a nominal slot to refer to the target of the demonstration. This is exactly what happens in cases like (4b): the speaker produces a sound and uses it at precisely the point where she wants the meaning of that demonstration to be referred to by the sentence.

PCs and quotation Applying such a view on quotation to PCs has some interesting ramifications. First, Wiese's original idea still holds: the quoted material is not really part of the sentence. But with Recanati's theory we have a more precise description of this. The quoted material is not part of the linguistic material, because it is merely demonstrated and demonstrations are not part of the linguistic material. However, the demonstration can be "recruited" into a nominal slots, as (closed) quotations in general are nouns (Pafel, 2011). This means, PCs with a nominal head are N + N-compounds after all.[2] It is not the fact that the first part is

---

[1]Wiese (1996, 188) already said that he had some more abstract notion of quotation in mind. Hence this argument against the quotation analysis seems to be like a straw man argument.

[2]This is compatible with Gallmann's conversion analysis of PCs (Gallmann, 1999).

phrasal that makes PCs special, but that it is quoted.

As already outlined above, this view on quotation does not require that the quoted phrases were uttered before; they only have to be demonstrated (in order that the recruitment can refer to a property associated with it). Hence, the first counter argument against the quotation analysis does not apply.

The second counterargument does also not apply. If the linguistic material is demonstrated, we actually expect it to be possible to refer to aspects of it, since it is not wrapped inside an "anaphoric island". That is, reference in cases like (5b) is only seemingly to parts of the word, whereas it is actually to the external demonstration.

Adopting Recanati's theory of quotations lets debunk the two counterarguments. However, there is also a strong argument in favor of the a quotation analysis, which is rather obvious but didn't receive much attention in the literature. The fact that indexical expressions in PCs do not refer to aspects of the utterance context (but to some imagined context) is directly accounted for if the phrasal part is a quotation.

(10)   Dieser "Meine-Mama-ist-die-Beste"-Kerl ist total unselbstständig.
        'The 'my mommy is the best'-guy is totally unindependent.'

        ≠   the speaker's mother
        =   mother of the guy

That indexicals can shift in quotations is a well known fact and hence not surprising to occur in PCs if the quotation analysis is adopted. In contrast, approaches that do not analyze the first constituent as a quotation cannot explain this without further assumption. We take this as additional evidence that a quotation analysis of PCs is on the right track.

References

Ackema, Peter & Ad Neeleman. 2004. *Beyond morphology* (Oxford Studies in Theoretical Linguistics 7). Oxford: Oxford University Press.

Clark, Herbert H. & Richard J. Gerrig. 1990. Quotations as demonstrations. *Language* 66(4). 764–805. http://www.jstor.org/stable/414729.

Gallmann, Peter. 1999. Wortbegriff und Nomen-Verb-Verbindungen. *Zeitschrift für Sprachwissenschaft* 18(2). 269–304.

Hein, Katrin. 2015. *Phrasenkomposita im Deutschen*. Narr.

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

Meibauer, Jörg. 2003. Phrasenkomposita zwischen Wortsyntax und Lexikon. *Zeitschrift für Sprachwissenschaft* 22. 153–188.

Meibauer, Jörg. 2007. How marginal are phrasal compounds? *Morphology* 17(2). 233–259. doi:10.1007/s11525-008-9118-1.

Pafel, Jürgen. 2011. Two dogmas on quotation. In Elke Brendel, Jörg Meibauer & Markus Steinbach (eds.), *Understanding quotation* (Mouton Series in Pragmatics 7), 249–276. de Gruyter Mouton.

Recanati, François. 2001. Open quotation. *Mind* 110. 637–687. doi:10.1093/mind/110.439.637. http://jeannicod.ccsd.cnrs.fr/docs/00/05/33/06/PDF/ijn_00000146_00.pdf.

Sproat, Richard. 1985. *On deriving the lexicon*. Cambridge, MA: MIT Press.

Wiese, Richard. 1996. Phrasal compounds and the theory of word syntax. *Linguistic Inquiry* 27(1). 183–193. http://www.jstor.org/stable/4178931.

The curious case of wandering case morphemes

Stefan Hartmann
University of Bamberg

1   Introduction

This paper investigates non-canonical uses of the German circumposition *um … willen* as well as the postposition *wegen*[1]. Both adpositions usually govern genitive case, although there is quite some variation, especially in the case of *wegen*. (1) illustrates typical uses of both adpositions. However, in non-standard data, we also find numerous cases like (2), where the genitive morpheme *-s* is "relocated" to the postposition.

(1)   um des Frieden-s   willen / des Frieden-s   wegen
      for the peace-GEN sake   / the peace-GEN due-to

      'for the sake of peace'

(2)   um des Frieden willen-s   / des Frieden wegen-s
      for the peace   sake-GEN / the peace   due-to-GEN

      'for the sake of peace'

The aim of the present paper is to test a hypothesis that can account for this phenomenon, based on data from the 20-billion-word webcorpus DECOW16B (Schäfer & Bildhauer, 2012). Specifically, I argue that these variants can be explained by the principle of cleft-formation (see e.g. Ronneberger-Sibold 1997), which Nübling et al. (2017, 117) see as the most important syntax-typological feature of German. On this view, German is characterized by cleft structures both at the phrasal and at the sentence level.[2] It is often assumed that cleft-formation facilitates language processing by highlighting syntactic structures. This is highly compatible with evidence from psycholinguistic studies, according to which language users continuously predict what their interlocutor is going to say next ("forward modelling", Pickering & Garrod 2013). Cleft structures offer a special potential for forward modelling as they invite language users to make predictions about the continuation of a phrase or sentence.

The cleft structure of noun phrases in German is particularly salient in genitive noun phrases with masculines or neuters belonging to the strong or mixed declension class, which have *-s* as genitive singular marker. Here, as Zimmer (2018, 67) points out, the rightmost element of the noun phrase is highlighted by the genitive-*s* that agrees with the determiner, e.g. **des neu-en Auto-s** 'of the new car'. In a way, the *wegens* and *willens* constructions discussed in the present paper mimic this structure: By relocating the case morpheme to the rightmost element of the noun phrase, the cleft structure is made salient. As such, it seems plausible to assume that a principle like "the genitive-*s* occurs at the rightmost element of the cleft" is part of German native speakers' implicit linguistic knowledge, which would offer an explanation for the phenomenon of the "wandering" genitive-*s* in the *um … willen(s)* and *wegen(s)* constructions.[3]

---

[1]*wegen* is also – and much more frequently – used as a preposition in modern German, but I focus on the postpositional use here: Prepositional *wegen* can arguably considered a different construction, and the non-canonical use discussed here is more characteristic of the postpositional use.

[2]Ronneberger-Sibold (1997) uses the term "framing" for this phenomenon and "frames" for the cleft structures in question. As these terms are used in a different sense in various other domains, I avoid them here.

[3]I have borrowed the metaphor of "wandering" case morphemes from Eisenberg (2013, 173), who in turn paraphrases Fourquet (1973).

2   Data, methods, hypotheses and predictions

The DECOW16B corpus was searched exhaustively for instances of *wegens* and for instances of *um … willens* in a distance of max. 6 words. After manual deletion of false hits and duplicates, 979 instances of *um … willens* and 351 instances of postnominal *wegens* remained in the data. For comparison, samples of 5000 attestations each for *um … willen* and *wegen* were extracted from DECOW16B. Again, false hits were deleted, and 4300 instances of *um … willen* as well as (only) 115 instances of postnominal *wegen* remained in the data. In the remainder of this paper, I will collectively refer to *um … willens* and *wegens* as the non-canonical constructions and to *um … willen* and *wegen* as the canonical ones. The data were coded for a number of variables to be discussed in more detail below. The aim is to test two predictions which follow from the hypothesis that the drive towards cleft-formation is the main motivation for the displacement of the *s*-morpheme:

(i) There are significantly more masculine and neuter nouns in the *wegens*- and *um…willens*-data than in comparison datasets with the canonical variants, as the genitive-*s* only occurs in the genitive singular of strong masculines and neuters. (ii) We find a significantly higher proportion of non-canonical s-less genitives in the *wegens*- and *um…willens* data than in the comparison datasets: While *s*-omission is quite common especially in the case of low-frequency words, loan words, and proper nouns (Zimmer, 2018), we can predict that it occurs more frequently in combination with the non-canonically *s*-suffixed postpositions.

The hypotheses are tested using logistic regression modeling. For prediction (i), a mixed binomial logistic regression model is fit to the data, operationalizing "gender" (feminine / non-feminine) as a binary response variable. For prediction (ii), the method of CART trees and random forests is used, which is well-applicable to data with relatively few observations but many variables (Tagliamonte & Baayen, 2012, 161). The presence or absence of a genitive-s is used as the response variable (only strong and mixed masculine and neuter nouns enter the model, i.e. only nouns where a genitive-s is possible). As predictor variables, the factors that have emerged as significant predictors for s-lessness in Zimmer's (2018) multifactorial study were used: a) whether the noun in the NP slot is inherently monoreferent (which is true for proper names but also for common nouns referring to unique entities like *Grundgesetz* '(German) constitution'), b) whether it is used as an apposition, e.g. *des **Bundeskanzler(s)** Adenauer* 'of-the Federal Chancellor Adenauer', c) whether it is a proper noun, a non-native word, or a short word (acronym/abbreviation) ending in -*s*. These lexemes often remain uninflected in German, a phenomenon sometimes referred to as *Schonungsbedarf* (roughly: 'in need of conservation', i.e. language users tend to "conserve" the phonological structure of the word), which is why Zimmer (2018) refers to this variable as "Schonungsbedarf + s". In addition, d) frequency has a significant effect on s-lessness: Middle- and low-frequency words drop the genitive-*s* more often. To weigh the impact of these variables against that of the variant, the aforementioned predictors are complemented by the predictor "Variant" in the random forest model used here. Note that only datapoints with masculine or neuter nouns belonging to the strong or mixed declension classes were included in the model (as only they can vary between variants with and without a genitive-*s*). Also, some further datapoints had to be excluded because the respective lemma does not occur in the DECOW16B lemma frequency list, which was used to determine the frequencies of the individual lexemes.

3 Results

Let us discuss the results for each of the two predictions in turn. Table 1 shows the distribution of grammatical gender across the variants *um ... willen(s)* and *wegen(s)*.[4] In line with prediction (i), the proportion of feminine nouns is much lower in the case of the non-canonical forms (6.6% vs. 37% for *willen(s)* and 6.6% vs. 59.1% for *wegen(s)*). The distribution is quite similar if every lemma type is only counted once. In simple binomial mixed regression models with "Gender" (f vs. m/n) as binary response variable, "Variant" as predictor variable, and "lemma" as random variable, "Variant" emerges as a highly significant predictor both for *um...willen(s)* (Estimate = 2.26, Pr(|z|) = 0.005**) and for *wegen(s)* (Estimate = 23.3, Pr(|z|) < 2e-16). Turning to prediction (ii), the random forest models show that for both constructions, the variant makes a clear difference for the presence or absence of *s*-less genitives (see the CART trees in Figure 1). In the case of *um ... willen(s)*, the variables that proved most influential in Zimmer (2018) emerge as significant predictors of *s*-lessness in the canonical variant. For the non-canonical variant, by contrast, only frequency makes a difference. This is reflected in the measure of conditional permutation variable importance (Strobl et al., 2008): here, "Variant" emerges as the most significant predictor by far. In the case of *wegen(s)*, there are so few canonical uses in the dataset that for the canonical variant, no impact whatsoever of the aforementioned variables can be found[5] This is also why frequency, rather than variant, emerges as the most significant predictor according to the variable importance measure. Still, the random forest model lends strong support to the hypothesis that the genitives of masculine and neuter nouns behave very differently in the non-canonical than in the canonical constructions.
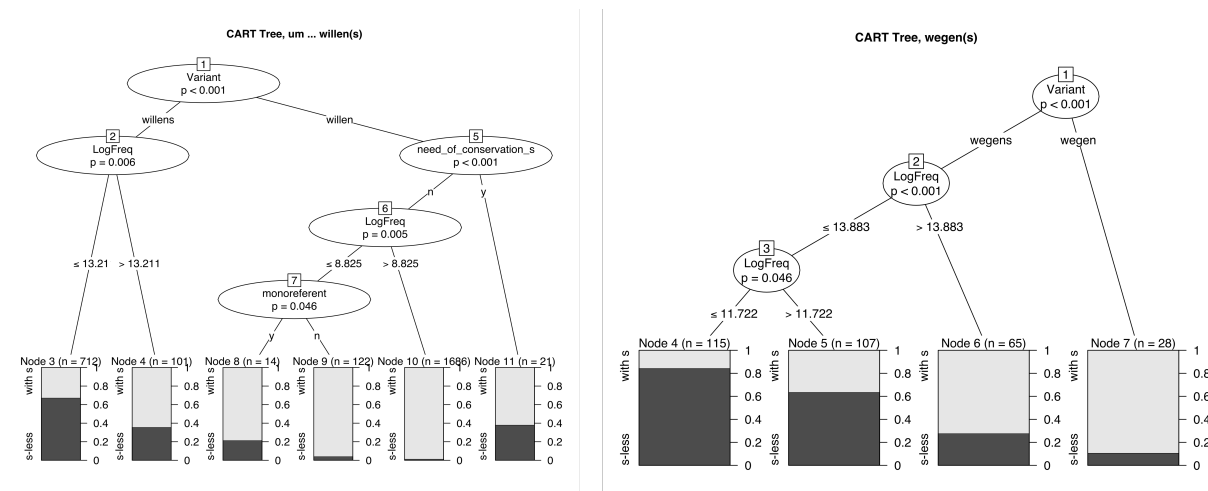
Figure 1: CART Trees for *um ... willen(s)* and *wegen(s)*. Response variable: *s*-lessness of genitives in masculines and neuters of the strong and mixed declension classes.

---

[4]Note that the numbers for *wegen* are based on a 5000-word sample, of which only the postnominal instances were used. This is why the numbers for *wegens* in the table are higher than those for *wegen*, but in fact the attestations for *wegen*, both pre- and postnominal, vastly outnumber those for *wegens*.

[5]They can, however, be found if one includes prepositional uses of *wegen*.

Variant	Freq in *willen* (Types)	Freq in *willens* (Types)	Freq in *wegen* (Types)	Freq in *wegens* (Types)
feminine	1344 (570)	61 (55)	68 (46)	23 (22)
masculine	1614 (294)	308 (89)	23 (22)	235 (125)
neuter	679 (208)	561 (304)	24 (10)	177 (177)

Table 1: Gender distribution

4 Conclusion

The principle of cleft-formation offers an adequate explanation for the emergence of the non-canonical variants discussed here. The distribution of grammatical genders in the *um ... willens* and *wegens* constructions support the hypothesis that the *-s* in the two postpositions actually has its origin in genitive markers that are either fully "relocated" to the rightmost element to the cleft (*um des Frieden willen-s*) or redundantly marked both on the head noun of the NP and the postposition (*um des Frieden-s willen-s*). However, a full account of the constructions discussed here would have to be more complex. For instance, phonological factors as well as the interaction with other, similar constructions seem to play a role as well and should be explored in more detail in future studies. As such, although *um ... willens* and *wegens* are quite infrequent constructions, they offer abundant research possibilities, and they show that small phenomena that can easily be overlooked provide a unique window into native speakers' linguistic intuition.

References

Eisenberg, Peter. 2013. *Grundriss der deutschen Grammatik. Bd. 1: Das Wort.* Metzler 4th edn.

Fourquet, Jean. 1973. *Prolegomena zu einer deutschen Grammatik.* Schwann-Bagel 4th edn.

Nübling, Damaris, Antje Dammel, Janet Duke & Renata Szczepaniak. 2017. *Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels.* Narr 5th edn.

Pickering, Martin J. & Simon Garrod. 2013. An Integrated Theory of Language Production and Comprehension 36(4). 329–347.

Ronneberger-Sibold, Elke. 1997. Typology and the diachronic evolution of German morphosyntax. In Jacek Fisiak (ed.), *Linguistic reconstruction and typology* (Trends in Linguistics 96), 313–335. Mouton de Gruyter.

Schäfer, Roland & Felix Bildhauer. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. In Cicoletta Calzolari, Khalid Choukri, Terry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of LREC 2012*, 486–493.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional Variable Importance for Random Forests 9(307). `http://www.biomedcentral.com/1471-2105/9/307`.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice 24(02). 135–178. doi:10.1017/S0954394512000129.

Zimmer, Christian. 2018. *Die Markierung des Genitiv(s) im Deutschen: Empirie und theoretische Implikationen von morphologischer Variation.* De Gruyter.

Conversion, structured inflection, and the ontological/semantic organization of the lexicon in Oneida

Jean-Pierre Koenig   *Karin Michelson*
University at Buffalo   University at Buffalo

Polysynthetic languages are well known for their inflectional complexity, but it is not clear whether this complexity should be modeled via a large set of position classes or by positing some internal structure. Oneida (Northern Iroquoian) is an interesting test case in this regard since going back to Lounsbury (1953), its verbal and nominal inflectional morphology has been analyzed as involving a flat (i.e. position class-based, slot-and-filler, templatic) combination of morphs as well as a hierarchically organized combination of morphs. (1) diagrams the inflectional structure of verbs: a stem combines first with an aspect suffix and then the result with one of fifty-eight pronominal prefixes (that all occur in one slot or position class) and up to five prepronominal prefixes (from a total of eleven modal and non-modal prefixes that themselves are distributed into 6 position classes). (2) diagrams the inflectional structure of nouns: a stem combines with a noun suffix and the result with a pronominal prefix. (The sets of pronominal prefixes on verbs and nouns are not always identical, hence the subscripts $_V$ and $_N$ in (1) and (2).)

(1)   [(prepronominals)-pronominal$_V$-[stem-aspect]]

(2)   [pronominal$_N$-[stem-noun.suffix]]

Despite the fact that the structure in (1) is traditionally assumed, very little evidence has been provided to support it. Lounsbury (1953) alludes to the fact that modal preponominal prefixes occur only when the punctual aspect suffix occurs and subsequent literature (Chafe 1967 for Seneca and Bonvillain 1973 for Mohawk) to the fact that the aspect suffix can determine (Agent versus Patient) pronominal prefix category. But, in neither case is the evidence particularly convincing as feature co-occurrence restrictions on morphosyntactic feature sets are enough to account for these two dependencies. In this paper, we provide new evidence for the structures in (1) and (2) as well as evidence that inflection in Iroquoian is interleaved with derivational conversion processes that form nouns from verbs. Such conversion processes play an important functional role in Oneida onomasiology, given the verb-centricity of Iroquoian languages noted since at least Cuoq (1866) for Mohawk. What is critical for our purposes is that conversion can target the stem or any bracketed constituent of the structure diagrammed in (1) to derive a constituent in (2).

Table 1 summarizes the properties of noun roots, as well as the classes of derived nouns, and identifies the inflectional potential of underived nouns and each kind of derived nouns as well as whether they are eligible for noun incorporation; the last column gives the number of each kind in Michelson & Doxtator (2002). Examples are provided in (3)-(6).

(3)   **o**-sahe?t-**a?**
NPF-bean-NSF
'bean(s)'

(4)   **lao**-hyatú-hsl-**i?**
3M.SG.POSS-write-NMZR-NSF
'his book, paper'

(5)   **akw**-an-isnuhs-ohlók-t-**a?**
1SG.POSS-SRF-finger-insert-CAUS-HAB
'my ring'

(6)   **yu**-t-wʌn-a-ta?-á-st-**a?**
3FL.A-SRF-voice-JN-put.in-JN-CAUS-HAB
'telephone'

	Nominal prefix (NPF)	Possessive prefix (POSS)	Nominal suffix (NSF)	Incorporation	# in Michelson & Doxtator (2002)
Noun root	+	+	+	Possible	305
Nouns derived from a verb stem	+	+	+	Possible	24
Nouns derived from a verb stem + aspect suffixes	+	+	-	Impossible	55
Nouns derived from fully inflected verbs	-	-	-	Impossible	333

Table 1: Classes of underived and derived nouns in Oneida, which inflectional processes they can undergo and whether they can incorporate into verbs or not.

Crucially, as Table 1 indicates, which constituent is targeted by the conversion process determines the inflectional potential of the derived noun. If an uninflected stem is the input, as in (4), the nominal output will have both nominal suffixes, nominal pronominal prefixes, and will be able to incorporate into a verb root, just as uninflected, underived noun roots, as in (3). If a stem plus an aspect suffix is the input, as in (5), only nominal pronominal prefixes are possible, since the stem already includes (verbal) aspect suffixes and only uninflected noun stems can incorporate. Finally, if a fully inflected word is the input, as in (6), the output has no nominal inflectional potential but includes all the verbal inflection included in the input.

Interestingly, about a dozen nouns in Michelson and Doxtator (2002) that are derived from verbs have two variants, one variant is derived from a verb stem that includes an aspect suffix, and the other variant is derived from a verb stem that includes a nominalizer suffix. The first variant behaves like the example in (5) whereas the other variant *must* incorporate. An example is given in (7): The forms in (7a) and (7b) include the habitual aspect suffix *-ha?* (*-hkw-ha?* becomes *-khwa?* by phonological rule), whereas the form in (7c) has the nominalizer suffix *-?tsl-*. Only this last form can (in fact, must) incorporate. The data summarized in Table 1 and (3)-(7) show that Oneida conversion is sensitive to and interleaved with the layered inflectional structure of verbs and nouns.

(7)   a.   **ate**-khw-a-hl-á-**khwaʔ**
           ∅,NPF-SRF-food-JN-set.on-JN-INSTR:HAB
           'table'

      b.   **akw**-ate-khw-a-hl-á-**khwaʔ**
           1SG.POSS-SRF-food-JN-set.on-JN-INSTR:HAB
           'my table'

      c.   waʔ-k-ate-khw-a-hl-a-**ʔtsl**-oˑkéw-eʔ
           FACT-1SG.A-SRF-food-JN-set.on-JN-NMZR-wipe-PNC
           'I wiped the table'

Although we have talked until now of Oneida nouns and verbs, the data on conversion supports an ontological reanalysis of those terms: *noun* stem or word reduces to stem or word describing entities or objects and *verb* stem or word reduces to stem or word describing events or situations. This is because conversion processes are ontologically transparent in Oneida: the change in inflectional potential is *always* accompanied by a change in ontological category

(from situation/event describing stem or word to object/entity describing stem or word). This contrast with ontologically opaque conversion processes in Indo-European and many other languages, where the input and output can share ontological type. For example, the noun *dance* in English can still describe the unfolding of an event through time, despite its nominal morphological and syntactic properties. The ontological transparency of conversion in Oneida supports the view that the Oneida lexicon (excluding particles) is organized along ontological classes (object vs. event describing stems or words), as inflectional potential strictly follows from the ontological class a stem belongs to: given an inflecting stem or inflected word's ontological category, its inflectional potential follows.

Finally, we show that a semantic organization of Oneida inflected lexemes is superimposed on its ontological organization. Whereas the ontological organization separates lexical entries into event/situation and object/entity denoting stems or words, the semantic organization separates inflecting stems or words on the basis of its arity, namely whether it denotes a one-place or two place predicate. Kinship terms and stems/words denoting possession relations exemplify the distinction.

Kinship relations describe entities in Oneida, as evidenced by the fact that pronominal prefixes that reference their arguments drop initial glides when relevant, just like stems that describe entities (see Table 1). But they select so-called transitive prefix like verbs, i.e. prefixes that reference both arguments of the relation as shown in (8). More generally, kinship stems share with situation-describing stems all inflectional properties characteristic of two-place predicates and with entity-denoting stems all inflectional properties characteristic of stems whose discourse referent are of sort entity rather than situation (see Koenig & Michelson (2010) for details).

(8)   lakeʔníha   aksótha   onatatyʌ́ha
        lake-ʔni-ha   **(w)ak**-hsot-ha   **(y)on**-atat-yʌ-ha
        3M.SG>1SG-father-DIM   3FZ.SG>1SG-grandmother-DIM   3FZ.DP.P-REFL-child-DIM
        'my father'   'my grandmother'   'mother and daughter'

Stems denoting possession relations are an exception to the previous generalization as only possessors are marked, as shown in (9). Importantly, this exceptional selection of intransitive prefixes despite the fact that their semantic content is relational stays constant whether that content is used to describe a member of that relation, an entity, (9), or the relation itself, a situation (10).

(9)   **la**-ʔnyú·ke
        3M.SG.A-nose-LOC
        'his nose'

(10)   Thoʔnʌ́   tékni   te-**wak**-atʌnoʔsʌ-·sh-ʌ́·
        and then two   DL-1SG.P-sibling-NMZR-have:STV
        'And then I have two brothers'

Although their differ in argument referencing, stems whose semantic content include a possession relation or a kinship relation behave similarly in one respect, how they are negated. In general, stems describing situations and entities differ in how they are negated in Oneida. Both require the presence of the uninflected particle *yah*, but whereas stems describing situations must also include the prepronominal negative (inflectional) prefix *teʔ-*, stems describing entities include the uninflected particle *té·kʌ*. Kinship terms and stems that describe entities that are possessed, though, behave differently; we illustrate with possessed stems. Possessed stems can follow both patterns, that is they can negate just like other stems describing objects/entities, as in (11) or they can negate like other stems whose semantic content is relational, as in (12).

(11) Yah né·   í·   ak-kʌ́h-aʔ   **té·kʌ** thi·<u>k</u>ʌ́.
      not assertion FIRST.PERSON 1SG.POSS-blanket-NSF it's not that

      'It's not my blanket.'

(12) Yah né·   í·   **teʔ**-wak-kʌ́h-aʔ   thi·<u>k</u>ʌ́.
      not assertion FIRST.PERSON NEG-1SG.POSS-blanket-NSF that

      'It's not my purse/blanket.'

The behavior of kinship terms and stems which contribute a possession relation in the sentence's semantic content shows that Oneida inflection is not only sensitive to structural considerations and based on an ontological distinction between what is being described, entities/objects or events/situations, but also to the kind of semantic type of the content used to describe those individuals, namely one place or two-place predicates (and in the case of two-place predicates, whether possession is involved or not). In brief, to properly infect a stem in Iroquoian, you need to know: what its structure is (as in (1)-(2)), the sort of its discourse referent (see de Swart 1998), and the type of its semantic content (simplifying somewhat, $< e, t >$ vs. $< e, e, t >$).

References

Bonvillain, Nancy. 1973. *A grammar of Akwesasne Mohawk (Canadian ethnology service paper no.8)*. Ottawa: National Museum of Man Mercury Series, National Museums of Canada.

Chafe, Wallace. 1967. *Seneca morphology and dictionary*. Washington: Smithsonian Press.

Cuoq, Jean-André. 1866. *Etudes philologiques sur quelques langues sauvages de l'Amérique.* Montréal: Dawson Brothers, reprinted 1966 by Mouton, La Hague.

de Swart, Henriëtte. 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory* 16(2). 347–385.

Koenig, Jean-Pierre & Karin Michelson. 2010. Argument structure of Oneida kinship terms. *International Journal of American Linguistics* 76. 169–205.

Lounsbury, Floyd. 1953. *Oneida verb morphology. Yale University Publications in Anthropology 48*. New Haven, CT: Yale University Press.

Michelson, Karin & Mercy Doxtator. 2002. *Oneida-English/English-Oneida dictionary.* Toronto: University of Toronto Press.

Description of verbal morphology of Asama: a realizational and implemented approach

Dimitri Lévêque   *Thomas Pellard*

INalCO - CNRS - EHESS, CRLAO   CNRS - INalCO - EHESS, CRLAO

In this presentation, we will give an overview of the verbal morphology of Asama, an endangered Japonic language, which includes non-concatenative phenomena implying mostly tone and vowel length. We will introduce the theoretical and practical framework we have used for this study, explain why we chose it and how we used it in order to analyze the verbal morphology of Asama, and conclude by showing how the programs implemented can be used to obtain more quantitative results by computing paradigm cells interpredictability and Shannon entropy.

1   Theoretical and practical framework

Asama is a Japonic language spoken in the village of the same name located in the island of Tokunoshima, Japan. While Japanese and Japonic languages are usually known for their relatively simple agglutinative verbal morphology, it is not the case for all of them, and some like Dunan (Pellard & Yamada, 2017) or Asama (Lévêque, 2017) show a more complex morphology. Asama's verbal morphology indeed displays several phenomena of stem alternation (see Table 1 for an illustration) and suffix allomorphy that are not found elsewhere in Japonic languages, and that make it far less easy to describe with a classical morpheme-based framework (Selkirk, 1982; Halle & Marantz, 1993). All the data used in this study (fifty complete and several hundred partial verbal paradigms) are taken from Okamura et al. (2009), Uwano (2001) and one of the author's own data collected during fieldwork in 2018 and 2019.

Table 1: Stem alternation in Asama

Form	TUBJUI "to fly"	KOORAKJUI "to dry"	MʔAARɨJUI "to be born"
NPST	tubj- (H)	koorakj- (H)	mʔaarɨj- (H)
PROH	tubj- (H)	koorakj- (H)	mʔaarɨj- (H) / mʔaarɨ- (H)
NEG	tub- (H)	koorak- (H)	mʔaarɨr- (H)
DES	tub- (H)	koorak- (H)	mʔaarɨ- (H)
PST	tud- (H)	kooracj- (H)	mʔaarɨt- (H)
SEQ	tud- (H)	kooraacj- (H)	mʔaarɨɨt- (H)
PROG	tud- (LH)	kooracj- (H) / kooraacj	mʔaarɨt- (H) / mʔaarɨɨt- (H)

The verbal morphology was fully implemented with finite-state transducers (Beesley & Karttunen, 2003) and with the foma software (Hulden, 1999).

Asama's verbal morphology has many features that make it far from being canonical, in the sense of Corbett (2009). Two of those features are listed above, and many of those phenomena involve tone and vowel length, two aspects that deserve a special attention in Asama verbal morphology.

• Overabundance: one case of a given paradigm has more than one surface form (ex. 1).

(1)

Meaning	{PROG PST} Var. 1	{PROG PST} Var. 2
"to begin"	hazɨmiitutaN (H)	hazɨmituuutaN (H)
"to be born"	maarɨitutaN (H)	maarɨtuuutaN (H)

- Multi-functional exponent: a single form can express more than one morphosyntactic property (ex. 2).

(2)

Meaning	{IMP.INDIR}	{SEQ}
"to read"	juum-ɨ (LH)	juud-ɨ (LH)
"to meet"	oor-ɨ (H)	oot-ɨ (H)

That is why for a great majority of inflectional forms, it is not possible to use a morpheme based analysis, since in many cases a given morpheme cannot be related to a fixed morphosyntactic property or set of properties. On the contrary, a Word-and-Paradigm approach (Hockett, 1954) such as Paradigm Function Morphology (PFM), a framework that has been first developed by Stump (2001), is more promising to account for inflectional morphology of Asama. In this study, we thus follow the six core assumptions posited by Bonami (2014) and Bonami & Stump (2016) to describe PFM. In particular, we base on the assumption of the definition of a language inflectional morphology as a paradigm function in order to analyze and compute verbal morphology.

2   Analyses and results

In order to describe verbal morphology in Asama, we use two different sets of inflectional classes. The first set is meant to take into account the segmental stem alternation alone, as it can be seen in the Table 2 below, that shows a partial description of verbal morphology. In this table, X, Y and Z are purely morphomic categories (Aronoff, 1994), X and Y being subdivised in sub-categories for only a few inflectional classes.

Table 2: First set of inflectional classes

Inflectional class	X_1	X_2	X_3	Y_1	Y_2	Z
m ~ d	mj	*NA*	*NA*	m	m	d
n ~ n	nj	*NA*	*NA*	n	nj	zj
k ~ c	kj	*NA*	*NA*	k	k	cj
k ~ z	kj	*NA*	*NA*	k	k	zj
s ~ s	sj	*NA*	*NA*	s	sj	sj
t ~ cc	cj	*NA*	*NA*	t	t	ccj
i ~ r ~ t	j	-	-	r	r	t
{a,o,u} ~ r ~ tt	j	-	-	-	r	tt
{s,z}ɨ ~ r ~ cj	j	*NA*	-	r	r	cj
{a,o,u} ~ r ~ cj	j	-	-	-	r	cj

The second set of inflectional classes accounts for the autosegmental part of stem alternation, implying mainly tone and vowel length. Table 3 shows how it is possible to represent those two classes with a single set of rules ("R" means "Root") and can be read as an explicitation of the realisational rules that lays at the core of the Paradigm Function Morphology.

Table 3: Second set of inflectional classes

Classe	{IMP.DIR}	{SEQ}	{PROG NEG}	{PROG NEG NPST}
I$_1$	R:Y$_2$ɨ (H)	R:Zɨ (H)	RZui (H)	RZuuraN (H)
I$_2$	R:Y$_2$ɨ (H)	R:Zɨ (H)	R:Zui (H) / RZui (H)	R:ZuraN (H) / RZuuraN (H)
II$_1$	RY$_2$ɨɨ (H)	RZɨɨ (H)	RZui (LHL)	RZuuraN (LHL)
II$_2$	R:Y$_2$ɨ (H)	R:Zɨ (H)	RZui (LHL)	RZuuraN (LHL)
III	R:Y$_2$ɨ (HL)	R:Zɨ (HL)	R:Zui (HL)	R:ZuraN (HL)
IV	R:Y$_2$ɨ (LH)	R:Zɨ (LH)	R:Zui (LHL)	R:ZuraN (LHL)

The verbal lexeme AMJUI "to knit", which is associated to the inflectional classes "m~d" and "III", can be used to illustrate this approach. To obtain the form of this lexeme, associated for instance with the following set of morphosyntactic properties {prog neg npst}, the following realizational rules apply:

- Rules of Stem Choice: RSC$_{md,III}$(⟨AMJUI,σ\{prog neg npst\}⟩) = ⟨aad,σ⟩

- Rules of Exponence:

  - Block 1: RE1$_{md,III}$(⟨aad,σ\{prog neg npst\}⟩) = ⟨aadu,σ⟩
  - Block 2: RE2$_{md,III}$(⟨aadu,σ\{prog neg npst\}⟩) = ⟨aaduran,σ⟩
  - Block 3: RE3$_{md,III}$(⟨aaduran,σ\{prog neg npst\}⟩) = ⟨aaduran (HL),σ⟩

Those rules correspond to what is found in the cell in grey color in Table 3. The implementation of all the transducers describing the complete verbal morphology are based on those two tables.

In the presentation, we first show with the help of detailed examples that the model outlined above is both an elegant and efficient way to describe the verbal morphology of the Asama language. It is moreover in line with the primary goal of this study, that is to say the description of an under-documented and undescribed language such as Asama, especially since it helps automating the gloss.

In the second part of the presentation, we present some quantitative results obtained from the transducers. Starting from the notion of paradigmatic structure, implicative morphology (Wurzel, 1989; Albright, 2002; Bonami, 2014) and interpredictability between the cells of paradigms, we have computed, as it has been done for example in Bonami & Luis (2015) or Pellard & Yamada (2017), interpredictability and conditional entropy calculations (Shannon, 1948; Blevins, 2013; Ackerman et al., 2009; Ackerman & Malouf, 2013), that give further strength to the qualitative explanations. All the calculations are made on the base of a distillation of the paradigm (Stump & Finkel, 2013, 42), that results in a sub-paradigm of five cells, each of them representing one area of full interpredictability of the full paradigm.

Entropy results are first used to highlight the main sources of uncertainty, and thus confirm what has been found in the qualitative analysis. Uncertainty mainly lies in neutralisations of oppositions based on tone and vowel length, and also in unpredictable segmental alternations. We also take advantage of the programs to attribute a weight to the uncertainty that can be attributed to segmental alternations and the uncertainty that can be attributed to suprasegmental alternations.

Finally, entropy calculations based on binary implications show that the best candidates to the status of principal parts of the verbal morphology (Stump & Finkel, 2013) are the nonpast and progressive forms.

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins & J. Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–81. Oxford: Oxford University Press.

Ackerman, Farrell & Robert Malouf. 2013. Morphological Organization: The Low Conditional Entropy Conjecture. *Language* 89. 429–464.

Albright, Adam C. 2002. *The Identification of Bases in Morphological Paradigms*. Los Angeles: University of California dissertation.

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

Beesley, Kenneth R. & Lauri Karttunen. 2003. *Finite State Morphology*. Stanford, CA: CSLI Publications.

Blevins, James P. 2013. The information-theoretic turn. *Psihologija* 46(3). 355–375.

Bonami, Olivier. 2014. *The fine structure of inflection paradigms*. Paris: Université Paris 7 - Denis Diderot. Habilitation à diriger des recherches.

Bonami, Olivier & Ana R. Luis. 2015. Sur la morphologie implicative dans la conjugaison du portugais : une étude quantitative. In *Mémoires de la société de linguistique de paris*, vol. 22, 111–151. Paris.

Bonami, Olivier & Gregory T. Stump. 2016. *Cambridge Handbook of Morphology* chap. Paradigm Function Morphology, 449–481. Cambridge: Cambridge University Press.

Corbett, Greville G. 2009. Canonical inflectional classes. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, Somerville: Cascadilla Proceedings.

Halle, Morris & Alex Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–231.

Hulden, Mans. 1999. Foma: a finite-state compiler and library. *Association for Computational Linguistics* 29–32.

Lévêque, Dimitri. 2017. *Description implémentée de la phonologie et de la morphologie verbale du dialecte des langues ryukyu parlé à Asama, Tokunoshima*. Paris: Paris III Sorbonne-Nouvelle / INaLCO M.A. Thesis.

Okamura, Takahiro, Motoei Sawaki, Yumi Nakajima, Chitsuko Fukushima & Satoru Kikuchi. 2009. *Tokunoshima hōgen nisen bun jiten*. Matsumoto: Shinshū daigaku Jinbun gakubu.

Pellard, Thomas & Masahiro Yamada. 2017. Verb Morphology and Conjugation Classes in Dunan (Yonaguni). In Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Perspectives on Morphological Organization*, 31–49. Brill. doi:10.1163/9789004342934_004.

Selkirk, Elizabeth. 1982. *The syntax of words*. Cambridge, MA: MIT Press.

Shannon, Claude E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27. 379–423 and 623–656.

Stump, G. & R. Finkel. 2013. *Morphological typology*. Cambridge: Cambridge University Press.

Stump, Gregory. 2001. *Inflectional morphology: a theory of paradigm structure*. Cambridge: Cambridge University Press.

Uwano, Zendō. 2001. *Tokunoshima asama hōgen no katsuyōkei akusento shiryō*. Hosei University.

Wurzel, Wolfgang Ulrich. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer.

When *sarkozysation* leads to the *hollandade,*
or the rejection of phonological well-formedness constraints by anthroponym-based derived words

Stéphanie Lignon[1], Fiammetta Namer[1], Nabil Hathout[2], Mathilde Huguin[1]

[1]UMR ATILF, [2]UMR CLLE-ERSS

The main issue addressed in this talk is the weakness of the classical well-formedness constraints such as OCP in wordforms coined on names of political personalities (hence PPNs) such as *Nicolas Sarkozy → sarkoz(y/i)ser* /sarkozize/ 'sarkozy-ize' where the /ziz/ sequence clearly violate OCP (*Obligatory Contour Principle,* Goldsmith (1976), McCarthy (1986), Yip (1988)). Similarly, size constraints seem to be weak in these derivatives such as *Najat Vallaud-Belkacem → vallaud-belkacemisation* `vallaud-belkacemization' where the stem is 5 syllables long and far exceeds the 2 syllables ideal size of the stem in French derivatives (Plénat, 2009b). These examples show that well-formedness constraints are overweighted for PPN-based derivatives by other constraints that are weaker in the general lexicon, one of them being the informativeness constraint (IC) that is satisfied by forms that allow the identification of the base. In this talk we argue that PPN-based derivatives are an ideal data-set for the observation of these constraints and many others. They also highlight the diversity of the constraints that compete and gang up. They not only concern the phonological form and the referential capacity of the derivatives, but also their morphological, lexical and discursive properties.

1   Data, methods

Our corpus is made up of derivatives coined on anthroponyms that belong to the contemporary French political field. 90 PPNs have been selected. They name people that have occupied a leading function in the 1981-2018 period in France: 25 women and 75 men. PPNs consist of a first name and a last name that identifies the referent of the anthroponym (*Nicolas Sarkozy*). In discourse, a PPN can be referred to by different expressions, some of them having an anaphoric vocation. In the following, we call these expressions *subnames,* because they are often subparts of the PPN. The subnames of a PPN include at least the first name (*Nicolas*), the last name (*Sarkozy*) and the first name + last name (*Nicolas Sarkozy*). When the PPN contains a compound last name, it can have up to 6 subnames, including the two components of the last name and sometimes an acronym (eg. DSK for Dominique Strauss-Kahn) or a nickname. As shown in the table below, all subnames can be used for derivation.

PPN	First name	Last name	First name + last name	Last name's 1st comp.	Last name's 2d comp.	Acronym
Nadine Morano	*Nadinette*	*Moranette*	*Nadinemoranien*			
Najat Vallaud Belkacem	*Najatou*	*vallaud-belkacemisation*	*Najat Vallaud-Belkacemien*	*Vallaudista*	*Belkacemien*	*NVBiste*
Ségolène Royal	*Ségolènerie*	*Royalie*	*ségolèneroyalitude*			
Dominique Strauss Kahn	*Dominiqueur*	*StraussKahnie*		*straussophile*	*Kahnerie*	*DSKphilie*
Christine Largarde	*Christinerie*	*lagardesque*	*Christinade Lagardinière*			
Nicolas Sarkozy		*sarkoziste*	*Nicolas Sarkozius*			

From these 90 PPNs we generated automatically about 130,000 candidate derivatives using a number of French suffixation exponents and only kept the ones that are attested online. The resulting corpus contains 5,000 derivatives and their 55,000 occurrences.

2   Well-formedness constraints

Numerous studies have demonstrated the existence of phonological and lexical well-formedness constraints on word formation. The main phonological constraints in French are OCP and the size constraint that give preference to derivatives with an ideal disyllabic stem (Plénat, 2009b). To reach this optimal shape, several strategies are implemented: stem adjustment (Plénat, 2009a); stem swapping (Dal & Namer, 2010); affix substitution (Koehl & Lignon, 2014; Lignon, 2013; Lignon & Plénat, 2009; Lindsay & Aronoff, 2013; Aronoff, 2016). Lexical (paradigmatic) constraints also apply: they are induced by the existing lexicon and can explain part of the variation observed in the output of various word formations (for French, see Roché 2011, Hathout 2011). Other constraints are more semantic. For instance, the informativeness constraint favors derivatives with a form that allow an optimal identification of the referent of the base. In the lexicon, the different constraints gang up and compete with each other (McCarthy & Prince, 1993; Prince & Smolensky, 1993). Individual trade-offs between them explain the observed lexical variations (Roché & Plénat 2014).

For instance, the interplay between the phonological and informativeness constraints can be observed in the verb in -iser derived from *Nicolas Sarkozy*. The verb could have been coined as *nicolasifier* 'nicolas-ify', which is phonologically better than *sarkoz(y/i)ser*. But because it is not informative enough this form is not attested online. When a speaker coins a PPN-based neologism, IC makes a form more likely to denote its referent unambiguously, even at the expense of the well-formedness constraints.

3   Derivation from PPNs: main cases

When it comes to anthroponyms, phonological constraints are weaker in the construction of their derivatives, with the exceptions exposed in §4. Examples (1) show that PPN-based wordforms are insensitive to dissimilative constraints.

(1) /sarkozize/
    Si elle pouvait se "__sarkoziser__" jusqu'à être élue, ça m'irait très bien. [Nicolas Sarkozy]
    'If she could "__sarkoz-ize__" herself until being elected, that would suit me very well.'
    /valsɛsk/
    La formule __Vallsesque__ a fini par percer [Manuel Valls]
            'The __Valls-esque__ formula finally broke through'

The table below confirms the tendency. In our corpus, the most frequent subname that surfaces in the derivatives is the last name (in more than 80% of them) and subnames containing the last name make up almost 90% of the stems. This is a direct consequence of IC since last names are the most informative subnames.

First name	Last name	First name + last name	Component 1	Component 2	Acronym
5,68%	82,45%	5,50%	2,49%	2,49%	1,37%

4   Derivation from PPNs: minority cases

Almost 18% of derivatives on PPNs are not coined on the last name. They result from a combination of conditions that include:

(a) **prosodic context**: when the last name is monosyllabic, the preferred stem is first name + last name. For example, nearly 58% of the words derived from the PPN *Rama Yade* use the stem /ramajad/ and only 16% use /jad/ alone. Here the IC is gang up with the size constraint.

(b) **discursive context**: when the PPN is mentioned in the text just before the derivative, the use of a subname different from the last name has an anaphorical function:

(2) *Faites pas votre **Sarkozy** (j'ai souvenir d'un coup de colère **nicolien** face à la question d'un jeune sur un plateau de tv, pendant une campagne*
'Don't be such a **Sarkozy** (I remember a burst of **Nicolas-ian** anger at a question asked by a young one on a TV set, during a campaign)'
*Le monde, en ce soir bien sombre, a grand besoin des paroles lumineuses d'**Harlem Désir**. Où es-tu, trou du heuh, **Harlemou** ?*
'The world, in this very dark evening, is in a great need of **Harlem Désir**'s illuminating words. Where are you, assho… hem… **little Harlem**?'

(c) **sociolinguistic context**: when the PPN denotes a woman, the first name is favoured especially when it is uncommon (eg. *Ségolène, Najat, Arlette, Roselyne, Rama, Rachida*): more than 80% of the words formed on a first name have a woman PPN base. This does not contradict IC, but weakens it when the last name is more informative than the first name such as in the case of *Bachelot* compared to *Roselyne*. On the other hand, *Royal*, which can be confused with the homonymous relational adjective, is less informative than *Ségolène*. In this case, the last name is used less frequently as a stem than both the first name and the first name + last name even if the latter violate the size constraint. This choice is sociologically marked, because for masculine PPNs the first name is used much less frequently even when it is rare and more informative than the last name. It actually echoes the way women politicians are referred to in texts.

(d) **evaluative context:** when the derivative is a hypocoristic (eg. in *-ette* or *-ou*), the preferred subname ends in /ɛ̃/ or /in/. The attraction between these stems and suffixes results from the paradigmatic pressure of the existing lexicon (Plénat, 2005; Plénat & Roché, 2004), where /in/ is the most frequent sequence that appears before these suffixes. The selection of the /ɛ̃/ ending favors last names, as expected (*Boutin*, /butɛ̃/, *Jospin*, /ʒospɛ̃/, *Autain* /otɛ̃/), but also some frequent masculine first names (*Alain* /alɛ̃/) as in (3a) (note that in contact with *–ette* or *–ou*, /ɛ̃/ is denasalized in /in/). More interestingly, the selection of /in/ endings favors the female first names, whether rare (3b, c) or frequent (3d):

(3) a. *Car, **Alinou** chéri, regardons les choses en face : tu as cent mille fois raison* [Alain Juppé]
'Because, **Alain-ou** darling, let us face things: you are absolutely right.

 b. *Qui se trouvait classe pour aller en conseil des ministres? C'est vrai qu'elle est tellement distinguée **Nadinette** avec ses perlouses…* [Nadine Morano]
'Who does find herself classy to go to the Council of Ministers? It's true that she is so distinguished, **Nadine-ette** with her fake pearls'

 c. *Pour une fois qu'une **marinnette** se presente ici , ca me fait rigoler* [Marine Le Pen]
'For once that a **marine-ette** presents herself here, that makes me laugh'

 d. *Retrouvez Ségo et **Martinette**, les deux sœurs haineuses* [Martine Aubry]
'Find out Ségo and **Martine-ette**, the two hateful sisters'

5   Conclusion

The data presented here show how constraints gang up and compete when the base of the derivative is a PPN: willingness to inform; gender of the PPN's referent; size of the subname; influence of the context; lexical pressure. Many of the derivatives in our corpus are nonce formations, and are only attested in online writing texts and are characterized by their spontaneity, volatility, willingness to play, etc. (Munat, 2007, Dal & Namer 2018). Our observations, analyses and results raise several questions: Are well-formedness constraints really operational in all contexts? If not, what are their limitations? Do PPNs form a subclass of anthroponyms? If not, what makes them special? Is it the size and frequency of their derivational families? Do they function as derivational bases in the same way as common nouns?

References

Aronoff, M. (2016). Competition and the lexicon. In: A. Elia, C. Iacobini & M. Voghera (eds.): *Atti del XLVII Congresso Internazionale della Società di Linguistica Italiana.* 39-52. Roma: Bulzoni.

Dal G. & F. Namer (2010) French Property Nouns based on Toponyms or Ethnic Adjectives: a case of base variation, F. Rainer, W. U. Dressler, D. Kastovsky & H. C. Luschützky (eds). *Variation and change in Morphology*, Amsterdam: John Benjamins: 53-74.

Dal G. & F. Namer (2018) Playful nonce-formations, creativity and productivity. In S. Arndt-Lappe, A. Braun, C. Moulin & E. Winter-Froemel (eds). *Expanding the Lexicon. Linguistic Innovation, Morphological Productivity, and Ludicity.* 203-228. Berlin: Mouton de Gruyter.

Goldsmith, J. (1976). *Autosegmental Phonology*. PhD Thesis. Cambridge, Mass., MIT.

Hathout, N. (2011). Une analyse unifiée de la préfixation en anti-. In M. Roché (ed) *Des Unités Morphologiques au Lexique.* 251-318. Paris: Hermès.

Koehl A. & S. Lignon. (2014) Property nouns with *-ité* and *-itude*: formal alternation and morphopragmatics or the *sad-itude* of the Aité$_N$. *Morphology* 24(4). 351-376.

Lignon, S. & M. Plénat. (2009) Echangisme suffixal et contraintes phonologiques. In B. Fradin, F. Kerleroux & M. Plénat (eds). *Aperçus de morphologie du français*: 65-81. Paris: PUV.

Lignon S. (2013) *–iser* and *–ifier* suffixations in French: Verify data to verize hypotheses? In N. Hathout, F. Montermini, J. Tseng (eds). *Morphology in Toulouse. Selected proceedings of Décembrettes 7.* 109-132. Munich: Lincom Europa.

Lindsay, M. & M. Aronoff (2013). Natural selection in self-organizing morphological systems, In N. Hathout, F. Montermini & J. Tseng (eds). *Morphology in Toulouse: Selected Proceedings of Décembrettes 7.* 133-153. Munich: Lincom Europa.

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17(2): 207-263.

McCarthy, J. J. & A. Prince (1993) *Prosodic Morphology: Constraint Interaction and Satisfaction.* Amherst, MA: Linguistics Department Faculty Publication Series.

Munat, J. (2007). *Lexical Creativity, Texts and Contexts*. Amsterdam:John Benjamins.

Plénat, M. (2005). Rosinette, cousinette, putinette, starlinette. Décalage, infixation et épenthèse devant *–ette*. In I. Choi-Jonin, M. Bras, A. Dagnac & M. Rouquier (eds). *Questions de classification en linguistique : méthodes et descriptions. Mélanges offerts au Professeur Christian Molinier.* 275-298. Bern: Peter Lang.

Plénat, M. (2009a). Le conditionnement de l'allomorphie radicale en français. *Mémoires de la Société de Linguistique, Nouvelle série* XVII. 119-140.

Plénat, M. (2009). Les contraintes de taille. In B. Fradin, F. Kerleroux & M. Plénat (eds). *Aperçus de Morphologie du français.* 47-64. Paris: Presses Universitaires de Vincennes.

Plénat, M. & M. Roché (2004). Entre morphologie et phonologie : la suffixation décalée. *Lexique* 16. 159-198.

Roché, M. (2011). Pression lexicale et contraintes phonologiques dans la dérivation en -aie du français. *Linguistica* 51. 5-22.

Roché, M. & M. Plénat (2014). Le jeu des contraintes dans la sélection du thème présuffixal. *4e Congrès Mondial de Linguistiaue Francaise*. Berlin. SHS Web of Conferences. 8: 1863-1879.

Prince A. & Smolensky P. (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report **2**.

Yip, M. (1988). The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19(1): 65-100.

Accounting for morphological complexity vs. simplification in situations of language contact: evidence from Cappadocian Greek

Dimitra Melissaropoulou
Aristotle University of Thessaloniki

1 Introduction

The last two decades the notion of complexity has become central to modern linguistic theorizing in terms of providing explanations for the architecture of grammar, language processing, language acquisition, language variation and change. Different perspectives have led to various proposals on the different forms (distinctions) of complexity such as absolute vs. relative complexity (Dahl 2004), local vs. global complexity (Miestamo 2006), I- vs. E-complexity (Di Sciullo 2012), enumerative vs. integrative complexity (Ackerman & Malouf 2013), formal vs. processing complexity (Culicover 2013), system complexity vs. complexity of exponence (Anderson 2015).

The aim of this paper is to offer further insights on the notion of local "autonomous" (cf. Aronoff 1994; Anderson 2015) morphological complexity vs. simplification in the light of the evidence provided by language contact, a parameter which is thought to be interrelated with change in the complexity of linguistic systems (simplification Trudgill 2011 vs. complexification Nichols 1992). For the purposes of this paper we use the notions of complexity vs. simplification a) in order to refer to the formal properties of the examined systems with particular emphasis on regularity, economy, and (formal) markedness relationships or lack thereof (cf. Trudgill 2011 and Siegel 1997 for pidgin and creoles) and b) as theory-internal concepts (i.e. as theoretical notions), although we take something like 'language simplification' to occur in language users' minds.

2 Data

To this end, we draw our data from Cappadocian, an Asia Minor Greek variety spoken for great many centuries in a situation of regressive bilingualism due to intense contact with the agglutinative Turkish language (among others Dawkins 1916; Karatsareas 2011; Janse forthcoming and references therein). Cappadocian is the most highly differentiated dialectal variety of Greek, due to the very long period of isolation from developments involving the rest of the Greek-speaking world, and to the long duration of contact with Turkish. Due to the historical circumstances under which Cappadocian evolved, Cappadocian is characterized by a significant number of linguistic innovations, some of which are generally attributed to the language-contact factor. These innovations brought research on

Cappadocian to the forefront of modern language (contact) studies (among others Dressler & Acson 1985; Thomason & Kaufman 1988; Winford 2005).

For the purposes of this paper, different morphological phenomena are chosen to be exemplified based on their autonomous morphological status as follows:

i. The loss of the originally three different grammatical gender distinctions in nouns:

(1) padiʃax.NEU   'king'
    baldəza.NEU   'sister-in-law'

ii. The tendency towards the establishment of a unique nominal inflectional paradigm with the generalized use of (the neuter) *-ja -ju* markers, substituting the several original subgroups of nouns:

(2) antropos 'man.NOM.SG' atropos-ju 'man.GEN.SG' atropoz-ja 'man.NOM.PL'

iii. The reduction of lexical stem allomorphy, as realized in derivational affixation, where imperfective stems instead of the (usually selected) perfective ones are selected for the formation of deverbal nouns:

(3) ðin(o)   'to give'   ðini$_{\text{IMPERF}}$-ma   'giving'   instead of *dosi$_{\text{PERF}}$-ma
    xan(o)   'to lose   xani$_{\text{IMPERF}}$-ma   'loss'   instead of *xasi$_{\text{PERF}}$-ma

iv. The loss of the morphological process of derivational prefixation. The vast majority of otherwise prefixed verbal forms are realized by loanword elements:

(4) ɣopartiz(u) < koparmak 'to break off, to tear off'   instead of   kse-kolo
    daɣuldiz(u) < dağılmak 'to scatter, to disperse'   instead of   ðia-lio

v. The loss of rivalry among competing derivational suffixes (e.g. *-ma, -simo, -si, -ja*) for the formation of deverbal nouns with the generalization of one productive suffix (i.e. *-ma*):

(5) ðin(o)   'to give'   ðini-ma 'giving'   instead of do-sim(o)
    çoru   'to see, to attend to' çori-ma 'attention' instead of çori-si
    vriz(o) 'to insult'   vrizi-ma 'insult'   instead of vris-ja

3 Proposal

All the observed phenomena, which constitute arbitrary aspects of morphological structure, are seen as instances of loss of morphological complexity, arguing in favor of language simplification in situations of intense language contact (contra Nichols 1992: 193). Nevertheless, while, in a broader perspective, all phenomena lead to a simpler morphological organization, either in terms of system complexity (loss of rivalry among derivational affixes, loss of prefixation as a derivational process, loss of gender) or in terms of complexity of exponence (elimination of allomorphy, loss of multiple inflection classes), following Anderson's (2015) categorization[1], they cannot not be treated adequately under a unified account.

We propose that some of them should be attributed to the direct influence, thus the direct reflex of complexity or simplicity of the model language, in terms of grammatical pattern replication, (loss of gender, the unique inflectional paradigm and loss of prefixation), while others (loss of rivalry among suffixes and elimination of allomorphy) as the result of the pressure that was exerted on the replica language by the dominant system into regression, paving the way into simplification i.e. regularity, economy and loss of redundancy, through minimization of rivalry among elements and categories with similar function.   We propose that the former notions would account for *language-dependent simplification phenomena* in terms of grammatical pattern shift, while the latter for *independent complexity vs. simplification phenomena*, acknowledging, however, that in some cases they may well join forces in the same direction of change.

Generalizing, our data seem to suggest that that in accounting for morphological complexity and the phenomena that are prone or resistant to loss in language contact situations, we should always take into account cross-linguistic divergence, the properties of the specific language set (model vs. replica language) and the compatibility or incompatibility parameter, as well as the notions of regularity, economy and markedness.

References

Ackerman, Farell & Richard Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language,* 89:429-464

---

[1] One should note, that the proposed taxonomy (between system complexity and complexity of exponence), does not appear to be able to make predictions on the amenability of the one over the other subgroup to change or on the conditions under which these phenomena are more or less easily susceptible to it.

Anderson, Stephen. R. 2015. Dimensions of Morphological Complexity. In M. Baerman, D. Brown & G. Corbett (eds.), *Understanding and measuring morphological complexity,* 11-28. Oxford: Oxford University Press.

Aronoff, Mark. 1994. *Morphology by Itself. Stems and Inflectional Classes.* Cambridge (MA): The MIT Press.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity.* Amsterdam: John Benjamins Publishing Company.

Di Sciullo, Anna Maria. 2012. Perspectives on morphological complexity. In Ferenc Kiefer, Mária Ladányi & Péter Siptár (eds.), C*urrent Issues in Morphological Theory: (Ir)regularity, analogy and frequency. Selected papers from the 14th International Morphology Meeting, Budapest, 13–16 May 2010,* 105-134. Amsterdam: John Benjamins Publishing Company.

Dressler, Wolfgang & Veneeta Acson. 1985. On the diachrony of subtractive operations: evidence for semiotically based models of Natural Phonology and Natural Morphology from Northern and Anatolian Greek Dialects. In Jasec Fisiak (ed.), *Papers from the 6th International Congress of Historical Linguistics,* 105-127. Amsterdam, Benjamins.

Culicover, Peter. 2013. *Grammar and complexity. Language at the Intersection of Competence and Performance.* Oxford: Oxford University Press.

Janse, Mark. forthcoming. Cappadocian. In Christos Tzitzilis, (ed.), *Η ελληνική γλώσσα και οι διάλεκτοί της* [Greek language and its dialects]. Thessaloniki: Institouto Neoellinikon Spoudon (Manolis Triantafyllides Foundation).

Karatsareas, Petros. 2011. A study of Cappadocian Greek nominal morphology from a diachronic and dialectological perspective. PHD Dissertation. University of Cambridge.

McWhorter, John. 2007. *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars.* Oxford: Oxford University Press.

Miestamo, Matti. 2006. On the feasibility of complexity metrics. *Finest linguistics* (Tallinna Ulikooli eesti filoloogia osakonna toimetised) 8, 1-26. Tallinn: Tallinn University Press.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time.* Chicago, London: The University of Chicago Press.

Siegel, Jeff. 1997. Mixing, leveling, and pidgin/creole development. In Arthur K. Spears & Donald Winford (eds.), *The structure and status of pidgins and creoles,* 111–150, Amsterdam: Benjamins.

Thomason, Sarah. G. & Terrence. Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics.* Berkeley: University of California Press.

Trudgill, Peter. 2011. *Sociolinguistic Typology. Social Determinants of Linguistic Complexity.* Oxford: Oxford University Press.

Winford, Donald. 2005. Contact-induced changes: Classification and processes, *Diachronica* 22, 373–427.

Playing with nonwords: morphological skills in dyslexia

Chiara Melloni and Maria Vender

Università di Verona

1 Background

While phonological and morphological abilities are both crucial for the acquisition of reading, morphology has been found to play a pivotal role especially at later stages of literacy acquisition, as established by several studies finding correlations between morphological skills and reading achievements in alphabetic orthographies (Brittain 1970; Tornéus 1987; Carlisle 1995). On the other side, research on reading impaired populations showed that their morphological abilities are severely compromised (Elbro & Arnbak 1996). In particular, deficits have been found across several tasks assessing the abilities to isolate and blend morphemes (Casalis et al. 2004), in gender and number agreement (Jiménez et al. 2004; Rispens 2004) and in other domains of inflectional morphology (Joanisse et al. 2000; Vender et al. 2017).

2 Research questions

In view of the above, the present study aimed to address the following research questions:

1. Is there (and what is the extent of) the dyslexic disadvantage across domains of Italian inflectional and derivational morphology?
2. Which are the most problematic domains for Italian dyslexic children?
3. Are morphological skills able to predict reading proficiency?

3 Method

A protocol comprising morphological tasks and preliminary measures was administered to 16 Italian dyslexic children (DC; 10;2 years old, SD = 1.15) and 18 typically developing children (CC; 10;6 y.o., SD = 0.88).

All children were tested along the following preliminary measures: non-verbal intelligence (*CPM Raven*, Raven, Court & Raven, 1998), receptive vocabulary (*Peabody Picture Vocabulary Test* by Dunn and Dunn 2000, Italian standardization by Stella, Pizzoli and Tressoldi 2000), word and nonword reading accuracy and speed (Tasks 2. and 3. of the *DDE-2* by Sartori, Job, & Tressoldi 2007). Eleven morphological tasks elicited the production of a derived or inflected form of a nonword (as in Berko's 1958 original Wug Test) or the retrieval of the base of a morphologically complex nonword.

The adoption of a test with nonwords is crucial for understanding whether children are able to capture and correctly apply the relevant word formation rules/patterns to possible, yet nonexistent, words. Besides pseudo-noun pluralization, the focus of most tasks was on verb or verb-based formation, as with past participles, deverbal adjectives and nominalizations (Table 1). Conditions in each task manipulated base allomorphy in compliance with the base verb conjugation class for the verb related tasks (all tasks except 1. and 7.), while they manipulated declension classes for noun pluralization in task 1. (as in Vender et al. 2017 and Melloni et al. in press) and type of evaluative affix in task 7.

Table 1. Summary of the morphological tasks and type of ability tested

Task 1.	singular N > plural N	INFLECTION
Task 2.	infinitive V > past participle V	
Task 3.	infinitive V > Agent N in *-tore*	DERIVATION
Task 4.	infinitive V > Action N in *-mento*	
Task 5.	infinitive V > Action N in *-ta*	
Task 6.	infinitive V > Adjective in *-bile*	
Task 7.	base N > evaluative N (*-ino, -one, -accio*)	
Task 8.	N in *-tore* > infinitive V	BASE RETRIEVAL
Task 9.	N in *-mento* > infinitive V	
Task 10.	N in *-ta* > infinitive V	
Task 11.	N in *-bile* > infinitive V	

To illustrate stimuli and elicitation procedure, we can consider an inflection task, Past Participle formation (task 2.). This task required the subject to derive the past participle of a nonce verb. The child was presented with a character, Goofy, who performed some invented actions. The elicitation formula was: "Qui si è messo a pindare. Cos'ha fatto?" (target: Ha *pindato*). ('Here he started to *pindare*$_{Inf}$. What has he done? (He has pindato$_{PastPart}$)'). There were 9 items, three for each condition, corresponding to the three Italian conjugations:

1. Condition 1: Infinitive *a-re* > -a-to, e.g. pind-are > pind-ato (I conjugation)
2. Condition 2: Infinitive *e-re* > -u-to, e.g. nov-ere > nov-uto (II conjugation)
3. Condition 3: Infinitive *i-re* > -i-to, e.g. call-ire > call-ito (III conjugation)

A typical derivation task, *-bile* adjective formation, requires the subject the subject to derive an adjective from the infinitive form of a nonce verb by adding the suffix *-bile*. The elicitation formula was: "Questa strada si può madare, quindi possiamo dire che è…(target: *madabile*)" ('This street can be madare$_{Inf}$, then we can say that it is…madabile$_{Adj}$'). There were nine items, three for each condition, built in compliance with conjugation classes as in task 2.

A base retrieval task like 8. required the subject to retrieve the infinitive form of the base verb from a nonce noun suffixed with the (agentive) *-tore*. The elicitation formula was: "Al pifatore piace…(target: *pifare*)" ('The *pifatore* likes…pifare$_{Inf}$'). In this case, we had six items, three for each of the two conditions formally corresponding to the I and III verb conjugations, since only derived forms in *-a-tore* and *-i-tore* are allowed in Italian (nominalizations from verb of the second conjugation are formally opaque, as the theme vowel -e- becomes -i- in the derived nominal, as in *miet-i-tore* 'reaper' < *miet-e-re* 'to reap').

As for the scoring system, one point was attributed for each correct item and no points for incorrect ones; no penalizations were given to mispronunciation errors if the target morphological operation was correctly performed (e.g. *pindare* > *pin̲tato*).

4 Results

The results of the study revealed that DC performed significantly more poorly than CC in the morphological tasks ($p < .001$).

To compare their performances in each task, independent sample t-tests were run considering the general accuracy in each task; then, a multivariate analysis of variance

(MANOVA) was conducted, with performance in each condition of every task as dependent variable and Group (DC; CC) as fixed factor. The analysis run on performance in each task revealed that DC underperformed CC in nine out of eleven tasks, especially in all of the inflection tasks and base retrieval tasks and in three out of five derivation tasks. No differences between CC and DC were found in task 4. and task 7., assessing the ability to form deverbal nouns in –*mento* and to add evaluative suffixes to base nouns.

Finally, correlation and simple linear regression analyses were run between the general mean accuracy in all morphological tasks and the preliminary measures. Morphological skills turned out to be a relevant predictor for all reading measures, especially for accuracy (Table 2).

Table 2. Summary of linear regression analyses predicting reading outcomes based on morphological skills

	B	*SE B*	ß	t	*p*
Word reading speed	14.054	4.086	.520	3.439	<.01
Nonword reading speed	6.754	2.916	.379	2.316	<.05
Word reading accuracy	8.597	2.530	.515	3.397	<.01
Nonword reading accuracy	8.394	1.763	.644	4.761	<.001

5 Discussion

The study provided clear answers to the research questions raised in section 2.

As for research question 1: Dyslexia emerges as a deficit severely affecting morphological skills, especially in (but not limited to) those tasks and conditions requiring fine morphological skills.

As for research question 2: Dyslexic children's performance was significantly poorer in inflection tasks, i.e. noun pluralization and past participle formation, and in tasks tapping the ability to retrieve the infinitival form of (invented) deverbal nouns.

As for research question 3: Morphological skills turn out to be a relevant predictor for all reading measures, especially for reading accuracy.

6 Implications

Morphological skills are impaired in dyslexic children and are relevant for predicting reading abilities, as measured by our Wug Test. Therefore, the results of this study could be taken as concrete indications for speech therapists and educators: morphology-based trainings should be further developed and deployed with the aim to improve dyslexics' reading skills (see Arnback & Elbro 2000; Bowers, Kirby & Deacon 2010). In the perspective of an inclusive education, instruction should insist on various aspects of metalinguistic skills, and especially on morphological skills, as a potential remediation strategy for reading deficits.

References

Arnbak, E., & Elbro, C. 2000. The effects of morphological awareness training on the reading and spelling skills of young dyslexics. *Scandinavian Journal of Educational Research*, 44, 229–251.

Berko, J. 1958. The child's learning of English morphology. *Word, 14,* 150–177.

Bowers, P.N., Kirby, J.R., Deacon, S.H. 2010. The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, 2, 144–179.

Brittain, M.M. 1970. Inflectional Performance and Early Reading Achievement. *Reading Research Quarterly*, 6, 34–48.

Carlisle, J.F. 1995. Morphological awareness and early reading achievement. In L. B. Feldman (Ed.), *Morphological aspects of language processing*, 189–209. Hillsdale, NJ: Erlbaum.

Casalis, S., Colé, P., & Sopo, D. 2004. Morphological Awareness in Developmental Dyslexia. *Annals of Dyslexia,* 54, 114–137.

Dunn, L. M., & Dunn, L. M. 2000. Peabody – Test di Vocabolario Recettivo – P.P.V.T.-R. Peabody Picture Vocabulary Test – Revised.

Elbro, C. & Arnbak, E. 1996. The role of morpheme recognition and morphological awareness in dyslexia, *Annals of Dyslexia*, 46, 209–240.

Jiménez, J. E., Garcìa, E., Estévez, A., Dìaz, A., Guzman, R., Hernandez-Valle, I., Rosario, M., Rodrigo, M. and Hernandéz, S. 2004. An evaluation of syntactic-semantic processing in developmental dyslexia. *Electronic Journal of Research in Educational Psychology*, 2, 127-142.

Joanisse, M.F., Manis, F., Keating, P., & Seidenberg, M.S. 2000. Language deficits in dyslexic children: speech perception, phonology and morphology. *Journal of Experimental Child Psychology, 77,* 30–60.

Melloni, C., Vender, M., Delfitto, D. In press. Inflectional morphology: evidence for an advantage of bilingualism in Albanian-Italian and Romanian-Italian bilingual children. In Amber Dudley, Amy Wallington and James Corbet (eds.), *Proceedings of the 14th Generative Approaches to Second Language Acquisition Conference*. Newcastle upon Tyne: Cambridge Scholars.

Raven, J., Court, J. H., & Raven, J. C. 1998. *Raven Manual, Section 1 (General overview) and Section 2 (Coloured Progressive Matrices)*. Oxford UK: Oxford Psychologist Press.

Rispens, J. 2004. *Syntactic and phonological processing in Developmental Dyslexia*. Doctoral dissertation. University of Groningen, Groningen.

Sartori, G., Job, R., & Tressoldi, P. E. 2007. Batteria per la Valutazione della Dislessia e della Disortografia Evolutiva-2. Firenze: Giunti OS Organizzazioni Speciali.

Stella, G., Pizzoli, C. & Tressoldi, P. 2000. PEABODY – Test di vocabolario recettivo –P.P.V.T. – R. Peabody Picture Vocabulary Test – Revised. Torino, Omega Edizioni.

Tornéus, M. 1987. The importance of metaphonological and metamorphological abilities for different phases of reading development. In *Third World Congress of Dyslexia*, Brete.

Vender, M., Mantione, F., Savazzi, S., Delfitto, D., & Melloni, C. 2017. Inflectional morphology and dyslexia: Italian children's performance in a nonword pluralization task. Annals of dyslexia, 67(3), 401–426.

French -age suffixation versus verb to noun conversion: quantitative approaches on surface and underlying properties

Alice Missud, Florence Villoing

Université Paris Nanterre, MoDyCo (UMR 7114)

1 Introduction

There are many rival morphological schemas that derive nouns from verbs using the same kinds of semantic operations. In French, the most extensively studied rivalry is between *-age* and *-ment* suffixations (*battage / battement* 'beating' (of the heart), *décollage / décollement* 'unsticking'). However, another productive rivalry that has remained unexplored is that of *-age* suffixation and verb to noun conversion (henceforth N-*age* and VN, respectively) (for example, the doublets in (1)). This will be the focus of our study.

(1) *accrochage / accroche* 'hook, hanging', *jetage / jet* 'throwing', *pesée / pesage* 'weighing', *rééquilibrage / rééquilibre* 'rebalancing'

In recent years, the work on rivalry has taken the turn of the new approaches in morphology against a representation of morphological relations by means of unidirectional rules and in favor of a surface-to-surface perspective (for example, Burzio 2002, and Hathout 2009, 2011, Roché 2011, Plénat 2011 for work on French) according to which the lexicon itself originates from a set of constraints involving several levels of linguistic analysis. This work led to the recognition that some rival suffixes organize their competition around a certain linguistic specialization.

While studying the productivity of English rival suffixes, Lindsay & Aronoff (2013) identified that a less productive suffix (*-ical*) is able to compete against a more productive one (*-ic*) thanks to its ability to select bases under certain conditions that only apply in its particular case. Thus, if phonological constraints apply for both suffixes and favor the shortest one, the disadvantaged suffix will find a way to remain productive by specializing in a different way, on a phonological, morphological, semantic or pragmatic level. We propose to study the rivalry between V→N conversion and *-age* suffixation from this perspective. Hence, a first point to note is that, compared to *-age* suffixation, V→N conversion does not add any new element to the selected base. Therefore, size constraints (Plénat & Roché 2003; Plénat, 2009) might play a role to a lesser extent on V→N conversion than on *-age* suffixation. In that case, on average, we expect *-age* derived noun bases to be shorter than converted noun bases. On a semantic level, V→N conversion mostly derives action nouns, as well as result, agent, instrument and locative nouns (Tribout, 2010), whereas *-age* suffixation prototypically constructs action and result nouns (Namer, 2009). Therefore, V→N conversion seems less restricted in base selection and deverbal interpretation than *-age* suffixation. In that case, we would expect the latter to become specialized in order to remain productive.

Two quantitative approaches can help reveal some of the constraints that might explain the coexistence of rival schemas. The first one, introduced by Arndt-Lappe (2014), considers analogy as an explanatory mechanism. Using Analogical Modeling (Skousen, 2002) on English *-ity* and *-ness* derivatives, Arndt-Lappe shows that *-ity* suffixation specialized in selecting adjectives that display different phonological patterns from those that select *-ness* on the surface level. The second one, proposed by Bonami & Thuilier (2019), uses logistic

regression to observe the statistical effects of multiple predictors in the selection of a base. In their study, these predictors are underlying properties identified for each base in the case of French *-iser* and *-ifier* rivalry. As these approaches have proven their worth, we propose to reinvest both of them in our study. We hypothesize that the interaction of surface level properties and underlying constraints could help shed light on what motivates the coexistence of rival V→N conversion and *-age* suffixation.

2 Predictors

We mapped several predictors reflecting relevant constraints on multiple linguistic levels in order to use them as variables in our statistical model.

Stem length: We hypothesize that *-age* suffixation will favor short stems because of the ability of V→N conversion not to add any exponent to the selected base.

Lexical aspect of the verb: Although the lexical aspect of the base verb is not necessarily transmitted to the derivative (Haas et al. 2008), *-age* derivatives can be expected to preferably select activity, accomplishment or achievement verbs, since they usually denote action or result nouns. V→N preferences in that matter need further investigation.

Argumental structure: The bases of *-age* derivatives were identified as prototypically agentive, sometimes ergative or unaccusative (Ferret et al. 2010). However, no preference was identified for V→N conversion. A constraint on the argumental structure of the base might apply for *-age* suffixation. As agentivity can be found amongst dynamic verbs that are frequently used in the first person (usually denoting an animated subject) (Lapraye, 2017 with *-age / -ment*), measuring the frequency of the use of the first person could possibly help to detect agentivity for a base verb automatically. Furthermore, following Hathout (2009) and Bonami & Strnadová's (2016) proposals in favor of a paradigmatic analysis of morphological schemas, we hypothesize that the presence of an agent noun in the morphological family of a verb could also be a predictor for agentivity.

Verb group: Lapraye (2017) has shown that *-ment* suffixation is able to select verbs from the 2nd and 3rd group whereas *-age* suffixation almost exclusively selects verbs from the 1st group. Such preferences need to be investigated in our case.

Reflexivity: As observed in the case of *-age / -ment* (Fradin, 2014), reflexivity of the base verb can sometimes be a criterion to distinguish two connotations of a single verb that derives doublets (*emballer* 'to wrap' → *emballage* 'wrapping', *s'emballer* 'to get excited' → *emballement* 'hype'). We aim to test if the same applies for our case, without taking doublets into account.

Surface form: Finally, surface form could also play a role. If analogical mechanisms are operating, phonological clues could favor one or the other suffix if it happens to be recurrent amongst multiple lexemes derived using the same morphological process.

3 Data and methodology

Because we wanted to test highly frequent derivatives and see the strongest constraints emerge, our hypotheses were tested using quantitative methods on a dataset limited to 200 derived nouns consisting of 100 converted nouns (from Tribout, 2010) and 100 *-age* derivatives extracted from frWaC (Baroni et al. 2009, 1.6 billion words). We selected derivatives with a frequency range from 211 000 (most frequent) to 800 (less frequent) as documented in frWaC. Doublets were excluded, except when frequencies were extremely unequal (*chauffage* 'heat' → *chauffe* 'warm up'). The corresponding verbs were paired manually.

Predictors were annotated for each base-derivative pair. Stem length, group and argumental structure of the base verb were annotated manually. The presence or absence of an agent noun in the morphological family was identified manually from Démonette (Hathout & Namer, 2014). We used Nomage (Balvet et al. 2011) and Haas & Marin's annotated lexicon (in preparation) to annotate the lexical aspect of some of the verbs; missing verbs were annotated manually. The frequency of use of the 1st and 3rd persons and the frequency of reflexive forms were extracted automatically from frWaC, dependency-parsed with Mind The Gap (Coavoux, 2017).

4 Experiments and results

4.1 Logistic regression

Binomial logistic regression was used to observe the relation between a dependent variable (V→N conversion / *-age* suffixation) and 6 variables. The 1st / 3rd person ratio was used as a normalized quantitative variable. The other 5 variables are categorical: verb group (1st, 2nd or 3rd), transitivity of the verb (transitive, labile, inaccusative or inergative), lexical aspect of the verb (state, activity, accomplishment or achievement), reflexivity of the verb (more or less than 5 occurrences) and the presence or absence of an agent noun in the morphological family of the derivative.

Our model showed that, compared to V→N conversion, *-age* suffixation significantly favors short stems (1 or 2 syllables, except for *paramétrer* 'to set' that has 3 syllables) and verbs that do not belong to the 3rd group. A slight preference for accomplishment verbs and non-reflexive verbs was also identified for *-age* suffixation. Longer stems are favored by V→N conversion (mostly 2 but up to 4 syllables), and even though it has a preference for verbs of the 1st group, verbs of the 2nd and 3rd group are selected as well (~30% of the data). The area under the curve (or accuracy) of the model attained 87%.

4.2 Analogical Modeling

Analogical Modeling was given each phonemes of a base at the infinitive form (except for the last one that indicates the verbal group) as categorical variables and applied commutation on each of them, in order to match other bases in the dataset and assimilate the base to either V→N conversion or *-age* suffixation. *-age* suffixation obtained an F-score of 66% whereas V→N conversion reached 57%. Predictions were slightly better than the ones presented in a random baseline (F-score of 57% for *-age* suffixation and 45% for V→N conversion). Although some phonemes can assist classification, the analogical gangs did not reveal any regular pattern of association in the surface form. Examining the matching phonemes did not provide insight into the formal properties of verbs that can distinguish V→N conversion and *-age* suffixation, except that verbs selected by *-age* suffixation tend to gather more in analogical niches.

5 Discussion

The underlying properties tested using logistic regression helped in the identification of semantic, morphological and phonological restrictions applying to *-age* suffixation. However, nothing significant came from transitivity, the 1st / 3rd person ratio and the presence of an agent noun in the morphological family of the derivative. The hypothetical correlation between reflexivity and agentivity of a verb needs further investigation. Combining these results with an analysis of the surface properties of the base verbs using Analogical Modeling was unable to reveal any relevant formal specialization for *-age* suffixation. These results can

be explained by the small size of our dataset. As a next step, investigating analogical mechanisms on a semantic level using distributional methods could eventually help reveal the different semantic domains that are invested by converted and -*age* suffixed nouns.

References

Arndt-Lappe, S. (2014). Synchronic and Diachronic Analogy in Suffix Rivalry: The Case of -ity and -ness. English Language and Linguistics, 18(03), 497-548.

Balvet, A., Barque, L., Condette, M.-H., Haas, P., Huyghe, R. et al.. (2011). La ressource Nomage. Confronter les attentes théoriques aux observations du comportement linguistique des nominalisations en corpus. Traitement Automatique des Langues, ATALA, 2011, Ressources Linguistiques Libres, 52 (3), pp.129-152.

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.

Bonami, O. & Strnadová, J. (2016). Derivational paradigms: pushing the analogy. 49th Annual Meeting of the Societas Linguistica Europaea.

Bonami O. & Thuilier J. (2019). A statistical approach to affix rivalry in lexeme formation: French –*iser* and *ifier*. *Word Structure*, pp. 4-41.

Burzio, L. (2002). Surface-to-Surface Morphology: when your Representations turn into Contrasts. In P. Boucher (ed.), Many Morphologies (142-177). Somerville, MA: Cascadilla Press.

Coavoux, M. (2017). Discontinuous constituency parsing of morphologically rich languages. Thèse de doctorat, Université Paris-Diderot.

Ferret, K., Soare, E., & Villoing, F. (2010). Rivalry between French –age and -ée: the role of grammatical aspect in nominalization. In M. Aloni, H. Bastiaanse, T. De Jager, & K. Schultz (eds.), Lecture Notes in Computer Science: Vol. 6042. Logic, Language and Meaning, 17th Amsterdam Colloquium, The Netherlands, December 2009, Revised Selected Papers (pp. 284–295). Berlin: Springer.

Fradin, B. (2014) Deverbal nominalization and the 'Means' interpretation. Paper read at International Morphology Meeting (IMM16), Budapest, May 29-June 1, 2014.

Haas P., Huyges R., Marin R. (2008) « Du verbe au nom: calques et décalcages aspectuels ». In *Actes du Congrès Mondial de Linguistique Française,* pp.2039-2053.

Hathout, N. (2009). Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie. Toulouse : Université de Toulouse le Mirail. Mémoire d'Habilitation à diriger des recherches.

Hathout, N. (2011) « Une approche topologique de la construction des mots : proposition 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,* Hermes-Lavoisier, Paris, pp.251-318.

Hathout, N. & Namer, F. (2014). Démonette, a French derivational morpho-semantic network. Linguistic Issues in Language Technology 11(5): 125-168.

Lapraye, A. (2017). Une approche statistique de la concurrence entre procédés constructionnels : la dérivation en -age et en -ment en français. Mémoire de Master, Université Paris-Diderot.

Lindsay, M. & Aronoff M. (2013). Natural selection in self-organizing morphological systems. In: F. Montermini, G. Boyé, J. Tseng (eds.): Morphology in Toulouse: Selected Proceedings of Décembrettes 7. Munich: Lincom Europa, pp. 133–153.

Plénat M. (2009) « Les contraintes de taille ». In : Fradin B., Kerleroux F. & Plénat M. (éds). *Aperçus de morphologie du français.* Saint-Denis : Presses Universitaires de Vincennes, 47-64.

Plénat M. (2011) « Enquête sur divers effets des contraintes dissimilatives en français ». In : M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique,* Hermes-Lavoisier, Paris, pp.145-190.

Plénat, M., Roché M. (2003) "Prosodic constraints on suffixation in French". In: Booij G., DeCesaris J., Ralli A. & Scalise S. (eds). *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting (Barcelona, September 20-22, 2001).* Barcelona, I.U.L.A.: Universitat Pompeu Fabra, 285-299.

Roché M. (2011) « Quel traitement unifié pour les dérivations en *-isme* et en *-iste* ? ». In : M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique,* Hermes-Lavoisier, Paris, 69-143.

Skousen, R., Lonsdale, D. W. & Dilworth, B. (2002). Parkinson (eds). Analogical modeling. Amsterdam: John Benjamins.

Tribout, D. (2010). Les conversions de nom à verbe et de verbe à nom en français, Thèse de doctorat, Université Paris Diderot-Paris 7.

Multiple event marking in the Seri verbal paradigm

Jérémy Pasquereau
Surrey Morphology Group (U. Surrey), U.K.

Patricia Cabredo Hofherr
UMR 7023 CNRS / U. Paris 8, France

The present study examines verbal morphology in Seri. Seri verbs have dedicated verb stems marking multiple events for singular and plural subjects respectively. We examine the semantics of multiple event forms to establish whether they express the same semantic category across sg and pl subject stems. We show that both sg and pl subject forms mark event plurality, but sg subject MULT-forms clearly differ from pl subject MULT-forms for older speakers: sg subject MULT-forms require distribution of multiple events in time, while the pl subject MULT-forms are distributive allowing simultaneous construals of the event multiplicity. Younger speakers tend to refuse contexts with clearly simultaneous construals. This suggests that the system is moving from two MULT-forms (a singular subject iterative and a plural subject distributive stem) to a system where both subject numbers have an essentially iterative MULT-form.

1 Introduction

Seri verbs generally have four forms whose morphology is not straight-forwardly predictable (1): there are many different exponents with a range of allomorphs, moreover even when different verbs share the same exponents, they often do not have the same function (Marlett, 2016; Baerman, 2016). Each verb expresses subject number agreement in a way that is not tied to specific exponents: two forms for singular subject and two for plural subject. Thus the Seri verbal inflectional system shows pervasive many-to-many mappings between form and function. Despite this lack of morphological homogeneity, subject number agreement reveals a homogeneous subject-number category across verbs.

(1) No one-to-one mapping between suffix form and function

Meaning	SG subject		PL subject	
intercept	-zac**tim**	-zaca**tim**	-zaca**toj**	-zaca**am**
lap up	-oaala	-oaala**tim**	-oaala**tam**	-oaala**toj**
fall	-poc	-poc**tim**	-poc**im**	-pocal**am**
unfasten	-aafp	-aafip**toj**	-atoofip**toj**	-atoofip**olca**

Within each subject number value, there is another distinction. Cabredo, Pasquereau & O'Meara (2019) (CPO) argue that the distinction is between a neutral form and a form marking event multiplicity (glossed MULT-forms). Prior work suggests that, like subject number, despite the morphological heterogeneity, this additional distinction also corresponds to a single feature (Marlett, 2016), i.e. a paradigm with subject number cross-cutting event multiplicity (2).

(2) Same-paradigm hypothesis: MULT.sg and MULT.pl are part of the same paradigm

run		Cat. 2: multiplicity	
		NEUTRAL	MULT
Cat. 1: subject number	SG	panzx	panozxim
	PL	pancojc	pancoxlca

However, given the complex morphology of Seri, this is not a trivial assumption. In this study, we compare SgSubj and PlSubj MULT-forms and show that for older speakers, event multiplicity is not homogeneous across subject number corresponding to a paradigm like (3).

Younger speakers' judgements indicate that they are levelling the paradigm to a single feature with consistent semantics across subject number like (2).

(3) Different-paradigms hypothesis: MULT.sg and MULT.pl do not express the same feature

run			NEUTRAL	MULT-1	MULT-2
Cat. 1: subject number	SG	panzx		panozxim	
	PL	pancojc			pancoxlca

We proceed as follows. In section 2, we examine the meaning of SgSubj MULT-forms. Section 3 then compares the SgSubj MULT-forms with PlSubj MULT-forms. Section 4 concludes.

2 Meaning of SgSubj MULT-forms

CPO show that SgSubj MULT-forms require a multiplicity of events. For instance, (4) is false in context A but true in context B.

(4) Context A: Yesterday, I went to Puerto Libertad and came back once.
     Context B: Yesterday, I went to Puerto Libertad several times.

   Moxima,   Xpanohax   conthayatim.
   yesterday Puerto_Libertad 3IO.DIR.1SBJ.RLYO.go.MULT

   Yesterday, I went to Puerto Libertad (several times).[EDSEI21OCT2018DRPM, elicitation]

In addition, SgSubj MULT-forms behave like pluractionals in other languages in two respects: the number of events cannot be counted by a cardinal numeral (CPO ex 19), and they do not multiply indefinites (CPO ex 21). Following CPO 2018 we conclude that SgSubj MULT-forms in Seri lexicalise a pluractional operator. Note, however, that Seri SgSubj MULT-forms are possible in contexts in which only two telic events take place (5). In this Seri SgSubj MULT-forms differ from other pluractional markers that do not allow contexts with a precise cardinality of events (cf. Van Geenhoven (2005) on the atelicity requirement of West Greenlandic pluractionals).

(5) Context: I hugged two children once one after the other.

   Xicaquiziil coi   isoj   cohyapxazl/   cohyapxazalim.
   children   DEF.PL 3POS.body 3IO.1SBJ.RLYO.cover 3IO.1.RLYO.cover.MULT

   I hugged the children (lit. I covered the children's body). [EDSEI23NOV2017DRPM, elicitation]

3 Comparison SgSubj MULT-forms vs PlSubj MULT-forms

CPO explicitly assume that SgSubj and PlSubj MULT-forms belong to a single category across the verbal paradigm and base their argumentation on SgSubj MULT-forms. As the present study examines precisely this assumption of a single category for SgSubj and PlSubj MULT-forms, we now proceed in two steps. First, we examine whether the arguments given in the previous section for SgSubj MULT-forms carry over to PlSubj MULT-forms, concluding that PlSubj mult-forms mark event-plurality, too (section 3.1). However, as is well-known, markers of event plurality are not a semantically homogeneous class (Dressler (1968); Yu (2003); Laca (2006) a.o.). In a second step we therefore compare the event pluralities marked by SgSubj MULT-forms with the event-pluralities marked by PlSubj MULT-forms (section 3.2). We show that there are systematic differences between SgSubj and PlSubj MULT-forms concerning simultaneous distribution over arguments. (For space reasons we only exemplify some diagnostics here).

3.1 Applying object number and aspect diagnostics to PlSubj mult-forms

Like SgSubj MULT forms, PlSubj MULT forms require an event plurality, are not compatible with cardinals counting events (6a) and do not multiply indefinite singulars (6b).

(6) a. Icatoomec hino coofin tintica xicacaziil quih sahmees pac
   week   1POS.to NMLZ.SBJ.happen MED.AW child.PL DEF orange INDEF.PL
   ihexej   / #ihexejam   isnaap yoozoj.
   INF.TRNS.buy.PL   INF.TRNS.buy.MULT.PL RLYO.6.times

   Last week, the children bought oranges 6 times. [EDSEIFLD3POST, elicitation]

   b. Context: Workers came to the village. Each man built his own house over the first few months.

   #Ctamcö coi haaco z iyaaizilca
   man.PL   DEF.PL ABS.house INDEF.SG 3;3.RLYO.make.MULT.PL

   The men built a house. SC: it sounds like they built one house together

We conclude that Seri PlSubj MULT-forms, too, are pluractional forms.

3.2 Distributive dependencies: SgSubj mult-forms vs. PlSubj mult-forms

Given that SgSubj and PlSubj MULT-forms behave as markers of event plurality, we now address the question whether they lexicalise the *same* pluractional marker by examining the distributive dependencies SgSubj and PlSubj MULT-forms allow.

The SgSubj MULT-form of the verb *hant quitox* 'drag' is not licensed by distribution over the object (7a) with simultaneous events, contrasting with the PlSubj MULT-form (7b). While for older speakers distribution over the object is sufficient to license the PlSubj MULT-form, younger speakers show variation and clearly prefer contexts imposing an iterative construal (i.e. suitcase-dragging events taking place one after the other).

(7) a. Context: At 2pm today, I saw Juan pulling his 3 suitcases behind him with 3 ropes.
   #Juan quih xiica an ihyaacalca quih hant iyootoxim.
   Juan   DEF suitcases   DEF down 3;3.RLYO.drag.MULT.SG

   Juan dragged the suitcases. [Questionnaire6FT3, elicitation]

   b. Context: At 2pm today, I saw Juan, Isaac and Manuel each pulling one suitcase.

   Xicacaziil quih xiica an ihyaacalcoj quih hant iyootyaxlca.
   child.PL   DEF suitcases   DEF down 3;3.RLYO.drag.MULT.PL

   The boys dragged the suitcases. (40 + : true, 40-: variation)

As in (7b), all older speakers accept distribution of the multiplicity of events required by the PlSubj MULT-form *cöcatooquelam* 'cross (intr)' over the plural subject without distribution over time (8), whereas younger speakers' judgements are more varied (sometimes they reject the truth of the example in the context, sometimes they accept it commenting that it would be better with distribution over time as well).

(8) Context: The women crossed the brook together, once. ([QuestionnaireFT3, elicitation])

   Cmajiic quih hant ipzx com imac cöyatooquelam.
   woman.PL DEF brook   DEF.SG.lying 3POSS.middle 3IO.cross.MULT.PL

   The women crossed the brook. (40 + : true, SC: because there's several, 40-: false)

These examples show that older speakers allow simultaneous construals for multiple events with PlSubj while younger speakers prefer iterative construals for SgSubj and PlSubj mult-forms (distribution in time).

The hypothesis that the preference for distribution in time is an innovation finds additional support in the meaning of certain lexicalised phrases which can still be analyzed as being composed of a MULT-form. For instance, the phrase meaning *gray whale* contains the MULT-form of the verb 'be white'. We hypothesise that at the time the phrase was lexicalised, the MULT-form was licensed by spatial distribution (the property of being white being predicated of spatially-distinct parts of the whale).

(9) Ziix cooxapoj   (Marlett, 2016)
     thing SBJ.NMLZ.white.MULT
     Gray whale (lit. whale that is spotted with white/ white here and there)

4   Conclusion

SgSubj and PlSubj MULT-forms in Seri are pluractional forms: they can only be used in contexts that convey a multiplicity of events. However, for older speakers, SgSubj and PlSubj MULT-forms do not have the same semantics: SgSubj MULT-forms are iterative requiring distribution over time whereas PlSubj MULT-forms are distributives, allowing distribution over time, locations or arguments. Younger speakers, in contrast, are associating the pluractional component of SgSubj and PlSubj MULT-forms to a uniform semantics of iterativity. The difference between older and younger speakers w.r.t. (7) and (8) can be interpreted as reflecting a linguistic change whereby for older speakers, distribution over the object/intr. subject argument (without obligatory distribution over time) is enough to license PlSubj MULT-forms suggesting that SgSubj and PlSubj do not belong in the same paradigm (cf. 3), while younger speakers prefer a uniform semantics of distribution in time for SgSubj and PlSubj MULT-forms consolidating the forms into an orthogonal paradigm (from two MULT-categories in (3) to one MULT-category in (2)) by interpreting SgSubj and PlSubj MULT-forms as expressing a single category of iterative pluractionality. Thus Seri exemplifies the development of a single cross-classifying feature value from two independent features.

References

Baerman, Matthew. 2016. Seri verb classes: morphosyntactic motivation and morphological autonomy. *Language* 92. 792–823.

Cabredo, Patricia, Jérémy Pasquereau & Carolyn O'Meara. 2019. Event plurality in Seri. In K. Johnson & A. Goebel (eds.), *Proceedings of SULA 10: Semantics of Under-represented Languages in the Americas,* 1–16. GLSA.

Dressler, Wolfgang. 1968. *Studien zur verbalen Pluralität.* Wien: Bühlau in Kommission.

Laca, Brenda. 2006. Indefinites, quantifiers and pluractionals: what scope effects tell us about event pluralities. In S. Vogeleer et al. (eds.), *Non-definiteness and plurality,* 191–217. Amsterdam: Benjamins.

Marlett, Stephen A. 2016. Cmiique Iitom: the Seri language. Unpublished grammar (draft).

Van Geenhoven, Veerle. 2005. Atelicity, pluractionality, and adverbial quantification. In H. Verkuyl et al. (eds.), *Perspectives on aspect,* 107–125. Berlin: Springer.

Yu, Alan. 2003. Pluractionality in Chechen. *Natural Language Semantics* 11. 289–321.

Verbal periphrasis in Bulgarian

Gergana Popova *Andrew Spencer*
Goldsmiths, University of London University of Essex

Introduction

Bulgarian has a rich set of periphrastic constructions (Spencer, 2003), including a perfect (1a, 2b) and a renarrated (evidential) series, usually, (1c), but not always, (2c), syncretic with the perfect indicative.

(1) a. Ti **si** **pisala** pisma do kmeta.
   you be.2SG.PRS write.L-PTCP.SG.F letters to mayor.DEF

   'You have written letters to the mayor'. [perfect indic]

 b. Ti **pisa** pisma do kmeta.
   you write.2SG.AOR letters to mayor.DEF

   'You wrote letters to the mayor'. [aorist indic]

 c. Ti **si** **pisala** pisma do kmeta včera.
   you be.2SG.PRS write.L-PTCP.SG.F letters to mayor.DEF yesterday

   '(Reportedly) you wrote letters to the mayor yesterday.' [renarr aor]

(2) a. Tja **pisa/pišeše** pisma do kmeta.
   she write3SG.AOR/IMPF letters to mayor.DEF

   'She wrote/was writing letters to the mayor'. [aor/impf indic]

 b. Tja **e** **pisala** pisma do kmeta
   she be.3SG.PRS write.L-PTCP.SG.F letters to mayor.DEF

   'She has written letters to the mayor'. [perfect indic]

 c. Tja **pišela** pisma do kmeta
   she write.IPFV.L-PTCP.SG.F letters to mayor.DEF

   '(Reportedly) she writes/was writing letters to the mayor.' [renarr pres/impf]

We propose a construction-theoretic analysis within a modification of Stump's (2016) PFM2 model.

Previous Word-&-Paradigm accounts

Periphrases resemble realizational inflectional morphology in that it is typically impossible to assign a constant meaning or feature value to the parts (auxiliary/ancillary element or (inflected) lexical form). They are also often integrated into the synthetic inflectional paradigm: the renarrations in (1c, 2c) alternate with synthetic indication constructions (1a, 2a) (Sadler & Spencer, 2001; Ackerman & Stump, 2004). Previous word-and-paradigm accounts (e.g. Ackerman & Stump, 2004) integrate periphrasis into the morphology directly. In Bonami's (2015) analysis, morphological rules specify that the lexical word component in a periphrase must occur in the context of a particular ancillary element.

However, for Bonami & Webelhuth (2018) the periphrase is essentially the head (the auxiliary verb) subcategorising for the lexical element (e.g. a participle) via ARG-ST. In the sense that the whole of the periphrase is contained in the lexical entry of the auxiliary, and the lexical entry is listed, this account comes close to listing a construction. To deal with non-compositionality, many previous accounts distinguish two kinds of features, along the lines of

the m-/s-features of Sadler & Spencer (2001) or the CONTENT/FORM paradigm distinction in Stump (2016), e.g. Bonami & Webelhuth's (2018) HEAD/INFL attributes. However, there is no direct access to the mother node of the construction in the Bonami & Webelhuth (2018) account, so the features expressed holistically by the construction have to be HEAD features of the syntactic head of the periphrase. When percolated to the next syntactic level they effectively serve as constructional features.

Where the auxiliary is the syntactic head, as in the Czech past tense construction, nearly all of its HEAD properties, including its Lexemic Identifier (LID), are those of its complement, the lexical verb. Thus for Czech *nekoupil jsem* 'I didn't buy' this induces a near-complete HEAD/INFL feature mismatch on the auxiliary: [HEAD|LID: koupit, VFORM: pst, POL: neg, AGR: m,3,sg] vs [INFL|LID: pst_aux, VFORM: prs, POL: pos, AGR: m,3,sg]. However, since the auxiliary's HEAD|LID value is that of the lexical verb complement, this makes the false prediction that the aux cannot scope over conjoined verbs, as in *ten dopis jsem [napsal ale neposlal]* 'that letter AUX.1SG [wrote but NEG.sent]'. More generally, this account predicts that the features associated with the periphrase as a whole will originate on at least one of its components and that in many cases the head of a periphrase is morphomic, in that the features governing its syntactic behaviour and semantic interpretation are not the features that it inflects for. We argue that the features components of the periphrase inflect for may be important for determining their syntactic behaviour.

Our account

We argue that no existing analyses take non-compositionality fully into account. It is especially problematic to assume that the features associated with the periphrase as a whole originate on one of its elements: no part of the periphrase in (1c) expresses the feature 'renarrated' or, indeed, 'aorist'. We argue therefore that periphrastic constructions should permit direct reference to the features associated with the construction as a whole (suspending the usual mechanism of bottom-up HEAD feature percolation for such constructions). We assume the existence in the grammar of special, periphrastic constructions, which themselves have HEAD/INFL features (see Figure 2). We then allow top-down mapping of constructional HEAD/INFL features to the HEAD/INFL features of the elements of the construction. We assume that the INFL features such constructions express are regulated by the syntax-morphology interface, so the paradigm function can map features directly to the INFL features on the construction level:

(3) $\mathrm{PF}(<\mathrm{L}, \sigma: \{\mathrm{TNS\text{-}ASP}: \mathrm{perf}\}>) = (<\mathrm{C}_{perf-cxt}, \sigma>)$

This ensures that periphrastic constructions are integrated into the (form) paradigms of lexemes.

We interpret the INFL attribute of Bonami (2015), Bonami & Webelhuth (2018) as a set of m/FORM features, by default identical to a set of (syntactically visible) HEAD features. To account for the syncretism of periphrastic constructions we noted in the beginning, we assume that a mechanism along the lines of the property mapping function ***pm*** of Stump (2016) maps INFL to HEAD features on the constructional level too, see also Figure 2:

(4) $\mathrm{pm}\,(<\mathrm{C}_{perf-cxt}, \sigma: \{\mathrm{TNS\text{-}ASP}: \mathrm{aor}, \mathrm{MOOD}: \mathrm{renar}\}>) =$
       $(<\mathrm{C}_{perf-cxt}, \tau:\{\mathrm{TNS\text{-}ASP}: \mathrm{perf}, \mathrm{MOOD}: \mathrm{indic}\}>)$

Our analysis enjoys a number of advantages:

1. Assuming that the distinctions encoded periphrastically are expressed on the constructional level obviates the need to consider the inflectional information expressed by the elements of the periphrase irrelevant to syntax. At the same time none of the components of the periphrase need be tasked with adopting the constructional features as their own.

2. The construction can encode various syntactic relationships, including shared argument structure, where necessary. However, the status of the construction as an exponent of certain constructional features is not dependent on lexical selection (much less 'reverse selection', Bonami 2015) and headedness.

3. Since the main featural mismatches are between the top-level construction and the *default* interpretation of its component words, not between HEAD INFL features, we do not encounter the problem earlier earlier accounts face with coordinated verbs.

$$
\text{HEAD } \sigma : \begin{bmatrix} \begin{bmatrix} \text{LID} & \boxed{1} \\ \text{TNS-ASP} & \text{PRF} \\ \text{AGR} & \begin{bmatrix} \text{AGR1} & \boxed{2}\{\text{P:2, N:sg}\} \\ \text{AGR2} & \boxed{3}\{\text{G:f, N:sg}\} \end{bmatrix} \\ \text{MOOD:} & \text{indic} \end{bmatrix} \end{bmatrix}
$$

$$
\text{INFL } \tau : \begin{bmatrix} \begin{bmatrix} \text{LID} & \boxed{1} \\ \text{TNS-ASP} & \text{PRF} \\ \text{AGR} & \begin{bmatrix} \text{AGR1} & \boxed{2}\{\text{P:2, N:sg}\} \\ \text{AGR2} & \boxed{3}\{\text{G:f, N:sg}\} \end{bmatrix} \\ \text{MOOD:} & \text{indic} \end{bmatrix} \end{bmatrix}
$$

$$
\begin{bmatrix} \text{HEAD} & \boxed{5} \\ \text{INFL} \boxed{5} & \sigma! \begin{bmatrix} \text{LID} & \boxed{4}\text{ be} \\ \text{TNS-ASP} & \boxed{5}\text{ PRS} \\ \text{AGR1:} & \boxed{2} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \text{HEAD} & \boxed{6} \\ \text{INFL} \boxed{6} & \sigma! \begin{bmatrix} \text{LID} & \boxed{1}\text{ write} \\ \text{AGR2:} & \boxed{3} \\ \text{VFORM} & \text{l-form-pfv} \end{bmatrix} \end{bmatrix}
$$

Figure 1: Bulgarian: 'you (2SG.F) wrote (Perfect)'

$$
\text{HEAD } \sigma : \begin{bmatrix} \text{LID} & \boxed{1} \\ \text{TNS-ASP} & \text{AOR} \\ \text{AGR} & \begin{bmatrix} \text{AGR1} & \boxed{2}\{\text{P:2, N:sg}\} \\ \text{AGR2} & \boxed{3}\{\text{G:f, N:sg}\} \end{bmatrix} \\ \text{MOOD:} & \text{renarr} \end{bmatrix}
$$

$$
\text{INFL } \tau : \begin{bmatrix} \text{LID} & \boxed{1} \\ \text{TNS-ASP} & \text{PRF} \\ \text{AGR} & \begin{bmatrix} \text{AGR1} & \boxed{2}\{\text{P:2, N:sg}\} \\ \text{AGR2} & \boxed{3}\{\text{G:f, N:sg}\} \end{bmatrix} \\ \text{MOOD:} & \text{indic} \end{bmatrix}
$$

$$
\begin{bmatrix} \text{HEAD} & \boxed{5} \\ \text{INFL} \boxed{5} & \sigma! \begin{bmatrix} \text{LID} & \boxed{4}\text{be} \\ \text{TNS-ASP} & \boxed{5}\text{PRS} \\ \text{AGR1:} & \boxed{2} \end{bmatrix} \end{bmatrix}
\quad
\begin{bmatrix} \text{HEAD} & \boxed{6} \\ \text{INFL} \boxed{6} & \sigma! \begin{bmatrix} \text{LID} & \boxed{1}\text{write} \\ \text{AGR2:} & \boxed{3} \\ \text{VFORM} & \text{l-form-pfv} \end{bmatrix} \end{bmatrix}
$$
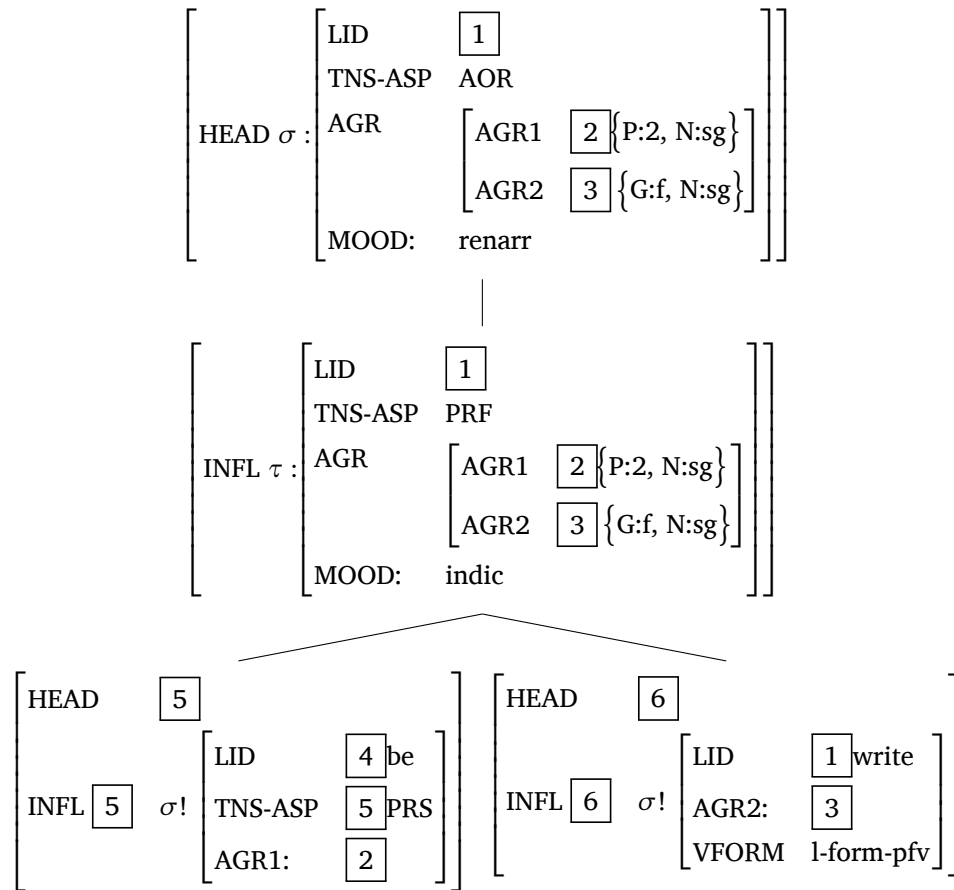
Figure 2: Bulgarian: 'you (2SG.F) wrote (Renarrated aorist)'

References

Ackerman, Farrell & Gregory Stump. 2004. Paradigms and periphrastic expressions. In Louisa Sadler & Andrew Spencer (eds.), *Projecting morphology* Stanford Studies in Morphology and the Lexicon, 111–157. Stanford, California: CSLI Publications.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25(1). 63–110.

Bonami, Olivier & Gert Webelhuth. 2018. Periphrasis and morphosyntactic mismatch in Czech. DGfS talk.

Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2000*, 71–96. Dordrecht: Kluwer.

Spencer, Andrew. 2003. Periphrastic paradigms in Bulgarian. In Uwe Junghanns & Luka Szucsich (eds.), *Syntactic structures and Morphological Information*, 249–282. Berlin: Mouton de Gruyter.

Stump, Gregory. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface.* Cambridge: Cambridge University Press.

Are French NNs variants of N-PREP-N constructions?

Jan Radimský

University of South Bohemia
in České Budějovice

1 Introduction

The aim of this paper is to provide a thorough empirical examination of the hypothesis that French subordinate Noun-Noun compounds, such as *roman photos* ('photo novel'), are nothing more than variants of corresponding syntactic phrases or phrasal lexemes (such as *roman* [*avec (des) photos*]$_{PP}$) (cf. Fradin, 2009:433, among others). With reference to data from both synchronic and diachronic corpora (*FrWac, Frantext, Google Books*), it will be argued that the competition between the N-N compounding pattern, on the one hand, and the corresponding phrasal lexemes or noun phrases with the N-PREP-/ART/-N structure, on the other, is a complex phenomenon that documents a certain change in naming strategies in French, especially from the 1960s onwards.

2 State of the art

Romance subordinate NNs are said to appear progressively during the 19th-20th Centuries as instances of a new pattern that enters in competition with older and extremely profitable patterns of phrasal lexemes (N + PREP + N and N + A). As far as French is concerned, a recent study carried out on the *Frantext* corpus has showed that the profitability of French subordinate NNs increased very slowly from the beginning of the 19th Century to 1960s with an exponential increase observed afterwards (Radimský, 2019). Therefore, many binominals, such as (1a), (2a), are attested also as phrasal lexemes with a preposition between the two nouns, the preposition being either bare (1b) or accompanied by a determiner (2b). In other cases, the attested 'prepositional' equivalent of a binominal seems to be a free syntactic phrase, as in (3b) compared to (3a).

(1a) stylo-bille

   pen$_N$ ball$_N$ – 'ballpoint pen'
(1b) stylo à bille

   pen$_N$ to$_{PREP}$ ball$_N$ – 'ballpoint pen'
(2a) bière pression

   beer$_N$ pressure$_N$ – 'draught beer'
(2b) bière à la pression

   beer$_N$ to$_{PREP}$ the$_{ART}$ pressure$_N$ – 'draught beer'
(3a) traffic voyageurs

   traffic$_N$ passengers$_N$ – 'passenger traffic'

(3b) traffic des voyageurs

    traffic$_N$ of$_{PREP}$ the$_{ART}$ passengers$_N$ – 'traffic of the passengers'

These facts lead to a commonly shared assumption that French subordinate NNs could be mere variants of the corresponding prepositional constructions. Fradin (2009:433) even goes so far as to claim that the prepositional constructions are older, which could, in turn, provide a support for the assumption put forward by Hatcher (1946), according to which the origin of French subordinate NN's lies in the instability of prepositions in N-PREP-N constructions.

Conversely, Arnaud (2015) claims that there are also many French subordinate NNs for which the corresponding prepositional construction is either attested in negligible numbers only, or even impossible to form, as in (4).

(4) *portrait robot*

    portrait$_N$ robot$_N$ – 'facial composite'

A thorough analysis of corpus data therefore seems necessary in order to determine, from both the qualitative and quantitative points of view, which NN's have (or even *can* have) prepositional counterparts and if so, to compare their diachronic frequency curves.

3 Data analysis

The analysis will start out from a sample of more than 1700 non-coordinate French NNs that comprise, in terms of the Scalise-Bisetto (2009) classification, attributive NNs, subordinate grounding compounds, and subordinate verbal-nexus compounds. Concerning these macro-classes, data drawn from the *FrWac* corpus allow us to argue that the prepositional counterparts of NNs are attested for most of verbal-nexus compounds (more than 90% of types), but only for 2/3 of subordinate grounding compounds, and for a minor part of attributive compounds, which corroborates the observations of Baroni, Guevara and Pirrelli (2009) concerning similar data from Italian. Indeed, for the subordinate verbal-nexus type the prepositional syntactic construction (3b) seems to be always available, while for attributive NNs (such as *guerre-éclair* – 'Blitzkrieg') it is often impossible, provided that the attributive relationship is rather paraphrasable by a copulative construction.

Within the subordinate grounding class, which is of major concern, NNs will be carefully divided into the following categories: the prepositional variant is impossible (5), the prepositional variant is extremely rare (6), both variants are in free competition (7), the NN variant is extremely rare (8).

(5) *confiture (Ø) maison,*   *yaourt (Ø) nature*
     homemade jam   natural yogurt

(6) *version (sur) papier*
     paper version

(7) *profil (d' / de l') utilisateur,*   *animateur (de) télé,*   *bière (à la) pression*
     user profile   TV entertainer   draught beer

(8) *boutique (de) souvenirs*
     souvenir shop

On this basis, diachronic data from Frantext corpus and Google books will make it possible to draw frequency curves for the respective variants, as exemplified in Fig. 1 and Fig. 2.
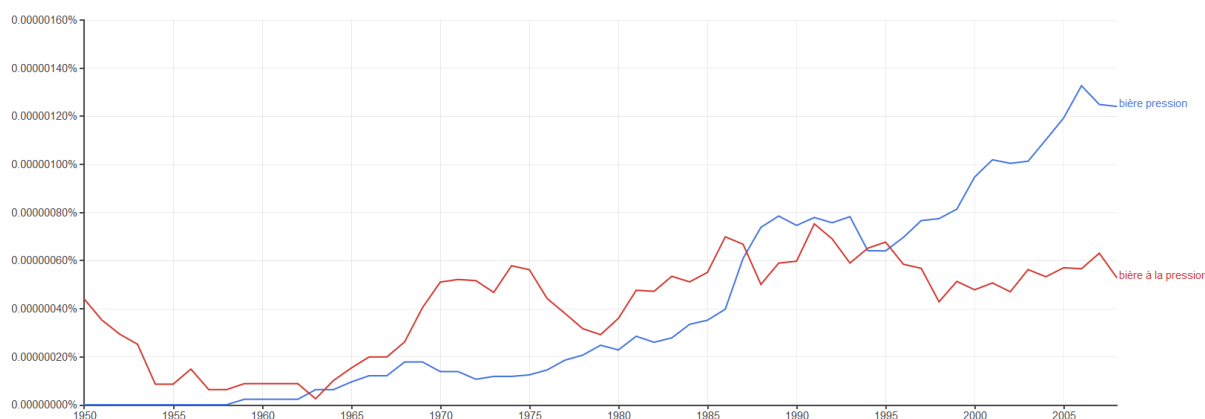
Fig. 1 – Relative frequencies of *bière pression* and *bière à la pression* from 1950s[1]
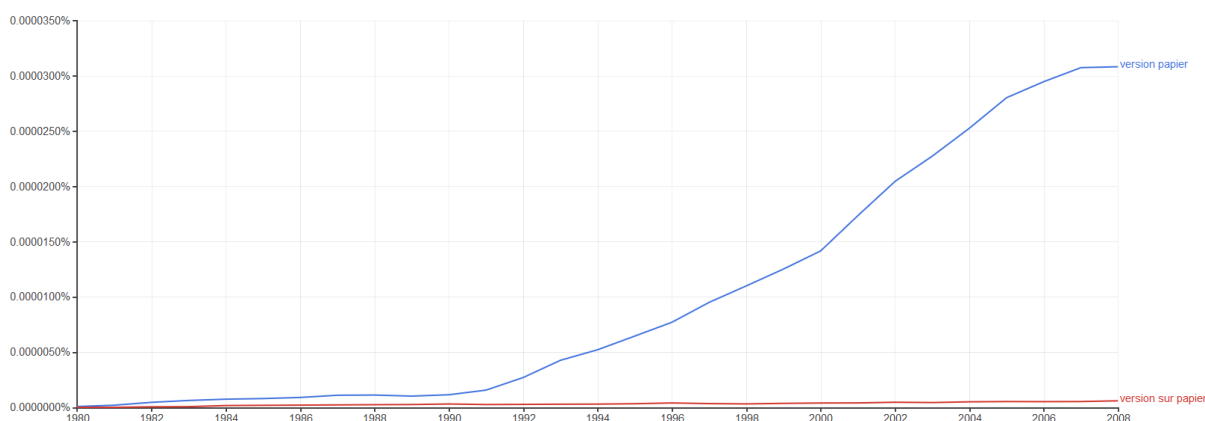
Fig. 2 – Relative frequencies of *version papier pression* and *version sur papier* from 1980s[2]

---

[1] https://books.google.com/ngrams

[2] https://books.google.com/ngrams

These data suggest that while some subordinate NNs might be considered as a reduced variant of a prepositional construction (Fig. 1), for others – and especially for those first attested after 1960s – the NN form was the default one from the origin (Fig. 2) or even the only form possible (see Examples (5)).

4 Conclusion

The data analysed in this paper suggest that during the second half of the 20th Century, French subordinate NNs progressively become instances of an independent word-formation pattern that cannot be directly linked to – or derived from – the corresponding N-PREP-N constructions, irrespective of the fact whether these prepositional constructions are intended to be phrasal lexemes or free syntactic phrases.

References

Arnaud, Pierre J. L. 2015. Noun-noun compounds in French. In Peter Müller et al. (eds.), Word formation. An international handbook of the Languages of Europe. Vol. 3. Berlin/Boston: Walter de Gruyter. 673-687.

Arnaud, Pierre. 2003. *Les composés "timbre-poste"*. Lyon: Presses universitaires de Lyon.

Baroni, Marco & Emiliano Guevara & Vito Pirrelli. 2009. Sulla tipologia dei composti N + N in italiano: principi categoriali ed evidenza distribuzionale a confronto. In: Ruben Benatti & Giacomo Ferrari & Monica Mosca (eds.), *Linguistica e modelli tecnologici di ricerca* (Atti del 40esimo Congresso della Società di Linguistica Italiana). Roma: Bulzoni. 73-95.

Ferraresi, Adriano et al. 2013. *frWaC*. Ústav Českého národního korpusu FF UK, Praha. Available at <http://www.korpus.cz>.

Fradin, Bernard. 2009. IE, Romance: French. In Rochelle Lieber, Pavol Štekauer, The Oxford handbook of compounding. Oxford: Oxford university press. 417-435.

Frantext. Corpus de langue française. <https://www.frantext.fr>

Google Books – Available at <https://books.google.com/ngrams>.

Google ngrams – Available at <https://books.google.com>.

Hatcher, Anna G. 1946. Le Type Timbre-Poste. *Word*, 2:3. 216-228.

Radimský, Jan. 2019. Les composés N-N de subordination : un paradigme emergent. *Studia Romanica Posnaniensia.* Poznań: Adam Mickiewicz University Press. Being printed.

Scalise, Sergio & Antonietta Bisetto. 2009. The classification of compounds. In: Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 34-53.

Modelling the interaction of regularity and morphological structure: the case of Russian verb inflection

Selena Rorberi

University of Pisa

Claudia Marzi

Institute for Computational Linguistics - CNR

1 Introduction

Modelling complex inflection systems, such as conjugation in Modern Greek, Italian or Russian, requires careful consideration of a number of factors, ranging from pervasive stem allomorphy to the identification of the appropriate inflection class and the inferential predictability of morpho-phonological processes. Descriptive approaches have taken different views on how to account for degrees of morphological (ir)regularity, while making different predictions about the way speakers process regular and irregular forms in highly-inflecting languages. In the present paper, we assess the psycholinguistic implications of two radically different approaches to the description of the Russian verb system: a more traditional approach dating back to Jakobson (1948), and a Words and Paradigm approach (Brown 1998). Based on recent fMRI evidence (Slioussar et al. 2014) and original results of a neural network simulation with recurrent self-organising maps (Ferro et al. 2011; Marzi et al. 2014; Pirrelli et al. 2015; Marzi et al. 2016), we suggest that both approaches are *prima facie* compatible with Russian data, while being in contrast with Pinker's claim that the regular-irregular distinction is an epiphenomenon of the storage-processing dichotomy in the human language faculty (Pinker & Ullman 2002). We argue that this evidence lends support to integrative models of the mental lexicon (Marzi & Pirrelli 2015), accounting for a graded interaction between regularity and morphological structure.

2 The Russian verb

Traditionally, the description of Russian verb inflection is cast into the classical derivational analysis first proposed by Roman Jakobson (1948). For each verb lemma, Jakobson postulates the existence of a unique, underlying stem, which may undergo a variety of morpho-phonological processes as a function of the class to which the verb belongs, and its specific set of endings (including class-specific thematic vowels, or "thematic ligatures"). A number of verb classes are identified, whose variety reflects the type and number of the morpho-phonological processes needed to turn an underlying base into a surface allomorph. Classes are identified by the suffix classifier in the verb stem: *-aj-, -ej-, -a-, -e-, -i-, -o-, -ova-, -avaj-, -nu-*. In particular, the classifier determines the conjugation class (i.e. the specific set of inflectional endings selected by the verb), the adjustment of the root final consonant (i.e. the root consonant immediately preceding the classifier), and the suffix alternation (e.g. *-ova-* alternates with *-uj-*). To illustrate, the stem *chitaj-* of the verb *chitat'* ('read') drops the final *-j* before an ending that begins with a consonant (e.g. past tense *chita-l*), but keeps *-j* when the ensuing ending begins with a vowel (e.g. *chitaj-u* 'I read'). In contrast, the stem *pisa-* of the verb *pisat'* ('write') drops its final vowel when the affixed ending begins with a vowel. In turn, this triggers consonant softening throughout the present indicative paradigm (e.g. *pish-u* 'I write'). As verb stems in any class are assumed to undergo some stem alternation, regularity is measured by the number of applicable processes. Accordingly, the *-aj-* class is more regular than the *-a-* class, since the former undergoes consonant truncation only, whereas the latter undergoes both vowel truncation and consonant softening.

More recently, Brown (1998) proposes a paradigm-based account of Russian verb inflection, cast into a Network Morphology framework (Brown & Hippisley 2012). The analysis focuses on the number of stem alternants associated with specific cells in a verb paradigm, independently of the degree of formal predictability or the number of processes involved in stem formation. Unlike Jakobson's analysis, in a verb like *chitat'* the stem is analysed as ending in a vowel (*chita-*); *-j-* is infixed when the stem is followed by a vowel-initial ending (*chita-j-u*). The so-called *-aj-* class in fact includes those verbs that keep their *a*-ending stem unaltered throughout the whole paradigm. This represents a kind of default class. In contrast, a verb paradigm with more stem alternants is less regular and more difficult to master and generalise than a verb paradigm with fewer or no stem alternants. Thus, regularity is expressed in terms of surface relations between paradigmatically-related verb forms. In regular paradigms, invariant stems are shared by all inflected forms, and are transparently perceived by the speakers. Conversely, irregular paradigms select more than one stem alternant, which are differently indexed, depending on the verb class.

2.1 Psycholinguistic implications

In spite of considerable differences in their formal apparatus, both approaches account for a graded notion of morphological regularity and its interaction with word processing. Following Jakobson, the more processes are involved in mapping allomorphs onto an invariant stem, the longer it takes a speaker to master them. In Brown's account, paradigms with more stem alternants are more difficult to process because their simultaneous availability in the speakers' long-term memory causes their co-activation and mutual competition during processing. For example, competing co-activation of *pisa-* and *pish-* as stem alternants of *pisat'* slows down their processing in recognition.

In general, paradigms with more stem alternants require stipulation of more morpho-phonological processes. Due to this correlation, the most regular class of *-aj-* verbs in Jakobson's approach (requiring one *j*-deletion rule) coincides with the regular class of invariant verb stems in Brown's account. In addition, both accounts predict that difficulty of processing, as well as ease of generalisation and learning, should vary continuously as a function of graded levels of regularity. Nonetheless, there is one point where the two accounts diverge. In radically amorphous versions of the *Word-and-Paradigm* approach (Blevins 2016), as well as in connectionist frameworks, the mapping of an input inflected form onto its sublexical constituents ((prefix +) stem + ending, for Russian verb forms) is a continuous function of the statistical regularities of inflectional paradigms. Accordingly, perception of morphological boundaries may vary as a result of the probabilistic support sublexical boundaries receive from frequency distributions of surface exemplars (e.g. Hay & Baayen 2005; Plaut & Gonnerman 2000; Rueckl & Raveh 1999). It follows that processing of regularly inflected forms should be more sensitive to their morphological structure and to type frequency effects than the processing of irregulars. Conversely, irregulars are processed holistically, in a way that is sensitive to token frequency effects.

3 Computational evidence

We provide data-driven evidence of the complex interaction in processing of a graded notion of (ir-)regularity and the morphological structure. Sixteen fully inflected verb forms have been selected for each of the 50 top frequency Russian verb paradigms (i.e. 50 aspectual pairs, which include 10 present and past tense imperfective forms, 6 perfective forms for the future tense) sampled from a reference corpus (Jakubíček et al. 2013). Without any information of morphological structure, they are learned by a recurrent self-organising neural network

(TSOM), consisting of a two-dimensional grid of artificial memory/processing nodes that dynamically memorise input strings as chains of maximally-responding processing nodes (Best Matching Units). The prediction-driven bias of its temporal layer of re-entrant connections makes strong expectations over upcoming symbols accounting for successful serial word processing. Figure 1 illustrates the dynamic of word access at the end of learning (i.e. epoch 100) by showing prediction rates at each letter position relative to the stem-ending boundary (or morpheme boundary, centred on *x*-axis = 0). Prediction scores are calculated by incrementally assigning each correctly anticipated symbol in the input a 1-point score. The more input symbols are anticipated, the easier the prediction of the verb form, the lower its processing load.

Our evidence suggests that perception of morphological structure interacts with regularity and formal transparency. The more prominent increase in prediction rates on more regular stems suggests a clear paradigmatic effect: the more verb forms share the same stem, the easier their processing. On the contrary, the drop in prediction reflects an increase in the processing effort made by the map in predicting an upcoming inflectional ending at the end of the stem. We take such a discontinuity to mark a clear structure-driven effect of processing "surprisal" (Levy 2008), due to an increase in entropy of the transitional probability from a regular stem to its grammatical endings. This is confirmed by the steeper increase in prediction rates for inflectional endings (positive *x* values in Figure 1) when they follow more irregular stems: stem allomorphs can anticipate inflection information thus reducing uncertainty for selection of ending.
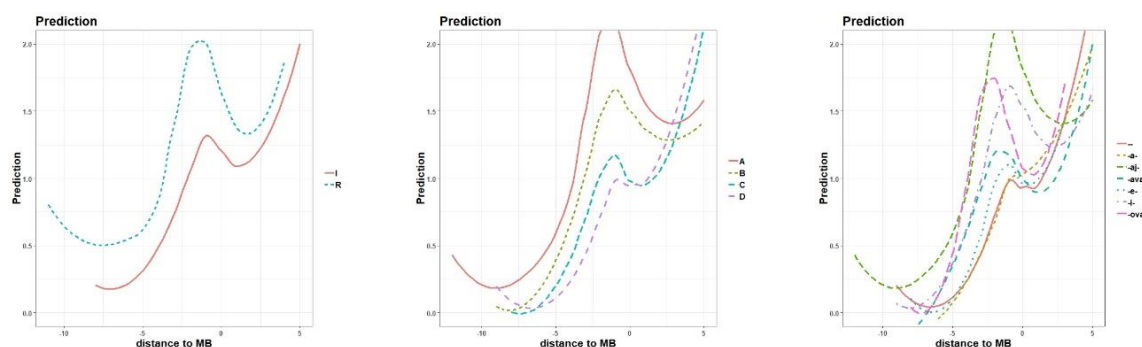
Figure 1. Regression plots of interaction effects between morphological (ir)regularity and distance to morpheme boundary (MB), in non-linear models (GAMs) fitting the number of symbols predicted by a TSOM: categorical fixed effect are (left panel) regularity (green dashed lines) vs. irregularity (red solid lines), (central panel) a gradient of regularity, and (right panel) suffixes classes.
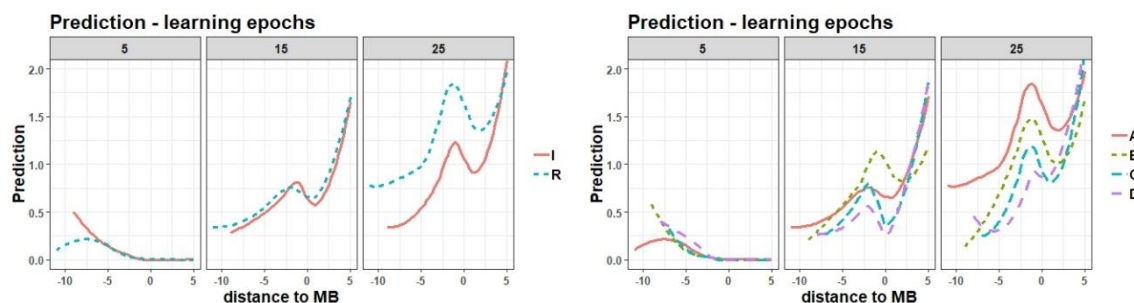
Figure 2. Regression plots of interaction effects between morphological (ir)regularity and distance to morpheme boundary (MB) for learning epochs 5, 15, 25, in non-linear models (GAMs) fitting the number of symbols predicted by a TSOM: categorical fixed effect are (left panel) regularity (green dashed lines) vs. irregularity (red solid lines), and (right panel) a gradient of regularity.

The effect is consistent. We observed it using three classification criteria for inflection regularity of different grain-size: (a) the traditional dichotomy between the class of *-aj-* verbs (Regular) and the class of non *-aj-* verbs (Irregular); (b) a more granular subdivision between *-aj-* verbs (class A), productive *-i-* and *-ova-* verbs (class B), *-a-*, *-e-*, *-avaj-* verbs (class C), and radically suppletive paradigms (class D); and (c) all suffix-based classes attested in our training set. In all cases, more regular verb classes, when compared with less regular classes, show higher prediction rates overall, while exhibiting a greater discontinuity in prediction at the stem-ending boundary. As shown in Figure 2, perception of morphological structure gradually emerges through the training epochs, as learning progresses.

Our results well agree with evidence of word processing load reported by Slioussar and colleagues in the task of generating 1Sg present tense forms of regular and irregular Russian verbs (Slioussar et al. 2014). In their experiment, regulars were found to require less attention, working memory and decision-making than irregulars. While their evidence appears to support an integrative model of word processing, our results address the important, related question of how similar effects may arise in a recurrent self-organising network that simulates the concurrent dynamic storage of paradigmatically related forms. In the end, the interaction between regularity and morphological structure appears to be more compatible with a *Word-and-Paradigm* account of Russian verb inflection, than with Jakobson's account, which does not make the same prediction.

References

Brown, Dunstan. 1998. Stem Indexing and Morphological Selection in the Russian Verb: a Network Morphology Account. In Ray Fabri, Albert Ortmann & Teresa Parodi (eds.), *Models of Inflection*, 196-224. Max Niemeyer Verlag.

Brown, Dunstan & Andrew Hippisley. 2012. *Network Morphology: A Defaults-based Theory of Word Structure*. Cambridge University Press.

Ferro, Marcello, Claudia Marzi & Vito Pirrelli. 2011. A Self-Organizing Model of Word Storage and Processing: Implications for Morphology Learning. *Lingue e Linguaggio*, X(2). 209-226.

Jakobson, Roman. 1948. Russian Conjugation. *Word*, 4(3). 155-167.

Jakubíček, Miloš, Adam Kilgarriff, Vojtêch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The tenten corpus family. In *7th International Corpus Linguistics Conference CL*, 125-127. https://www.sketchengine.eu/rutenten-russian-corpus/

Levy, Roger. 2008. Expectation-based Syntactic Complexity. *Cognition* 106. 1126-1177.

Marzi, Claudia, Marcello Ferro & Vito Pirrelli. 2014. Morphological structure through lexical parsability. *Lingue e Linguaggio*, XIII(2). 263-290.

Marzi, Claudia, Marcello Ferro, Franco Alberto Cardillo & Vito Pirrelli. 2016. Effects of frequency and regularity in an integrative model of word storage and processing. *Italian Journal of Linguistics*, 28(1). 79-114.

Pinker, Steven & Michael Ullman. 2002. The past and future of the past tense. *TRENDS in Cognitive Sciences* 6(11). 456-463.

Pirrelli, Vito, Marcello Ferro & Claudia Marzi. 2015. Computational complexity of abstractive morphology. In Baerman, Matthew, Dustan Brown & Greville Corbett (eds.), *Understanding and Measuring Morphological Complexity*. Oxford: Oxford University Press. 141-166.

Slioussar, Natalia, Maxim V. Kireev, Tatiana V. Chernigovskaya, Galina V. Kataeva, Alexander D. Korotkov & Svyatoslav V. Medvedev. 2014 An ER-fMRI study of Russian inflectional morphology. *Brain & Language* 130. 33-41.

Word formation with loanwords: A case of "Japanese English"

Masaharu Shimada   *Akiko Nagano*

University of Tsukuba   Tohoku University

1 Introduction

The topic of this paper is how English has been influencing contemporary Japanese lexicon and Japanese word formation, especially compounding. Many English words are used in daily communication among native speakers of Japanese, producing such compounds as in (1).

(1) a. *bakku-miraa*   (lit. back-mirror) 'a rear-view mirror'

　　b. *teeburu-supiichi* (lit. table-speech) 'a dinner speech'   (Shibatani (1990: 151))

The Japanese lexicon consists of three types of words: native Japanese words, Sino-Japanese words, and foreign words (called *gairaigo* in Japanese), such as English words. The complex word *bakku-miraa* in (1a) is made up from two English words, *bakku* and *mirra,* for example. However, the expressions in (1) are not shared with native speakers of English and often labelled as "wasei eigo." *Wasei* means 'made in Japan' and *eigo* means 'English,' so the literal translation of "wasei eigo" would be "Japanese English." We use these terms exchangeably.

Similar phenomena are also observed in other languages than Japanese as a result of language contact with English. For instance, Renner and Fernández-Domínguez (2015) discuss the influence of English on Romance languages under the notion of False Anglicisms. Contact-induced expressions like "wasee eigo" are usually taken as wrong expressions in English. Irwin (2011) gives the compound *goo-sutoppu* 'traffic light' as an example of "wasee eigo," which is composed of two independent English loanwords *goo* 'go' and *sutoppu* 'stop.' He takes it as a "semantically remodeled" compound, since a native speaker of English would interpret a hypothetical English compound *go-stop* as a state of stopping and starting, or juddering. Under this view, "wasei eigo" can be characterized as expressions occurring as a result of a deviant use and interpretation of English by non-native speakers.

In this paper, focusing on contact-induced compounds of *goo-sutoppu* type, we will argue that they are never anything like semantically remodeled expressions of English, and that their existence is a matter of morphology, but not semantics. Based on the notions of matter (MAT) and pattern (PAT) and their distinction in borrowing (Matras & Sakel (2007)), and the morphological difference between English and Japanese, we will answer the question of why Japanese speakers, but not English speakers, use such compounds, and how they are produced.

2 Matter borrowing in compounding in "Japanese English"

Compounding is a process of creating a new complex word by merging at least two lexemes. And they are right-headed in most languages. Although the compounds in (1) are not acceptable among native-speakers of English, it can be said that they satisfy the requirements for compounds. The loanwords in the compound in (1a), *bakku* and *miraa,* are independently used as nominal elements in Japanese lexicon. *Bakku* means the area of something that is furthest from the front, paraphrased as *koohoo,* 'the back of something,' and *miraa* means a mirror. Also, (1a) is interpreted as a kind of mirror, showing its right-headed property. Likewise, the constituents *teeburu* and *supiichi* in (1b) have the independent lexemic meanings of 'table' and 'speech,' respectively, and the complex expression *teeburu-supiichi* conveys the meaning consistent with the right-headedness of compounds.

We point out here that the considerations so far suggest that MAT borrowing is involved in the compounding in (1). It is a borrowing process in which only a phonological material or a surface form of an expression is borrowed from one language to another, and no grammatical or semantic aspects are transferred (see Sakel (2007), for example). If they are replicated in a recipient language, a borrowing process is identified as PAT borrowing. In the case of (1), only the phonological information or the surface word form of English words *back*, *mirror*, *table* and *speech* are replicated in Japanese, and they are exploited in compounding in Japanese lexicon. The compounds in (1) are formed conforming to the frame of Japanese lexicon and Japanese word formation, with only phonological features and MAT borrowed from English for phonological realization. Since there is no PAT borrowing involved in (1), it is rather difficult for native speakers of English to find the compounds acceptable.

3   A problem

The MAT-PAT distinction enables us to give a straightforward explanation for the compounds in (1). They are formed through compounding with only MAT borrowed from English. So native speakers of English do not use them, while they are natural for native speakers of Japanese. Turning our attention to the compound *goo-sutoppu* which Irwin (2011) gives, the situation does not seem to be so simple. As he mentions, its meaning shared among Japanese speakers is 'traffic light.' While the right-headedness of compounds is easily identified in (1), based on their meanings, the right-headedness of *goo-sutoppu* does not seem to immediately follow from its meaning, and it is not determined at this point that it has a structure of compounds. The questions to be addressed are thus as follows:

(2) a. Is such an expression as *goo-stoppu* a compound?
     b. Why is there a difference in its acceptability between English and Japanese
         speakers?

4   Dvandvas in Japanese

We will pursue the possibility here that the complex expression of *goo-sutoppu* type is a coordinate compound, more specifically, a dvandva compound. Assume that dvandvas are headless. We can then take *goo-sutoppu* as a headless compound, if it is proved to be a member of dvandvas. Before a detailed discussion about whether it is a dvandva, let us summarize the background knowledge about dvandvas and their classification relevant to our discussion.

4.1 Theoretical background

Bauer (2008) gives five types of coordinated compounds, and one of them is called a dvandva. The typical example in Japanese is *eda-ha* (lit. branch-leaf) 'branches and leaves.' According to him, "the dvandva is understood as being a new unity made up of the whole of the two entities named. (Bauer (2008:2))" *Eda-ha* thus means the union of branches and leaves, each of which is referential.

Bauer further subdivides dvandvas into five types. Three of them, which are relevant to the discussion below, are listed in (3).

(3) a. Additive types
         eda-ha   (lit. branch-leaf)   'branches and leaves'   (Japanese)
     b. Co-hyponymic types
         bas-kaar   (lit. bus-car)   'vehicles'   (Punjabi)
     c. Co-Synonymic types
         đường sá   (lit. road street)   'roads'   (Vietnamese)

The additive type in (3a) is understood as the union or sum of the sets denoted by the constituents, as mentioned above. Note that the constituents with opposite meanings can also be combined to form this type of dvandvas. In Japanese, for example, the word *sa* 'left' can form the additive type with the opposite word *yuu* 'right,' as shown in *sa-yuu* (lit. left-right) 'a left side and a right side.' The constituents have the different meaning, but they can be regarded as a member of the same set. The branch and the leaf are both parts of a tree, and the notions of left and right are both related to the direction.

The co-hyponymic type in (3b) denotes more abstract concepts than what each constituent denotes. The constituents of the compound *bas-kaar* means 'bus' and 'car,' and they are instances of vehicles. Then the compound itself convey this more general meaning of 'vehicle.' It can be said that the meaning of the co-hyponymic type is extended as a result of conceptualization. Note that we can observe a similar phenomenon in the additive type as well. The additive type *eda-ha* in (3a), for example, can refer to 'trivial things.' Both the branch and the leaf are not the trunk and the main part of a tree, inducing the meaning extension.

The co-synonymic type in (3c) is a compound composed of the constituents with the identical meaning, contrasting with the additive type and the co-hyponymic type. Though the constituents are identical in meaning, this type has no emphatic or repetitive interpretation.

4.2 MAT borrowing for producing dvandvas

Remember that considering its headedness, the issue is whether *goo-sutoppu* is a compound or not and how it is interpreted as a traffic light. Our claim is that it is a dvandva compound of additive type, which broadens the interpretive possibility. There are five reasons to believe that this analysis is on the right track.

First, the constituents *goo* 'go' and *sutoppu* 'stop' are both notions related to movements or traffic affairs, and can be considered to be in the same set. This membership property of the constituents is needed for forming dvandvas, and the expression *goo-sutoppu* satisfies it.

Second, from the semantic point of view, its interpretation seems to involve conceptualization or metonymy. *Goo-sutoppu* does not simply mean going and stopping. But it refers to a signal, a device controlling a traffic condition, showing when we must go and when we must stop. It can be said that the meaning of a traffic light is equipped to this expression through conceptualization. This strongly suggests that it is a dvandva compound.

Third, there are so many examples in Japanese similar to *goo-sutoppu*, as shown in (4).

(4) a. *hamu-eggu*   (lit.) ham-egg   'ham and egg, bacon and egg'
    b. *appu-daun*   (lit.) up-down   'ups and downs'
    c. *macchi-pompu*   (lit.) match-pump   'a person who intentionally causes trouble and settles down'
    d. *T.P.O.*   (lit.) time-place-occasion   'taking the time, the place and the occasion into consideration'
    e. *D.P.E.*   (lit.) developing-printing-enlargement   'a photo shop'
    f. *inkamu-gein*   (lit.) income-gain   'capital gains'

The English counterparts of *Hamu-eggu* in (4a) and *appu-daun* in (4b) have the coordinate structure with the conjunction *and*, as shown in their translations. This indicates that the complex expressions without *and* in (4a) and (4b) are dvandva compounds. Simultaneously, this implies that they are produced through MAT borrowing, but not PAT borrowing. Only the phonological materials are borrowed, without the coordinate structure involving *and*. As for (4c) and (4e), meaning extensions are observed, as shown in their translations. (4f) is an example of a co-synonymic dvandva. Its constituents are identical in meaning. (4) thus

indicates that dvandvas like *goo-suttoppu* are rather productive. Other Languages can be a donor language. The examples in (5) are produced through MAT borrowing from French.

(5) a. *figu-noa*   (lit. fig-nut)   'bread mixed with figs and nuts'
    b. *furomaaju-figu* (lit. cheese-fig)   'bread mixed with cheese and figs'

These expressions are often found in bakery shops in Japan. *Figu, noa* and *furomaaju* in (5a) and (5b) come from the French words *figue* 'fig,' *noix* 'nut' and *fromage* 'cheese,' respectively. The phonological materials are borrowed from French, suggesting that dvandvas are naturally formed with foreign words in Japanese through a process of MAT borrowing.

Finally, our proposal accounts for why native speakers of English feel difficulties in interpreting the complex words of *goo-sutoppu* type. Bauer (2008) points out that the coordinated compounds to be labeled with the term "dvandva" are much more limited than generally assumed. Typological studies, such as Arcodia et al. (2010), point out that most of the European languages resist dvandva formation in principle. A morphological frame of dvandvas are not in English lexicon and its native speakers do not analyze them in Japanese.

5 Final remarks

We have shown that both headed-compounds like (1a) and (1b) and non-headed dvandva compounds like *goo-sutoppu* are formed in Japanese lexicon with MAT borrowing. Native speakers of English thus feel these expressions to be deviant.

Finally, I will make a brief remark on parametric variations in the occurrence of dvandvas. If dvandva formation is a system in the lexicon, they should be morphological in nature. Kastovsky (2006) suggests that word-basedness and stem-basedness are key notions to consider parametric variations of languages. If English is a word-based language and Japanese is a stem-based language, it is highly likely that dvandvas are licensed only in stem-based languages. Given that Kastovsky characterizes Old English as a stem-based language, dvandvas are expected to be found in OE. Interestingly, even our rough dictionary search has found many instances of co-synonymic type of dvandvas in OE. A few of them are listed in (6).

(6) a. *ellencræft*   'strength, power' (*ellen* 'strength, power' + *cræft* 'power, strength)
    b. *friþowaru* 'protection' (*friþu* 'protection' + *waru* 'protection')

Thorough research from this perspective thus seems to be necessary.

References

Arcodia, F. G., Grandi, N. and Wälchli, B. 2010. Coordination in compounding. In Scalise, S. and Vogel, I. (eds), *Cross-disciplinary issues in compounding*. Amsterdam: John Benjamins. 177-197.

Bauer, L. 2008. Dvandvas. *Word Structure* 1, 1-20.

Irwin, M. 2011. *Loanwords in Japanese,* Amsterdam: John Benjamins.

Kastovsky, D. 2006. Typological changes in derivational morphology. In van Kemenade, A. & Los, B. (eds), *The handbook of the history of English*. Oxford: Blackwell Publishing. 151-176.

Matras, Y. & Sakel, J. (eds.). 2007. Grammatical borrowing in cross-linguistic perspective. Berlin: Mouton de Gruyter.

Renner, V. & Fernández-Domínguez, J. 2015. False Anglicization in the Romance languages: A contrastive analysis of French, Spanish and Italian. In Cristiano, F. & Henrik, G. (eds), *Pseudo-English: Studies on False Anglicisms in Europe*. Berlin: De Gruyter. 147-157.

Sakel, J. 2007. Types of loan: Matter and pattern. In Y. Matras & Sakel, J. (eds.), Grammatical borrowing in cross-linguistic perspective, 15–29. Berlin: Mouton de Gruyter.

Shibatani, M. 1990. *The languages of Japan*. Cambridge: Cambridge UP.

Competition in the bilingual lexicon and cross-language priming asymmetries: A morphological connection?

Madeleine Voga

Univ. Paul-Valéry Montpellier III

EA LLACS-ReSO

Morphological processing and the variables influencing it have been studied through several experimental protocols, among which the masked priming technique has a special status. Although the technique is quite sensitive to orthographic factors (Forster, Mohan & Hector, 2003), which is normal in the extent that it examines visual processing, it has a great potential, with the adequate design, to explore the various kinds of relations between morphologically simple or complex lexical units (LU henceforth). Given that LU share several of their characteristics (orthographic, phonological, semantic, morphological) in both dimensions, syntagmatic and paradigmatic, a model of the bilingual lexicon has to take into account not only the processes but also the way in which these intertwining characteristics influence representation and processing.

Early studies on morphological processing with the masked priming technique, for ex. Grainger, Colé & Segui (1991) demonstrated that orthographic similarity of the prime inhibits lexical access of morphologically complex targets, despite (or because of) the absence of any morphological relation between them, e.g. the prime *mûrir* 'ripen' inhibits the target mural 'wall'. This inhibition reaches 27ms when LU share their initial letters and is accounted for in terms of "preactivation of lexical representations during the processing of the prime, which interferes with the processing of the target" (p. 380). The complex flow of activation/inhibition taking place in the lexicon during processing is central for Interactive Activation models (McClelland & Rumelhart, 1981) and the above observation has led to models such as the supra-lexical model of morphology (Giraudo & Grainger, 2001).

Other variables related to interference and competition have been documented in the literature, for instance Coltheart's N (Coltheart, Davelaar, Jonasson, & Besner, 1977), which refers to the number and relative frequency of neighbours, i.e., words differing by a single letter (e.g., banish and vanish). For a model such as the SOLAR model (Self-Organizing Lexical Acquisition and Recognition, Davis, 1999), lexical competition orientates the acquisition. As noticed by Bowers, Davis & Hanley (2005), it is important to have a psychologically accurate definition of what is a neighbour and considering as such only words of the same length that differ by one letter (Coltheart's N) is based on simplicity rather than perceptual similarity.

As far as monolingual morphological processing is concerned, the role and the influence of competition is illustrated in several experiments, for instance, Voga & Giraudo (2017) report evidence in favor of the variable called 'pseudo-family size' in French. This variable reflects the number of words that function as antagonists during processing, e.g. the prime *portons* 'we carry', will potentially activate (at least) all words that share its initial letters (i.e. the letters of the stem) at the lexical level. These 'pseudo-relatives' (e.g., *portail* 'portal', *porte* 'door', *port* 'harbour', *portier* 'porter', *portion* 'portion', *portique* 'porch', *portrait* 'portrait', *portière* 'door', *postons* 'we mail') interfere with the representation of the target, delaying thus its recognition. The above variable which is, in a certain way, the opposite of the MFS (Morphological family Size, de Jong, Schreuder & Baayen, 2000) highlights the fact the complex flow of activation and inhibition giving rise (or not) to morphological priming effects, does not exclusively depend on what happens inside the word (its morphemes), but also beyond the-word-to-be

identified, in its lexical environment, i.e., other LUs exhibiting different kinds of similarities with the word under study.

The aim of this presentation is to highlight the fact that lexical interference and competition (Hoffmann & Jacobs, 2014) also take place in the bilingual lexicon, influencing the pattern of morphological effects. To put it in terms of language co-activation, the lexicon is shown to be sensitive to the 'larger chain of morphological relations' (Mulder, Dijkstra, Schreuder & Baayen, 2014), and this influence may be positive (inducing facilitation effects) or negative (inducing inhibition effects and/or preventing the occurrence of positive effects).

Two sets of bilingual evidence will be presented, both using cross-language cross-script priming (in both directions, L1 to L2 and L2 to L1) with Greek - French bilinguals who had Greek as L1, had formal education on French and had lived at least 6 months in France. In the first experiment, four categories of materials were tested in a lexical decision task (108 words and 108 pseudowords), in which the duration of the prime (48ms SOA) prevented the participant from consciously perceiving the prime. These categories were the following:

a) 27 pairs of cognates such as κατακλυσμός /kataklysmós/ 'cataclysm' - *cataclysme* 'cataclysm', known to induce robust cross-language cross-script effects;

b) 27 prefixed non-cognate pairs such as συνέπεια /sinépia/ 'consequence' - *conséquence* 'consequence'. They had the same meaning as LUs but did not overlap on form, and their base wasn't a word in both languages. In this category of stimuli, the segmentation into morphemes was difficult, given that the 'base' of (at least) one member of the non-cognate pair does not mean anything, e.g. the segment -έπεια /épia/ in συνέπεια /sinépia/ 'consequence'. These words shared nevertheless a prefix with a common meaning, ex. συν- /syn/ - *con/com*, από- /apó/ - *ex*, υπό- /ypó/ - *sous/sup*;

c) 27 prefixed non-cognate pairs, sharing the same meaning at the level of the whole unit. These pairs had synonymic prefixes (as in b), as well as the same proportion of formal saliency of the prefix as in (b). In this category, the base corresponds to a word in both languages, ex. διάθεση /diáthesi/ 'disposition' – *disposition*, in which the base θέση /thesi/ 'position' is an existing LU. Morphological segmentation was thus possible here;

d) 27 non-cognate pairs without prefix, such as λογαριασμός /logariasmós/ 'bill' - *facture* 'bill'. This group corresponds to the classic category of non-cognates, tested in several studies (e.g., Dimitropoulou, Duñabetia & Carreiras, 2011).

The above stimuli were primed by three conditions of priming: translation in the other language (i.e., if the prime was in Greek the target was in French); morphological (mostly adjectives and adverbs) and the unrelated condition. They were tested in both directions of priming, L1 to L2 priming direction and the opposite. Results (RTs) for the L1-to-L2 direction, show that translation priming occurs for all three categories, excepted for non-cognate with word-base (c). This first result suggests that stimuli of category (c) triggered competition, probably related to the big morphological families they activate in the other language, e.g., θέση /thesi/ 'position' and θέτω /theto/ 'to posit' belong to very broad morphological families, not necessarily transparent. Moreover, the first two types of materials (a and b) induce translation (57 and 29ms respectively) and morphological effects (50 and 34ms) which are concomitant and of equivalent amplitude. In other words, no difference is found between cognates overlapping in form and non-cognates, which do not have similar form, but share a common morphological structure and a non-word base. This result is interpreted in the direction of shared prefix units (synonymic prefixes such as δια- /dia/ and *trans-*) which manage, despite the lack of any formal overlap, to activate each-other cross-linguistically, from the L1 to the L2 of the participant. This type of organisation corroborates Bybee's (1988; 2010) position following which the languages of the bilingual are organised morphologically,

in terms of morphological clusters. In such an organisation, there is no reason why not supposing connections between synonymic prefixes such as those considered here. Note however, that the positive influence of shared meaning (at the level of the prefix or of the whole LU) manages to emerge as long as there is no competition from other LUs, such as the one exerted on the stimuli of category (c) seen above. The results of the L2 to L1 direction of this first set of experiments, also show evidence in favour of competition. The statistical analysis shows significant inhibition (-23ms) for categories (c) and (d) taken as a group, versus cat. (a) and (b), which is clear evidence in favour of competition in the L2 to L1 direction of priming.

This last result is especially important because it relates to the well-documented and hotly debated, especially for cross-script bilinguals, asymmetry between the two priming directions, an asymmetry which is nevertheless not always found. While it is found in some studies (e.g., for Chinese-English: Jiang & Forster 2001; Chen, Zhou, Gao & Dunlap 2014; Allen, Conklin & van Heuven 2015, with Japanese-English cognates) it is not found in others (e.g., Duyck & Warlop, 2009, for Dutch-French non-cognates).

In some other studies (Voga, 2014), an asymmetry is found for one type of stimuli but not for the other. The second set of experiments compares (in the same experiment and with the same participants), two types of Greek - French cognates: etymologically French (Latin), e.g., ρόλος /rólos/ 'role' - *rôle* or κουζίνα /kuzína/ 'kitchen' - *cuisine*, and etymologically Greek ones, e.g., ιδέα /idea/ 'idea'- *idée*. The etym. French cognates, while they manage to prime in the L1 to L2 direction (34ms for cognate and 28ms for morphological priming), fail to induce any significant effect in the opposite direction. This pattern is quite different to that of etym. Greek cognates, which manage to induce robust translation and morphological priming in both directions (e.g. translation priming effect: 24ms in the L2 to L1 direction and 56ms in the 'easier' direction, i.e., L1 to L2). To sum up, the etymologically L2 cognates, contrary to their L1 counterparts, confirm the asymmetrical pattern between the two priming directions and behave similarly to non-cognates (Dimitropoulou et al., 2011, with Greek learners of Spanish).

The asymmetry-between-the-languages-of-the-bilingual issue is to be put in relation to competition between LU. Moreover, as the above data clearly show, competition between LUs contained in the 'unified lexico-semantic architecture' (Schoonbaert, Duyck, Brysbaert & Hartsuiker, 2009) that forms the bilingual lexicon, depends on several intertwining factors, springing from different parts of words: a) the influence of the base, e.g., the role of 'base-word lexicality' (Exp. 1); b) the influence of the morphological family of the base, e.g., the 'secondary morphological family size' (Mulder et al., 2014). This type of variable also relates to etymology, whose influence was demonstrated in Exp. 2; c) the influence of the prefix, both as a structural and perceptual unit (denoting morphological complexity, at least when it is salient) and as a meaning unit, given that prefixes' salience is related to the multiplicity of their different senses (Voga, Nikolaou & Anastassiadis-Symeonidis, 2018, for Greek L2 data). The prefix's semantic instruction can be obscured by various factors, e.g., by the application of metaphoric or metonymic semantic rules on the base-word before that of Word Construction Rules (Corbin, 1987/1991).

The above results and suggestions highlight the fact that morphological processing does not exclusively depend on what happens inside the word, but also beyond the-word-to-be identified, in its lexical and morphological environment. This environment corresponds to a dynamic network based on the 'larger chain of morphological relations', compatible with facts such as productivity, lexical interference and competition, as well as facts related to discrimination (Milin, Feldman, Ramscar, Hendrix & Baayen, 2017) and categorisation

(Bybee, 2010) processes. This lexical environment does not only exert positive influence but can also lead to inhibition. Visualising the complex sum of activation/inhibition in terms of a unique, purely bottom-up linear process whose function would consist of decomposing the word-to-be-identified into pieces independently of words' characteristics seems insufficient.

References

Allen D., Conklin K. & van Heuven W.J.B. 2015. Making sense of the Sense Model: Translation priming with Japanese-English bilinguals. *The Mental Lexicon* 10. 32-55.

Bowers, J.S., Davis, C.J., & Hanley, D. 2005. Is there a "hat" in "that"? *Journal of Memory and Language* 52. 131–143.

Bybee, J. 1988. Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical Morphology. Approaches to modern linguistics*, 119–142. San Diego: Academic Press.

Bybee, J. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Chen, B. Zhou H. Gao Y. & Dunlap S. 2014. Cross-language translation priming asymmetry: a test of the Sense model. *Journal of Psycholinguistic Research* 433. 225–40.

Colthheart, M., Davelaar, E., Jonasson, J. T., Besner, D. 1977. Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*, 535-555), New York: Academic Press.

Corbin, D. (1987/1991). *Morphologie dérivationnelle et structuration du lexique* (2 vols.) Tübingen / Villeneuve d'Ascq : Max Niemeyer Verlag / Presses Universitaires de Lille.

De Jong, N. & R. Schreuder & R. H. Baayen. 2000. The morphological size effect and morphology. *Language and Cognitive Processes* 15. 329–365.

Davis C. J. 1999. *The Self-organizing lexical acquisition and recognition (SOLAR) model of visual word recognition*. Unpublished Doctoral dissertation, University of New South Wales.

Dimitropoulou, M. Duñabetia J.A. & Carreiras M. 2011. Masked translation priming effects with low proficient bilinguals. *Memory & Cognition* 392. 260–275.

Duyck, W. & Warlop N. 2009. Translation priming between the native language and a second language: new evidence from Dutch French bilinguals. *Experimental Psychology* 563. 173–9.

Forster, K. I. Mohan K. & Hector J. 2003. The mechanics of masked priming. In S. Kinoshita & S. J. Lupker (eds.), *Masked priming: State of the art*, 3–37. Hove: Psychology Press.

Giraudo, H., & Grainger, J. 2001. Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review* 8(1). 127–131.

Grainger, J., Colé, P., & Segui, J. 1991. Masked morphological priming in visual word recognition. *Journal of Memory and Language* 30. 370–384.

Hofmann, M. J. & A. M. Jacobs. 2014. Interactive activation and competition models and semantic context: from behavioral to brain data. *Neuroscience and Biobehavioral Reviews 46*. 85–104.

Jiang, N., & Forster, K. I. 2001. Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language* 44. 32–51.

McClelland, J. L. & Rumelhart D.E. 1981. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 88. 375–407.

Milin, P., Feldman, L.B, Ramscar, M., Hendrix, P. & Baayen, H. 2017. Discrimination in lexical decision. *PLoS ONE* 12(2):e0171935.

Mulder, K. Dijkstra T. Schreuder R. & Baayen R.H. 2014. Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language* 72. 59–84.

Schoonbaert, S., Duyck, W., Brysbaert, M. & Hartsuiker, R. 2009. Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition* 37(5). 569–586.

Voga, M. & Giraudo, H. 2017. Word and beyond-word issues in morphological processing. *Word Structure* 10 (2). 235-254.

Voga, M., Nikolaou, G. & Anastassiadis-Symeonidis A. 2018. Prefix conceptual salience in L2 acquisition and processing: a reading-time study within context. In J. Audring, N. Koutsoukos, F. Masini, & W. Sandler (eds.) *Online Proceedings of the 11th Mediterranean Morphology Meeting* (MMM11), 81-92. Nicosia: University of Cyprus. http://mmm.lis.upatras.gr/index.php/mmm/article/view/2875/3167

Voga, M. 2014. Διαγλωσσική ομοτυπία και μεταφραστική προτεραιοποίηση: ο ρόλος του ορθογραφικού δείκτη και της ετυμολογίας μέσα από δύο ελληνο-γαλλικά πειράματα [Cognateness and masked priming: the role of the orthographic cue and etymology through two Greek - French experiments]. *Studies in Greek Linguistics 34*. 104-125.

Poster session A:
Wed 24 September 15:00–16:00

Deux propositions pour la description sémantique des noms de systèmes d'idées en -ISME

Agabalian Grigory

Université Paris-Descartes

Nous appelons « noms de systèmes d'idées » (NSI) les noms de doctrines, de religions et de théories. L'objectif de notre présentation est de mettre en évidence deux particularités sémantiques chez les NSI construits en -ISME :

1) Leur sens est un discours et devrait être décrit comme tel.

2) Leur description sémantique demande qu'on distingue base et expression linguistique de ce qui est valorisé.

Nous partirons de l'analyse de Roché (2011) qui distingue trois patrons de construction pour les noms en -ISME dont un qui sert à décrire les NSI. Ce patron construit un sens qui consiste en la valorisation (au nom du vrai, du bien, du beau, de ce qui est important) de ce que représente la base. POPULISME, par exemple, est construit selon ce patron et signifie 'valorisation du peuple'. D'après Roché, la base de ce patron est toujours un nom, même lorsque la forme radicale est celle d'un adjectif. Ainsi, la base d'ÉGALITARISME est le nom ÉGALITÉ et non l'adjectif ÉGALITAIRE[1].

1 La description sémantique des NSI comme noms de discours

Dans un premier temps, nous situerons les NSI au sein d'une distinction entre deux types de valorisations. Puis, nous exposerons notre approche pour les décrire comme noms de discours.

1.1 Deux types de valorisations

Nous distinguerons deux types de sens valorisationnels que la suffixation en -ISME peut construire :

1) Valorisation modale : la valorisation est en forme[2] d'acte d'énonciation, par exemple une affirmation (que la base représente ce qui est vrai) ou une réclamation (de voir advenir ce que représente la base). Les ismes de valorisation modale réussissent plusieurs tests. Ils acceptent d'être les sujets d'un verbe locutoire (*L'évolutionnisme affirme que* p) ainsi que de compléter des syntagmes prépositionnels comme *Selon le N* ou *Pour le N* qui fonctionnent comme des adverbes d'opinion et servent à introduire des discours rapportés (*Selon l'évolutionnisme,* p). Ces ismes mettent en lumière un fait remarquable quant à la relation langue/discours. Habituellement, lorsque les linguistes s'intéressent à cette relation c'est pour observer comment la langue est mise en discours. Avec les ismes de valorisation modale, nous sommes en présence de la situation exactement inverse : ce sont des discours qui ont été mis en langue ou, plus exactement, des discours entiers qui ont été transformés en un lexème grâce à un procédé morphologique. En effet, ÉVOLUTIONNISME, NOMINALISME ou FÉMINISME ne sont ni plus ni moins que des affirmations ou des

---

[1] Pour l'analyse des ismes (formation et/ou sens) voir : en français Dubois (1962), Corbin (1988), Dubois & Dubois-Charlier (1999), Roché (2011), Lignon, Namer & Villoing (2014) ; en anglais Marchand (1960), Plag (2003), Baeskow (2004), Rainer (2018) ; en espagnol Beniers Jacobs (1992), Santiago Lacuesta & Bustos Gisbert (1999), Muñoz Armijo (2010).

[2] « Forme » au sens saussurien du terme, c'est-à-dire ce qui est opposé à la substance.

réclamations devenues des noms. Il existe donc une opération morphologique (la suffixation en -ISME) qui permet de transformer des discours en lexèmes. Notons, au passage, que ces ismes de valorisation modale pourraient rejoindre la classe sémantique des noms d'idéalités de Flaux & Stosic (2015).

2) Valorisation attitudinale : la valorisation consiste en une évaluation de tout ce qu'on peut observer chez un humain et qui constitue son attitude. C'est, par exemple, le cas du sens construit d'ÉGOÏSME ('valorisation de moi-même' (égo-)). Les ismes de valorisation attitudinale ne réussissent pas les mêmes tests que les ismes de valorisation modale (*L'égoïsme affirme que* p / *Selon l'égoïsme,* p) tout simplement parce que les valorisations dont ils sont les noms ne sont pas en forme d'actes d'énonciation. En revanche, ils réussissent d'autres tests que, cette fois, les ismes de valorisation modale ne réussissent pas (ou difficilement) comme celui avec *Faire preuve de* N (*Faire preuve d'égoïsme* / ??*Faire preuve d'évolutionnisme*). Sur le plan de la classification sémantique, les ismes de valorisation attitudinale pourraient rejoindre la classe des noms de qualité (Flaux & Van de Velde 2000) ou des noms d'attitude (Anscombre 1995).

Les NSI relèvent évidemment d'une valorisation modale. Indiquons, toutefois, que certains ismes cumulent les deux types de valorisations, ce qui se traduit généralement par une polysémie doctrine/attitude (INDIVIDUALISME, MATÉRIALISME, RATIONALISME, etc.).

1.2 Interprétation des NSI dans des contextes dialogaux reconstitués

Sachant que pour les NSI la valorisation prend la forme d'un acte d'énonciation, le sens construit de ces noms exige d'être situé dans au moins l'un de ces deux contextes dialogaux :

1) un contexte de désaccord lorsque l'acte d'énonciation consiste en l'énonciation de la certitude que la base représente ce qui est vrai (ÉVOLUTIONNISME, NOMINALISME) ;

2) un contexte que nous avons qualifié « de rupture » lorsque l'acte d'énonciation consiste en l'énonciation de la volonté de voir advenir ce que représente la base (ABOLITIONNISME, LIBÉRALISME).

Nous présenterons uniquement le cas du contexte de désaccord. Celui-ci est un contexte reconstitué comprenant trois variables pragmatiques qui s'enchaînent de manière à former un dialogue : 1) une assertion initiale qui ouvre le dialogue, 2) une réfutation qui répond à cette assertion initiale et 3) une rectification de l'assertion initiale. Décrire le sens valorisationnel d'un NSI comme NOMINALISME équivaut à saturer ces trois variables de manière à situer son sens dans un contexte de désaccord reconstitué :

N°	Variables : désaccord	Saturations possibles : exemples pour NOMINALISME
1	assertion initiale	L-E_1/E_1[3] affirme que les idées générales sont des réalités
2	réfutation de 1	L-E_2/E_2 affirme que les idées générales ne sont pas des réalités
3	rectification de 1	L-E_2/E_2 affirme que les idées générales sont des noms

Ce contexte de désaccord montre que le sens de NOMINALISME n'est pas juste une valorisation de ce que représente sa base. Son sens est, en fait, une rectification située dans un dialogue.

Nous défendrons donc que les ismes modaux, tels que NOMINALISME, doivent leur sens valorisationnel à un contexte de désaccord qui est leur raison d'être en tant que NSI. Ce contexte montre également que leur sens doit être décrit au sein d'un interdiscours

---

[3] L signifie locuteur et E énonciateur.

reconstitué dans la mesure où l'on parle d'un sens qui est, ici, la rectification d'une assertion initiale.

2 La base et l'objet valorisé

Chez les NSI ce qui est valorisé et ce dont la base est le nom ne coïncident pas toujours. Dans notre modèle, nous appelons « objet valorisé » non pas l'objet mondain qui est valorisé mais son expression linguistique (sous la forme d'un syntagme ou d'une proposition).

2.1 La base ne représente pas l'objet valorisé mais l'agent qui valorise

De manière générale, le sens valorisationnel dont parle Roché (et qu'il qualifie d'« *axiologique* ») est une structure actancielle à deux actants : {agent *valorise* objet}. On peut penser que la base est censée toujours représenter l'objet (ce qui est valorisé). C'est faux. Elle peut aussi représenter l'agent qui valorise. En fait, on trouve deux configurations actancielles dans le sens construit des NSI :

1) {agent = BASE *valorise* objet = ?} pour les NSI dont la base est un anthroponyme désignant une source intellectuelle (MARXISME) ou un nom d'adepte (PROTESTANTISME). Dans un tel cas, l'objet valorisé est totalement imprédictible (?). Il n'existe pas morphologiquement.

2) {agent = ? *valorise* objet = BASE} pour les autres NSI (ABOLITIONNISME, AUTONOMISME, FÉMINISME, etc.).

Lorsque la base est le nom de l'agent qui valorise, elle n'est, de fait, plus du tout le nom de ce qui est valorisé (l'objet). Pour pouvoir encore parler de valorisation de ce que représente la base dans un tel cas, il faut obligatoirement distinguer l'objet valorisé de la base.

2.2 La base représente l'objet valorisé de manière incomplète

Par exemple, NÉGATIONNISME ne signifie pas 'valorisation de la négation' mais 'valorisation de la négation des chambres à gaz / du génocide des Juifs', c'est-à-dire la valorisation de quelque chose qui est la base ('négation') plus des modificateurs ('des chambres à gaz / du génocide des Juifs') qui spécifient cette base. Or, l'ensemble 'négation des chambres à gaz / du génocide des Juifs' n'est pas l'un des sens du lexème NÉGATION. Il y a donc spécification de la base mais pas sélection de l'un des sens d'une base polysémique (comme avec HUMAIN dans INHUMAIN par exemple). Ce qui est valorisé ne peut pas correspondre à l'un des sens d'une base polysémique et donc la base ne suffit pas, à elle seule, à connaître ce qui est valorisé. Ce qui est valorisé est, en fait, une idée complexe dont l'expression linguistique peut prendre la forme d'une proposition ou d'un syntagme multinominal. La base ne constitue qu'une petite partie de cette proposition ou de ce syntagme. Le reste est complètement imprédictible et ne relève pas de l'un de ses sens. Voilà donc une deuxième raison, moins forte que la première, qui peut inciter à distinguer l'objet valorisé de la base. Naturellement, c'est là une proposition qui intéresse surtout le travail du sémanticien (pas tant celui du morphologue dans la mesure où l'identification de la base n'est pas concernée).

3 Conclusion

Nous avons essayé de montrer que les NSI sont des noms de discours (affirmations, réclamations, préconisations…) et, en tant que tels, leur sens ne peut pas être décrit de manière adéquate si on ne le situe pas dans un interdiscours où il est, notamment, la

rectification d'une assertion initiale. Plus généralement, cette description cherche à montrer qu'avec les NSI la relation habituelle entre la langue et le discours est inversée : ils sont des discours mis en langue. Dans un second temps, nous avons essayé de montrer que l'identification de la base ne suffit pas toujours, pour le sémanticien, à connaître l'objet valorisé parce que cette base peut être le nom de l'agent qui valorise ou parce qu'elle peut nécessiter d'être complétée par des arguments.

Références

Anscombre Jean-Claude. 1995. Morphologie et représentation événementielle : le cas des noms de sentiment et d'attitude. *Langue française* 105. 40-54.

Baeskow Heike. 2004. *Lexical properties of selected non-native morphemes of English*. Tübingen: Gunter Narr Verlag.

Beniers Jacobs Elisabeth. 1992. El sufijo *-ismo* en el español de México. In Elizabeth Luna Traill (dir.), *Scripta Philologica : in honorem Juan M. Lope Blanch* (Vol. II), 313- 337. Mexico: Universidad Nacional Autónoma de México.

Corbin Danielle. 1988. Une hypothèse à propos des suffixes *-isme, -ique, -iste* du français : la troncation réciproque. In Ronald Landheer (éd.), *Aspects de la linguistique française. Hommage à Q.I.M Mok*, 63-74. Amsterdam: Rodopi.

Dubois Jean. 1962. *Étude sur la dérivation suffixale en français moderne et contemporain. Essai d'interprétation des mouvements observés dans le domaine de la morphologie des mots construits*. Paris: Larousse.

Dubois Jean & Dubois-Charlier Françoise. 1999. *La dérivation suffixale en français*. Paris: Nathan.

Flaux Nelly & Stosic Dejan. 2015. Pour une classe des noms d'idéalités. *Langue française* 185. 43-57.

Flaux Nelly & Van de Velde Danièle. 2000. *Les noms en français : esquisse d'un classement*. Paris: Ophrys.

Lignon Stéphanie, Namer Fiammetta & Villoing Florence. 2014. De l'agglutination à la triangulation ou comment expliquer certaines séries morphologiques. In Frank Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost (éds.), *Actes du 4e Congrès Mondial de Linguistique Française*, 1813-1835. Paris: Institut de Linguistique Française.

Marchand Hans. 1960. *The categories and types of present-day English word-formation*. Wiesbaden: O. Harrassowitz.

Muñoz Armijo Laura. 2010. *La historia de los derivados en -ISMO e -ISTA en el español moderno*. Thèse de doctorat. Universidad Autónoma de Barcelona.

Plag Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Rainer Franz. 2018. Word formation and word history: the case of CAPITALIST and CAPITALISM. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (éds.), *The lexeme in descriptive and theoretical morphology*, 43-65. Berlin: Language Science Press.

Roché Michel. 2011. Quel traitement unifié pour les dérivations en *-isme* et en *-iste* ?. In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (éds.). *Des unités morphologiques au lexique*, 69-143. Paris: Hermes Science-Lavoisier.

Santiago Lacuesta Ramón & Bustos Gisbert Eugenio. 1999. La derivación nominal. In Ignacio Bosque & Violeta Demonta (éds.), *Gramática descriptiva de la lengua española* (vol. III). Madrid: Espasa Calpe. 4505-4594.

Revisiting inflectional morphology: Towards a new paradigm for teaching nominal inflection in Modern Greek as a second language

Anna Anastassiadis-Symeonidis
Aristotle University of Thessaloniki

Maria Mitsiaki
Democritus University of Thrace

1   Abstract

This paper seeks to resituate Modern Greek (MG) nominal inflection within a broader continuum of theoretical and applied linguistic analysis. The goal of our work is thus twofold. Firstly, we discuss the morphological structure of the MG nominal system and we propose an analysis which provides a formal account of inflected noun formation within the Paradigm Function Morphology framework (PFM, Stump 2001). Secondly, we emphasise the need for a simple and economical description of nominal inflection in L2 teaching material, one which goes beyond the conventional analyses, by taking into consideration the very principles of Paradigm Functions and Information Theory (Ackerman & Alouf 2013).

2   Introduction

The Greek nominal system comprises the three grammatical categories of gender, case, and number. It distinguishes three genders (MASC, FEM, and NEU), four cases (NOM, GEN, ACC, and VOC), and two numbers (SG and PL, Stephany & Christofidou 2009). Most of the existent proposals to the description of the MG nominal inflectional system are (a) gender-driven (following Triantafyllidis 1941, Table 1), and (b) morphologically-driven, based on the distribution of inflectional endings (among others Mackridge 1985, Ralli 2003, Klairis & Babiniotis 2005).

INFLECTIONAL CLASS A	INFLECTIONAL CLASS B	INFLECTIONAL CLASS C[1]
Masculine nouns in:	Feminine nouns in:	Neutral nouns in:
-os: *anθropos* 'human'	*-a*: *θalasa* 'sea', *mama* 'mum'	*-o*: *vutiro* 'butter'
-is: *piitis* 'poet', *manavis* 'greengrocer', *prezvis* 'ambassador'	*-i*: *anaɟi* 'need', *poli* 'city'	*-i*: *peði* 'child'
	-os: *psifos* 'vote'	*-ma*: *cima* 'wave'
-as: *filakas* 'guard', *tomeas* 'sector', *psaras* 'fisherman'	*-u*: *alepu* 'fox'	*-os*: *kratos* 'state'
-es: *kafes* 'coffee'		*-as*: *peras* 'end'
-us: *papus* 'grandfather'		*-i* (<υ>): *oksi* 'acid'

Table 1. Gender-driven analysis of the MG nominal inflectional system (Triantafyllides 1941)

Ralli's morphologically-driven proposal for a division of MG nouns into eight inflectional classes (IC) is a very popular one among linguists. Such a division is based on (a) the

---

[1] More than one representative nouns of the three ICs are given, so as to cover various inflectional paradigms of (im)parisyllabicity.

presence of a systematic allomorphic variation of the stem, and (b) the form of the whole set of inflectional endings that are combined with the stems (2003: 86, see Table 2).

The previous inflectional modeling leaves us with the inflectional endings portrayed in Table 3 (see also Ackerman & Malouf 2013).

IC	
IC 1	*anθropos, psifos*
IC 2	*filakas, tomeas, psaras, piitis, manavis, prezvis, kafes, papus*
IC 3	*θalasa, mama, anaɟi, alepu*
IC 4	*poli*
IC 5	*vutiro*
IC 6	*peði*
IC 7	*kratos*
IC 8	*cima*

Table 2. MG ICs proposed by Ralli

IC	NOM-SG	ACC-SG	VOC-SG	GEN-SG	NOM-PL	ACC-PL	VOC-PL	GEN-PL
1	-os	-∅	-e	-u	-i	-us	-i	-on
2	-s	-∅	-∅	-∅	-es	-es	-es	-on
3	-∅	-∅	-∅	-s	-es	-es	-es	-on
4	-∅	-∅	-∅	-s	-is	-is	-is	-on
5	-o	-o	-o	-u	-a	-a	-a	-on
6	-∅	-∅	-∅	-u	-a	-a	-a	-on
7	-os	-os	-os	-us	-i	-i	-i	-on
8	-∅	-∅	-∅	-os	-a	-a	-a	-on

Table 3. MG inflectional endings (Ralli, 2003)

However, both gender and morphologically-driven approaches do not reflect on issues of frequency and availability of inflectional suffixes, leaving aside the crucial matter of the inflectional system core-periphery continuum. Moreover, they do not offer a systematic and economical determination of where to put the stem/affix boundary, allowing for a needless abundance of inflectional suffixes, e.g. *filak-as, piit-is, anθrop-os* instead of *filaka-s, piiti-s, anθropo-s*. Such a lack of descriptive economy impinges on the theory's predictability.

Therefore, a challenging research question we have to address is whether the morphological gradience of MG inflectional system can be captured by the PFM framework.

3   An alternative model for MG nominal inflection

Our main concern is to propose an integrated theoretical model of MG nominal inflection, and at the same time a robust learning model, which facilitates speakers/learners to handle morphological complexity, i.e. to make accurate guesses about unknown forms of words, based on exposure to known forms (see also Ackerman & Malouf 2013).

To achieve our goal, we had to revisit the issue of the stem space. In order to test the stem formation process, we proceeded with the analysis of about 82.000 MG nouns lemmatized into the *Reverse Dictionary of Modern Greek* (Anastassiadis-Symeonidis 2002), using as a theoretical basis previous work by Anastassiadis-Symeonidis (2012). The stem formation, indexing and selection processes led to a variety of distinct stems (see Table 4).

Stem (S)	Mode A	Mode B	Mode C
S1	X + V(owel) <α/η/ε/ου/α>	X + *i* <ι>	X + *i* <η>
S2a	X (bare stem)	X	X
S2b	X + V + C(onsonant) ð/*t*	X + *j*	X + *e*

Table 4. MG noun stem formation

Where: X is a lexeme's stem. S2a is derived from S1, after removing the terminal vowel. S2b is derived from S1, by adding the consonant ð for masculine and feminine nouns, and the consonant *t* for neutral nouns. Mode A & B are the prototypical modes of stem formation. Mode B applies only to neutral nouns, subject to morphophonological processes due to palatalization. Mode C applies to instances of MG nouns, subject to rules of learned formation traced to diachrony (Ancient Greek), e.g. deverbal nouns in *-si/-ksi/-psi* or MASC-VOC-SG formation in *-e*, as it is found in the paradigms of a closed IC (IC3), without new members, handed down to MG from the Ancient Greek.

The data analysis yielded by our study provides strong evidence that MG nouns constitute 6 inflectional classes (IC1 to IC6), presupposing a minimal amount of segmentation into stems and exponents. An overall classification of MG noun stems and exponents is illustrated in Table 5 and 6:

IC	Stem	Examples	IC	Stem	Examples
IC1	S1	*filaka, piiti, tomea, psara, manavi, papu, kafe, prezvi*	IC4	S1	*vutiro, peði*
	S2a	*filak, piit tome, psar, manav, kaf*		S2a	*vutir, peð*
	S2b	*psarað, manavið, papuð, kafeð, prezve*		S2b	*peðj*
IC2	S1	*θalasa, anaɉi, mama, alepu, poli*	IC5	S1	*cima, peras, oksi*
	S2a	*θalas, anag/ɉ, mam, alep, pol*		S2a	*oks*
	S2b	*mamað, alepuð, pole*		S2b	*cimat, perat, okse*
IC3	S1	*anθropo*	IC6	S1	*kratos*
	S2a	*anθrop*		S2a	*krat*
	S2b	*anθrope*		S2b	

Table 5. PFM approach to MG ICs

	NOM-SG	ACC-SG	VOC-SG	GEN-SG	NOM-PL	ACC-PL	VOC-PL	GEN-PL	TYPE FREQ.
IC1	*s*				*es*	*es*	*es*	*on*	13000
IC2				*s/os*	*es*	*es*	*es*	*on*	33000
IC3	*s*		*o/e*	*u*	*i*	*us*	*i*	*on*	9000
IC4				*u*	*a*	*a*	*a*	*on*	20000
IC5				*os*	*a*	*a*	*a*	*on*	6000
IC6				*us*	*i*	*i*	*i*	*on*	200

Table 6. PFM approach to MG noun inflectional exponents

Such an analysis innovates in several ways:

1. It provides a unifying description of lexemes, inflected, derivative or compounds, since the stem space has been determined in a morphologically-sound mode.
2. It reveals the core-periphery inflectional classes, as some stems are more readily adopted than others, for instance stems in Mode A & B (Table 4) are prototypical compared to those in Mode C, S1 is prototypical compared to S2.
3. It contributes to the emergence of a prototypical and simple inflectional system, where syncretism reigns, i.e.: -*s* marker for MASC NOM SG and FEM GEN SG, -*on* for GEN PL, -*es* for MASC and FEM NOM/ACC/VOC PL, and -*a* for NEU NOM/ACC/VOC PL.
4. It is economical, since it allows for recognizable stems and a limited number of inflectional suffixes, without necessitating structural zeros or allomorphs.
5. It displays high predictability (Corbin 1987), e.g. [-learned]/low register suffixes attach to stems of the S2b type, especially for the highly-frequent classes IC1, IC2, IC4.

4 Towards a new model of nominal inflection for L2 teaching

Next, the PFM account of the MG nominal inflectional system is quantified in information-theoretic terms, by making use of INTEGRATIVE-COMPLEXITY, a metric that reveals to what extent morphological systems are organized in ways that allow them to be learned and used by native speakers (Ackerman & Malouf 2013). The researchers apply the measure of entropy to the nominal inflectional model provided by Ralli. In turn, we apply it to the PFM model, giving a more refined quantification, since we factor type frequency into our calculations. For instance, the probability of any lexeme belonging to any one class is not 1/6, as certain ICs appear to be highly-frequent (e.g. IC2 FEM and IC4 NEU, Table 6). Such measurements shed light into the possible inflectional class membership for unknown lexemes and the paradigm cell interpredictability (conditional entropy). It is obvious that, since our model allows for less ICs, easily recognizable stem blocks, and less exponents, entropy appears to be much reduced, and thus learnability is substantially higher.

A question that is raised is whether the previous probabilistic measures could be applicable also for L2 learners. They certainly work for Independent (Levels B1-B2) and Proficient Users (Levels C1-C2), but what about Basic Users who are not able to rely on previous language input? Our basic hypothesis is that low entropy morphological systems reflect patterned grammatical organization, which may be of great help to novice L2 learners.

As a next step we recorded and analyzed approaches to teaching inflection in ten MG second language textbooks targeting basic users. What comes as a conclusion is that even the newest textbooks exhibit a traditional gender-based approach, insensitive to paradigm generalizations and inflectional ending coreness/peripherality. Thus, we propose a staged instruction of MG nominal inflection to L2 learners, i.e. S1 as the most salient stem precedes in language teaching, S2a and S2b are explicitly derivable from S1 and they come next. Mode A & B are given instructional priority. The PFM robust system of lexeme realization ensures that stems, inflectional endings, rules of derivation and compounding fall under a unified and pedagogically sound description.

5 References

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The Low Conditional Entropy Conjecture. *Language* 89. 429–464.

Anastassiadis-Symeonidis, Anna. 2012. The SMG Inflectional System of Nouns and their Trends. In Zoe Gavriilidou, Angeliki Efthymiou, Evangelia Thomadaki & Penelope

Kampaki-Vougioukli (eds), *Proceedings of the 10th ICGL*, 23–40. Komotini: DUTh. [in Greek]

Anastassiadis-Symeonidis, Anna. 2002. *Reverse Dictionary of Modern Greek*. Institute of Modern Greek Studies (Manolis Triantafyllidis Foundation). [in Greek]

Corbin, Danielle. 1987. Morphologie dérivationnelle et structuration du lexique, 2 vol. Tübingen, Max Niemeyer Verlag, (1992) Villeneuve d'Ascq: Presses Universitaires de Lille.

Ralli, Angela. 2003. Morphology in Greek Linguistics: The State of the Art. *Journal of Greek Linguistics* 4. 77–129.

Stump, Gregory, T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: CUP.

Triantafyllidis, Manolis. 1941. *Modern Greek Grammar (of Demotic)*. Athens: Greek Government Schoolbook Press. 2nd edition (1978) Thessaloniki: Institute of Modern Greek Studies.

Auxiliary selection in Romance and inflectional classes

Xavier Bach

University of Oxford

Pavel Štichauer

Charles University, Prague

1   Introductory remarks

In this paper, we start out from the premise that alternating auxiliary selection, such as the well-known HABERE/ESSE (H/E) distinction in some Romance languages, could also be viewed in terms of partially arbitrary lexical stipulation, giving rise to two different inflectional classes (see, e.g., Bonami 2015: 97; Baerman, Brown & Corbett 2017: 28-29).

Such a view is relatively novel; it has only been touched upon in recent work (see, e.g., Bonami, 2015, Bach 2018, Štichauer 2019) where the general process of grammaticalization is invoked: auxiliary selection may become grammaticalized and may thus give rise to - more or less - arbitrary inflectional classes. This might have been a problematic view for various reasons. First, the two classes of lexemes selecting one or the other auxiliary are widely held to be well motivated (and, indeed, to a large extent it is the case in Romance, but see Sorace 2000, Bonami 2015), while the canonical inflection classes are entirely unmotivated (Corbett 2009). Second, auxiliary verb constructions were not considered to be genuinely inflectional. But recent work in inflectional periphrasis has clearly demonstrated that they are to be taken as exponence strategies (see, e.g. Ackerman & Stump 2004, Brown et al. 2012, Spencer & Popova 2015)

2   Mixed perfective auxiliation systems as IC

We intend to demonstrate that grammaticalization of auxiliary selection – giving rise to partially unmotivated inflectional classes – can be seen on the example of mixed auxiliary perfective systems, attested in some Italo-Romance varieties, where we find a more intricate alternation of the two auxiliaries H/E within one and the same TAM paradigm (see, e.g., Loporcaro 2007; 2014; 2016), as in examples (1) and (2) (cf. Torcolacci 2015: 52; Manzini & Savoia 2005: 682, respectively):

(1)

	SINGULAR	PLURAL
1	**sɔ f'fatt** I.am done.PTCP	**am ˈfatt** we.have done.PTCP
2	**a ˈfatt** you.have.SG done.PTCP	**avet ˈfatt** you.have.PL done.PTCP
3	**a f'fatt** he/she.has done.PTCP	**an ˈfatt** they.have done.PTCP

(2)

	SINGULAR	PLURAL
1	**sɔ durˈmito** I.am slept.PTCP	**semo durˈmito** we.are slept.PTCP
2	**si durˈmito** you.are.SG slept.PTCP	**sete durˈmito** you.are.PL slept.PTCP
3	**a durˈmito** he/she.has slept.PTCP	**a durˈmito** they.have slept.PTCP

In particular, we intend to discuss those interesting cases where one coherent class of lexemes, such as reflexives, is split between the two auxiliaries giving rise to *mixed paradigms which seem to be canonical instances of heteroclisis in a system where we have non-canonical inflectional classes* (see, e.g., Stump 2006, Kaye 2015, Bach 2018). But before, we lay out the notion of inflectional classes that we adopt.

3   Content *versus* form paradigms

We espouse Stump's (2016) distinction between content paradigms and form paradigms, positing that in the case of only one auxiliary, as in Spanish (and also in some Italian dialects), we have one content paradigm and one form paradigm for all perfective inflection (i.e. within the periphrastic screeve); in the case of two auxiliaries we would thus have two form paradigms (and, accordingly, two different realizations), as in the following scheme (inspired by Baerman, Brown & Corbett 2017: 77):

(3)

PRESENT PERFECT

	SG	PL
1		
2		
3		

	SG	PL
1	H	H
2	H	H
3	H	H

	SG	PL
1	E	E
2	E	E
3	E	E

	SG	PL
1	ho dormito	abbiamo dormito
2	hai dormito	avete dormito
3	ha dormito	hanno dormito

	SG	PL
1	sono partito/a	siamo partiti/e
2	sei partito/a	siete partiti/e
3	è partito/a	sono partiti/e
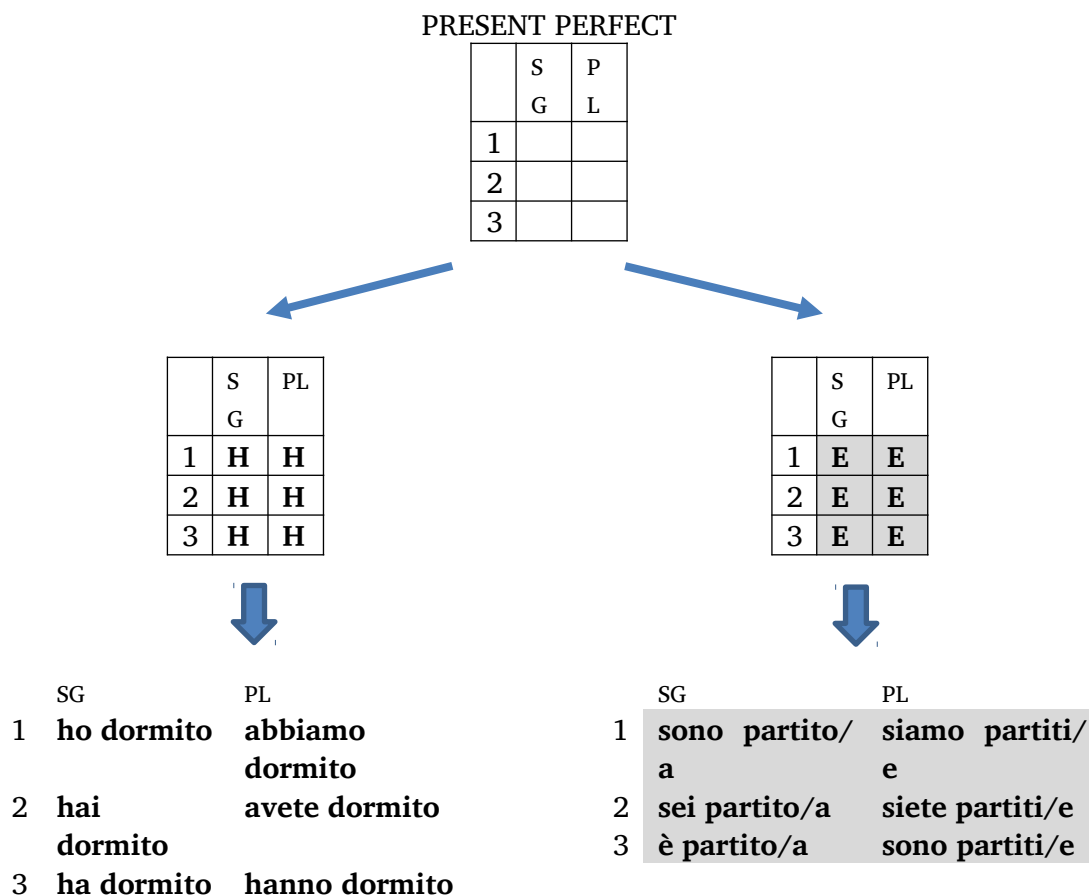
Of course, to adopt this view for the standard auxiliary selection in Romance is problematic since, as can be seen in (3), the lexemes selecting ESSE do not realize, strictly speaking, the same morphosyntactic property set (the gender and number of the subject is overtly expressed here, while this is not the case with HABERE). As we shall see, there are cases, within the mixed auxiliation systems, where exactly the same morphosyntactic set is realized regardless of the auxiliary selected (much as in example (2) above). In such cases, we do have two different form paradigms with distinct realizations corresponding to one content paradigm. It is best to treat instances such as (3) as a case of non-canonical inflectional class: one of the classes realizes an additional feature with respect to the other class. In synchrony, it would also be possible to treat such instances uniformly, with a system of syncretism; in such a system one would posit that the HABERE class also realizes the gender of the subject but that the realization for both genders is syncretic. In diachrony though, we know that it is not what happened and that the features involved were different.

However, we also find much more intricate cases where heteroclisis is at play involving unexpected mismatches that will turn out to be relevant for the content/form paradigm distinction. In many varieties we find a standard split between transitives/unergatives and unaccusatives selecting one or the other auxiliary (as in the scheme above), but within the class of reflexives a heteroclite system of periphrastic exponence can be found, as shown in example (4), reconstructed after Manzini & Savoia 2005, II, 652–653, and consulted with native speakers of the dialect:

(4)

	TRANSITIVES		*REFLEXIVES*		UNACCUSATIVES	
1 sg.	**aju laˈvatu**	⟸	*m aju laˈvatu*		sugnu viˈnutu/a	
2 sg.	**a laˈvatu**		*ti si llaˈvatu/a*	⟸	**si vviˈnutu/a**	
3 sg.	**a llaˈvatu**		*s ɛ llaˈvatu/a*	⟸	**ghɛ vviˈnutu/a**	
1 pl.	**amu laˈvatu**	⟸	*n amu laˈvatu*		simu viˈnuti/e	
2 pl.	**ati laˈvatu**	⟸	*v ati laˈvatu*		siti viˈnuti/e	
3 pl.	**anu laˈvatu**		*si su llaˈvati/e*	⟸	**su vviˈnuti/e**	

In this variety (Altomonte, reg. Calabria, southern Italy), transitives (*lavare* 'to wash') select H throughout the paradigm, and unaccusatives (*viniri* 'to come') consistently require E (as in Standard Italian or French). However, reflexives (*lavarsi* 'to wash oneself') are curiously split between the two auxiliaries and thus give rise to a third class of lexemes (reflexives), where the intraparadigmatic distribution of the two auxiliaries does not seem to be motivated.

Under the notion of inflectional classes that we are adopting, we could say that we have here two main IC, while the third class is heteroclite taking over part of the realisations from the two main classes. If we adopt Stump's (2016) approach, we could say that to one content paradigm (broadly defined as PRESENT PERFECT, see (3) above) correspond two different form paradigms (with H and E, respectively), as seen above in (3), and the third is a heteroclite class. Yet there is again the problem that we have just touched upon, namely the fact that the classes of periphrastic realizations do not correspond to only one content paradigm as the two paradigms realize different morphosyntactic property sets (with E, as already remarked, the gender and number of the subject are overtly marked).

But there is more to say about this particular situation. In fact, if we subscribe to this view, we face here an interesting and unexpected mismatch: part of the paradigm with H realise one property set, while the rest of the paradigm with E realise a different property set. This is at odds with how content paradigms are defined. In general, in fact, reflexives seem to be the locus of such mismatches in the Romance languages. In Occitan, reflexives seem to involve a mixture of the features of both auxiliaries, in that, in cases where reflexives have a direct object, they use an ESSE auxiliary but the past participle agrees in gender and number with the object, which is normally the agreement pattern of the HABERE auxiliary.

On the basis of the example of such mixed paradigms, as well as the Occitan example, we attempt not only to redefine the relationship between content and form paradigms, but also to put forward a more elaborate typology of inflectional classes induced by such intraparadigmatic auxiliary alternations.

References

Ackerman, Farrell & Gregory T. Stump. 2004. Paradigms and periphrastic expression. In Louisa Sadler & Andrew Spencer (eds.), *Projecting* Morphology, 111–157. Stanford: CSLI Publications.

Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2017. *Morphological Complexity.* Cambridge: Cambridge University Press

Bach, Xavier. 2018. *The Origins of Inflectional Classes*. Unpublished Ph.D. Thesis, University of Oxford.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25. 63–110.

Brown, Dunstan, Marina Chumakina, Greville G. Corbett, Gergana Popova & Andrew Spencer. 2012. Defining 'periphrasis': key notions. *Morphology* 22. 233–275.

Corbett, Greville G. 2009. Canonical inflectional classes. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected proceedings of the 6th Décembrettes*, 1–11. Somerville, MA: Cascadilla Proceedings Project.

Kaye, Steven J. 2015. *Conjugation class from Latin to Romance. Heteroclisis in diachrony and synchrony*. Unpublished PhD Thesis, University of Oxford.

Loporcaro, Michele. 2007. On triple auxiliation in Romance. *Linguistics* 45. 173–222.

Loporcaro, Michele. 2014. Perfective auxiliation in Italo-Romance. The complementarity of historical and modern cross-dialectal evidence. In Paola Benincà, Adam Ledgeway & Nigel Vincent (eds.), *Diachrony and Dialects. Grammatical Change in the Dialects of Italy*, 48-70. Oxford: Oxford University Press.

Loporcaro, Michele. 2016. Auxiliary selection and participial agreement. In Martin Maiden & Adam Ledgeway (eds.), *The Oxford Guide to the Romance Languages*, 802-818. Oxford: Oxford University Press.

Sorace, Antonella. 2000. Gradients in Auxiliary Selection with Intransitive Verbs. *Language* 76. 859–890.

Manzini, Maria R. & Leonardo M. Savoia. 2005. *I dialetti italiani e romanci: Morfosintassi generativa I-III*. Alessandria: Edizioni dell'Orso.

Spencer, Andrew & Gergana Popova. 2015. Periphrasis and Inflection. In Matthew Baerman (ed.), *The Oxford Handbook of Inflection*, 197–230. Oxford: Oxford University Press.

Stump, Gregory T. 2016. *Inflectional Paradigms. Content and Form at the Syntax-Morphology Interface*. Cambridge: Cambridge University Press.

Štichauer, Pavel. 2019. Mixed paradigms in Italo-Romance: a case of morphologization of auxiliary selection? In Silvio Cruschina, Adam Ledgeway & Eva-Maria Remberger (eds.), *Italian Dialectology at the Interfaces*, 79–100. Amsterdam: John Benjamins.

Torcolacci, Giuseppe. 2015. *Marking the Default. Auxiliary selection in Southern Italian dialects*. PhD Thesis, University of Leiden. Utrecht: LOT.

Distributivity as incremental plurality

Matthew Baerman,[1] Jérémy Pasquereau,[1] Carolyn O'Meara[2]

[1]Surrey Morphology Group, University of Surrey; [2]National Autonomous University of Mexico

1 Introduction

Verbs in Seri (isolate; Mexico) have up to five forms distinguishing different number values.[1] The individual exponents (suffixes and stem modifications) have no fixed value within the system, but appear to mark incremental degrees of plurality along a single scale, employing for the most part the same morphological means as found with nominal plurality (Baerman 2016). Previous studies (Marlett 1988, 2016, Cabredo, Pasquereau, and O'Meara 2018) have suggested that this paradigm is the realization of two cross-classifying features, subject number (SINGULAR vs PLURAL) and event number (NEUTRAL vs MULTIPLE), where MULTIPLE indicates that the event is extended over time, or iterated or distributed over subevents.

(1) a. *singular subject neutral*   b. *plural subject neutral*

      ih-yo-ohit   ha-yo-iit-oj

      1SG.TR-RLSYO-eat[2]   1PL-RLSYO-eat.PL-PL

      'I ate something.'   'We ate something.'

      c. *singular subject multiple*   d. *plural subject multiple*

      ih-yo-ohit-im   ha-jo-iit-olca

      1SG.TR-RLSYO-eat-MULT   1PL-RLSYO-eat.PL-PL

      'I ate something (over time).'   'We ate something (over time).'

However, recent fieldwork has shown that there are more distinctions than can be accounted for by this system, and we propose incorporating a third feature, distributivity. Distributive marking is well known from other languages of North America (Mithun 1999); what makes the Seri system noteworthy is that the category has no unique expression, but rather, it is conflated with subject number, where distributivity is manifested as one increment of plurality above the baseline. Unlike subject and event number, which are reliably marked, distributivity is only exceptionally distinguished, licensed by conditions which remain somewhat obscure.

---

[1] Other inflectional features – e.g. TAM and subject person – are prefixally marked and orthogonal to this paradigm.

[2] The gloss RLSYO stands for 'realis-YO', which is a tense-mood value (Marlett 2016).

2 Contexts for distributive marking

Deriving a transitive causative from an intransitive frequently licenses a distributive distinction: the subject plurality of the base verb in effect becomes object plurality of the derived transitive causative. For example, from the intransitive verb *-oiij* 'be tubular' (singular subject) ~ *-oiilc* 'be tubular' (plural subject), a causative can be formed in which the plural form indicates a plurality specifically of the object, provided the subject is singular (2). This we interpret as distributive.

(2) Gabriel quih **hesen** **pac** i-y-ah-oiilc.
Gabriel DEF ironwood INDEF.PL 3;3-RLSYO-CAUS-tubular.PL
'Gabriel made the ironwood logs into tubes.' (distributive)

Where there are multiple objects, distributivity may apply to any of them. For example, with the verb 'tie' (causative of 'be attached'), it applies either to the object (3a) or indirect object is plural (3b). Note that the marking of distributivity is optional: the singular form of the verb found in (4c) would be acceptable in (4a, b) too.

(3) a. Kika quih poosj quih **hehet** **pac** iiqui iyahizlca.
Kika DEF rope DEF stick.PL indef.PL [3POS]towards 3;3.RLSYO.CAUS.tie.PL
'Kika attached the rope to a few sticks.' (distributive)

b. Kika quih **poosilca quih** hehe quih tazo iiqui iyahizlca.
Kika DEF rope.PL DEF stick DEF one [3POS]towards 3;3.RLSYO.CAUS.tie.PL
'Kika attached the ropes to one stick.' (distributive)

c. Kika quih poosj quih hehe quih tazo iiqui iyahizj.
Kika DEF rope DEF stick DEF one [3POS]towards 3;3.RLSYO.CAUS.tie
'Kika attached the rope to one stick.' (neutral)

With intransitive (non-causative) verbs, what is normally a plural subject form (4a) can be used with a singular subject (4b). In this case the plurality inherent to the form is interpreted as distributed over the subject, i.e. a single bread is burned in multiple places.

(4) a. Siimet coi yitalc. b. Siimet zo yitalc
bread DEF.PL RLSYO.burn.PL bread INDEF.SG RLSYO.burn.PL
'The breads have burned.' (PL SBJ) 'The bread has burned (a little).' (distributive)

This construal of distributivity may itself be absorbed into a transitive configuration through causative derivation, as in (5), where the same plural verb form as in (2) is used, with the property of 'being tubular' distributed over a singular object ('ironwood').

(5) Gabriel quih **hesen** **z** iyahoiilc.
     Gabriel DEF ironwood INDEF.SG 3;3.RLSYO.CAUS.be_tubular.PL (distributive)
     'Gabriel made an ironwood log into tubes.' SC: if it's a big log, a trunk

3 Morphology

Number marking in the Seri verb involves dozens of polyfunctional allomorphs whose sole consistent function is to mark incremental plurality. For example, the suffixes *-c*, *-ca*, *-coj* and *-cam* in (6) can be either neutral or multiple, depending on the verb, but along a scale of *-c* < *-ca* < *-coj* < *-cam*, the ones to the right always indicate a greater degree of plurality. An exception to this is the suffix *-(t)im*, which is a dedicated SINGULAR SUBJECT MULTIPLE marker.

(6) *plural subject neutral*   *plural subject multiple*

     -teepzaj-c   -teepzal-ca   'sit in'
     -axnal-ca   -axnal-coj   'scold'
     -anaml-coj   -anaml-cam   'hurry'
     -azaail-cam   -azaiil-cam   'anchor'   (note stem alternation)

With a singular subject, the distributive is characterized as being somewhere in the scale between a SINGULAR SUBJECT NEUTRAL form and a PLURAL SUBJECT NEUTRAL form. Consider the two verbs in (7), from Moser & Marlett (2010), for which a distributive distinction has not been identified. The plural subject suffixes reflect a number scale of *-ca* < *-coj* < *-am*. The verb 'tie' (7) has all three suffixes, with the lowest on the scale being used for singular subject distributive (see example 3).

(7)

SG NEUT	SG MULT	PL NEUT	PL MULT	
yihinej	yihinel-im	yihinel-ca	yihinel-coj	'be exposed'
iyoqueetij	iyoqueetil-im	iyoqueetil-coj	iyoqueetil-am	'lift partially'

(8)

SG NEUT	SG MULT	SG DIST	PL NEUT	PL MULT	
yahizj	yahizal-im	yahizl-ca	yahizl-coj	yahizal-am	'tie'

With plural subjects we find no comparable distinction of distributivity. Plural subjects have at most two forms, one which is neutral and the other which can be interpreted either as multiple or as distributive, or indeed as both. When all three types of plurality coincide: subject plurality, event plurality (multiple) and distributivity, they are systematically conflated and realized by a form which expresses the maximal degree of plurality.

4 Parallels with nominal plurality

The same system of plural markers (minus the explicitly verbal SG MULT suffix -*(t)im*) are found with nouns. Comparison with possessed nouns is particularly instructive: these may distinguish two plural forms. The maximally plural form marks plurality of EITHER possessor or possessed item, while the one which expresses a lesser degree of plurality is used specifically for a singular possessor and plural possessed item. This intermediate form is not only morphologically comparable to the verbal distributive, when such nouns are verbalized (a regular process), they must be interpreted as such; thus, the form in (9b) with its plural object (thus distributive) contrasts both with the SINGULAR NEUTRAL *iyasleepec,* and the SINGULAR MULTIPLE *iyasleepejquim.*

(9) a. *noun*   b. *derived verb*

i-sleepejc   i-y-a-sleepejc

POSS-place.behind.ear.PL   TR-RLSYO-CAUS-place.behind.ear.PL

'places behind his/her ears'   'S/he removed lice from them' (from behind their ears)

5 Conclusion

Seri verbs distinguish three kinds of plurality, compressed into a single scale of incremental number marking. While subject and event plurality are reliably distinguished, distributivity emerges intermittently as an extra distinction within this system, partly by exploiting available slots in the hierarchical number marking system, and partly through coercion with subject-verb mismatches; the exact conditions are still a matter for investigation. A striking property of this system is that although distributivity is orthogonal to subject number, it is expressed as a degree of plurality along the same scale: singular subject distributive is expressed as being more plural than an unmarked (neutral) singular subject, but less plural than a plural subject, thus comparable to a paucal in a nominal number system.

References

Baerman, Matthew. 2016. Seri verb classes: Morphosyntactic motivation and morphological autonomy. *Language* 92(4). 792–823.

Cabredo Hofherr, Patricia, Jérémy Pasquereau & Carolyn O'Meara. 2018. Event plurality in Seri. *Proceedings of Semantics of Under-Represented Languages in the Americas* 10: 1-16.

Marlett, Stephen A. 1981. *The structure of Seri.* PhD thesis, UCSD.

Marlett, Stephen A. 2016. *Cmiique Iitom: The Seri language.* Ms., University of North Dakota.

Mithun, Marianne. 1999. *The languages of native North America.* Cambridge: CUP.

Moser, Mary B. & Stephen A. Marlett. 2010. *Diccionario seri-español-inglés.* Mexico City and Sonora: Plaza y Valdés Editores and Universidad de Sonora.

A re-analysis of verb morphology in South Bolivian Quechua: a case study of the Uma Piwra rural variety

Gladys Camacho Rios

The University of Texas at Austin

This study reanalyzes the verb morphology for *Uma Piwra* Quechua variety (UPQ), spoken in the Southern part of Bolivia. The analysis is supported by 20 hours natural speech corpus collected in the everyday conversations and daily work activities. This paper contributes providing a thorough description of the verb morphology and considers the standing of UPQ in recent work of typology of morphologically complex verbs (Mattissen 2017 and Fortescue 2017). It challenges three descriptive claims that are either made explicitly or presupposed in studies of South Bolivian Quechua (Garland D., et al 1971, Muysken 1986, Herrero & Sanchez 1978, Van de Kerke 1993 and Plaza 2009): *(i)* there are no lexical restrictions on the productivity of derivational suffixes; *(ii)* some suffix-suffix combinations are built incrementally and compositionally rather than as complex lexemic chunks; and *(iii)* variable affix ordering always covaries with a distinction in scope. Additionally, the data in previous studies are not based on naturalistic speech and hence many examples are infelicitous or only marginally acceptable. My analysis primarily uses natural speech followed by native speaker judgements to arrive to a deep understanding of UPQ morphology. Concerning recent work on typology of morphologically complex verbs, following Mattissen (2017) and Fostescue (2017), I argue that Quechua displays typical polysynthetic verb structure with regard to its rich morphological elaboration, but is atypical because it does not allow any compounding nor verbal suffixes with typical adjectival or nominal meaning types.

Contrary to claims made in previous studies, most derivational suffixes are not in fact fully productive in natural speech. For example, the suffix -***kipa*** in (1a) appears described as productive suffix meaning *"repetition"* in Herrero and Sanchez (1978) and Van de Kerke (1993); however, according to my analysis this suffix is a lexeme-building suffix since it only appears combined to two verbs and it denotes the meaning: *'to change texture of grains by softly grinding or toasting',* indeed it only applies to grain type objects as follows:

(1) *jank'a-spa (quinoa)*   *kuta-**kipa**-nchik,*   *maylla-nchik*
toast-GER   grind-TRANS-1PL   wash-1PL

'we softly grind it (quinoa) after toasting, we wash it'

This study also shows that productive suffixes are built incrementally and compositionally. For example, in (2) and (3) the suffix -***chi*** meaning *'X' causes/makes 'Y' to do V* and -***pu*** 'benefactive' each suffix bears a compositional meaning when they co-occur.

(2) *achkha*   *llant'a-s-ta-pis*   *apa-**chi**-ni*
A lot   wood-PL-ACC-also   carry-**CAUS**-1SG > 3SG
'I also made someone carry a lot of wood' (Asteria)

(3) *kachi* *yaku-pi*   *chullu-**chi**-**pu**-ni*
salt   water-LOC   sock-**CAUS-BEN**-1SG > 3SG
'I make it (chicken meat) soak in salt water' (Angela)

However, certain productive suffixes in UPQ, when combined, form suffix clusters whose meaning is no longer compositional. Previous studies by Solá and Lastra (1964), Bills et al., (1971), Herrero & Sanchez (1978), Muysken (1986), Van de Kerke (1993) and Plaza (2009), do not observe suffix-suffix clusters. For example, in (4) the meaning of the cluster -*kampu* '*to do V, for good/safeness to avoid the bad presentiment the subject has*' cannot be predicted from the individual meanings -*ku* '*reflexive*' -*mu* '*motion*' -*pu* '*benefactive*'. In my analysis, when these three suffixes co-occur they necessarily involve phonological changes and occupy single slots rather than different slots in the verb template. In fact, such non-compositional meanings remain understudied in wider literature (Fortescue 1980).

(4) *imapis*   *jap'i-nman*   *ni-spa,*
     we never know what   catch-3SG.COND   say-GER,
     *qati-**kampu**-ni*   *ah*
     take animals-**back.for.good**-1SG   ah
     'We never know what it can catch it (the brooding turkey sitting far away
     from home) saying that I took/take it back for good' (Angela)

This study also uncovers examples where variable affix ordering does not covary with a difference in scope. While the literature suggests more suffixes freely varying with respect to each other, the analysis found restricted free-variation among single productive suffixes. Only -*rqu* and -*rpa* can freely vary with respect to -*ysi*. Indeed, the free affix variation occurs without an obligatory difference in scope, as exemplified in (5) and (6).

(5) t'iqpa-**ysi-rpa**-lla-sqayki
     peel-ASSIT-suddenly-keep-1>2SG.FUT
     'I help you keep peeling (dry corn)'

(6) t'iqpa-**rpa-ysi**-lla-sqayki
     peel-suddenly-ASSIT-keep-1>2SG.FUT
     'I help you keep peeling (dry corn)'

This study presents a holistic description for UPQ morphology. Even though UPQ has limited polysysnthesis typologically because it lacks compounding, and adjectival- and nominal-type verbal suffixes, productive non-inflectional suffixes display very concrete and well elaborated adverbial meanings. In the following examples, we can see that concrete adverbial meanings are expressed in the way the verb is performed. The fact that only adverbial-type suffixes are well elaborated makes UPQ typological atypical.

(7) *chiri-pi-chá*   *wañu-**rpa**-saq*   *ni-chka-ni*   *jajaja*
     cold-LOC-DUB   die-**suddenly**-1SG.FUT   say-PROG-1SG   hahaha
     'I am saying that "I will suddenly drop dead in the cold"' (Angela)

(8) *runt-it-u*   *ka-pti-n*   *runtu*   *wayk'u-**rqu**-ku-ni*
     egg-DIM   exist-if-3SG   eggs   cook-**nimbly**-REF-1SG
     'If there are eggs, I nimbly and happily cook eggs myself' (Genoveva)

(9) *'talaq, talaq'* *ni-**ri**-chi-chka-nku* *imata-chá*
    'talaq, talaq' sound-***nicely***-CAUS-PROG-3PL   something-DUB
    'talaq, talaq, it seems they are making sound on something' (Copertino)

References

Adelaar, Willem F H. n.d. "Quechua I y Quechua II: En defensa de una distinción establecida," 21. 2004 (Willem Adelaar, in collaboration with Pieter Muysken) The Languages of the Andes. Cambridge: Cambridge University Press

Bills, Garland D., et al (1971) An introduction to Spoken Bolivian Quechua. Institute of Latin American Studies. *University of Texas Press, Austin and London.*

Fortescue, Michael. 2017. *The Oxford Handbook of Polysynthesis*. Oxford Handbook of References. New York, NY: Oxford University Press.

Fortescue, Michael. 2017. <What are the Limits of Polysynthesis?> Pp.1-25. in Fortescue, Michael. 2017. *The Oxford Handbook of Polysynthesis*. Oxford Handbook of References. New York, NY: Oxford University Press.

Fortescue, M. D. 1980. "Affix Ordering in West Greenlandic Derivational Processes." *International Journal of American Linguistics* 46 (4): 259–78.

Haspelmath, Martin, and Andrea D. Sims. 2010. *Understanding Morphology*. 2nd ed. Understanding Language Series. London: Hodder Education.

Herrero S.J., J & Sáanchez de Lozada (1978). *Gramática del Quechua*: estructura del Quechua Boliviano contemporaneo, Edotorial Universo Ltda., Cochabamba.

Lastra,Y. (1986). Cochabamba Quechua Syntax, Mouton The Hague.

Mattissen, Johanna (2004). A structural typology of polysynthesis. Word volume 55. Number 2

Matissen, Johanna. 2017. <Sub-Types of Polysynthesis> Pp.1-42 in Fortescue, Michael. 2017. *The Oxford Handbook of Polysynthesis*. Oxford Handbook of References. New York, NY: Oxford University Press.

Muysken, P. (2009). Approaches to affix order. *Linguistics*, 24(3), pp. 629-644. Retrieved 18 Feb. 2019, from doi:10.1515/ling.1986.24.3.629

Muysken, Pieter. (1981) <Quechua Word Structure> Pp.279-327 in Heny, Frank. (1981). *Binding and filtering.* Cambridge, Mass : MIT Press

Muysken, Pieter. (1988) <Affix order and interpretation: Quechua> in M. Everaert, M.Thommelen (eds.), Morphology and modularity: in honour of Henk Schiltink, pp. 259-280. Dordrecht [etc.]:Foris.

Rice, Keren (2011). Principles of affix ordering: An overview. Word Structure 4.2 (2011): 169-200. Edinburg University Press.

Van de Kerke, Simon Cornelis. 1996. Affix order and Interpretation in Bolivian Quechua.Printed by Krips Repro, Meppel, the Netherlands

Postverbal *liaisons* in traditional songs: a morphological reanalysis?

Gasparde Coutanson

MoDyCo – UMR 7114 et Université Paris

Nanterre

Côté (2005: 67) synthesizes the main analyses that have been suggested for the morphophonological phenomenon of *liaison*, a consonantal external sandhi occurring for example in French and which has been seen either as a truncation, a suppletion or a floating segment (or an epenthesis). Bürki and Laganaro (2014: 169) give the following definition of *liaison*: "some words in French entail two realizations (or variants), the second differing from the first one by an additional consonant in specific contexts" [our translation], thus taking sides with proponents of a suppletion hypothesis. Studies of *liaison* in morphology (e.g. Bonami & Boyé, 2003 or Plénat & Plénat, 2011) have mostly focused on suppletion of adjectival forms occurring before nouns. Among the epenthetic hypothesis, some *liaisons* could also have been given a morphological role, i.e. [z] *liaisons* are often seen as a plural marker (cf for instance the review of literature made by Mallet 2008: 59-61).

Postverbal *liaisons* usually occur after verbal suffixes (which constitute a *portemanteau* morph as French verbal suffixes reflect at the same time tense, mood, number and person in a unit that is not dividable anymore) and can be produced, under certain conditions, almost exclusively before a vocalic-initial word. As an illustration, consider example (1) from our ethnomusicologist corpus of traditional French songs collected in the 1960s among adult singers in Central France[1]:

(1)[2]   nous꞊couch-er-ons   ensemble

        nu꞊kuʃ-ər-ɔ̃   zãsãblə

        SBJ.1.PL꞊sleep-FUT.IND-1.PL[3]   together

        'We will sleep together'

If produced, postverbal *liaisons* after verbs conjugated in first or second person involve the segment [z] while after verbs conjugated in the third person, another segment is involved, [t]. Previously, this has partly been noticed by Morin & Kaye, who have proposed [t] to be a "verbal marker" (Morin & Kaye, 1982: 323). A more fine-grained analysis has been presented by Tranel, who noted that [t] was "the linking consonant for verbs conjugated in the third

---

[1] Primary data come from the Jean Dumas fund (*fonds Jean Dumas*) and are available online on the Interregional base of French oral heritage (*Base inter-régionale du patrimoine oral*) at http://patrimoine-oral.org/dyn/portal/index.seam?aloId꞊11898&page꞊alo&fonds꞊&cid꞊1535. Songs are being transcribed and annotated for *liaisons* and *pataquès* in an ongoing project using CLAN (MacWhinney, 2014).

[2] All examples presented in this paper come from our sung corpus.

[3] Here the morphemic gloss we use is based on the Interlinear morphemic gloss proposed by Lehmann (see Lehmann, 2004).

person (singular and plural)" (Tranel: 1987, 175) while [z] was "the first and second person marker for both singular and plural conjugated verbs" (Tranel: 1987, 176).

In our corpus, some *liaisons* seem quite unexpected regarding the written form of the verb. These *liaisons* are called *pataquès* and are generally unwritten (see examples (2), (3) and (4) below). As these occurrences are produced in songs, some may say it is related to an anti-hiatus strategy, that is producing a consonantal epenthesis between two vowels to respect the canonical syllabic structure of the language. Morin did not observe this in the historical development of liaison for spoken French (see Morin, 2005), but it leads us to the following question: is *pataquès* just a phonological process? Postverbal *liaisons*, expected or not, would be studied in this presentation as we can notice that the choice of the consonant produced and its place does not seem random.

Was there a reanalysis of postverbal *liaisons* as morphemes? Although these *liaisons* are quite rare in current everyday language (for example in the PFC corpus[4] only 23% of the liaisons are realized in postverbal contexts[5])[6], they can be noticed in distance varieties (see Koch & Oesterreicher, 2001 on the concept of distance/immediacy) such as songs or among politicians. In a usage-based perspective, it could be hypothesized that hearers extract some regularities from their linguistic input and assign them a morphological role before reusing them analogically in similar syntactic contexts. We could rise the hypothesis that for our sung corpus a morpheme –z would have been extracted in postverbal contexts as a marker dealing with speech-act participants (speaker and addressee) and that a morpheme –t would have been extracted in these contexts as a marker dealing with non-speech-act participants (sometimes referred to as other). Typologically, this would be related to the concept of person hierarchy, sometimes proposed in the literature as 1, 2 > 3 (Siewierska 2014: 151). This ranking fits our case the most, as the subject would be indexed on the verb with a suffix: -z, attributed to subjects for whom the referents are speech-act participants and -t, attributed to subjects for whom the referents are non-speech act participants. That first and second persons would be indexed differently on the verb could be linked to another typological concept: the animacy hierarchy where human > animate > inanimate > abstract (Siewierska 2004: 149), as speech-act participants are necessarily human contrary to non-speech act participants for whom it is not an obligation, thus assuring a direct indication of the humanness of the subject. For the singular in written Old French, Andrieux & Baumgartner (1983: 58 [Morin & Bonin,

---

[4] See Durand et al., 2011 for more information on the project Phonology of Contemporary French (*Phonologie du Français Contemporain*).

[5] The *liaisons* contexts were extracted from the online research database in June, 2019 when occurring after a word tagged as a present (VER:pres), a conditional (VER:cond) or an imperfect form (VER:impf). Part-of-speech tagging has not been controlled and uncertain occurrences have been excluded.

[6] We have very few spoken data in the Jean Dumas fund. Thus comparing spoken vs sung productions of our singers, especially since we knew that French was not the only language used by our informants who also spoke some regional language, did not seem relevant.

1992: 36]) have already proposed that the expansion of graphic ‹ s › at the end of verbs conjugated in first person singular could be linked to a wish to morphologically distinguish the couple I-You (*personnes d'allocution*) from the third person.

Two arguments could be provided against our hypothesis. First of all, contexts in which *liaison* can appear, upon certain conditions among which the next word should be vocalic-initial, are quite rare as vocalic-initial words are less frequent than consonantal-initial words in spoken French, so the phenomenon would not constitute a systematic verbal agreement with the subject. Furthermore, consider example (2) where a *pataquès* is produced. After *chantait* it is not the expected segment with respect to our hypothesis that will be pronounced, as [z] does not reflect the usual segment used after verbs conjugated in the third person, [t]:

(2) le plus jeune   des   trente chantait   une   chanson

 l-ə = ply = ʒoenə   de   tʁɑ̃tə ʃɑ̃t-ɛ   zynə   ʃɑ̃sɔ̃

 DEF-M = SUP = young ASS:DEF-PL thirty sing.IND-IMPF.3.SG INDEF:F song

 'The youngest of the thirty was singing a song'

Nonetheless, some arguments can also be put forward to back up this position. First, some *pataquès* point in this direction. In (3), some *liaison* that cannot be influenced by the written form of the verb *reviendrai* is produced:

(3) je reviendrai   au   pays

 ʒə = ʁəvjɛ̃-dʁ-ɛ   zo   pɛi

 SBJ.1.SG = return-FUT.IND-1.SG   at:DEF.M   country

 'I will come back to homeland'

In (4), another argument in favor of our position is that this segment can also occur not directly after verbs but also after their pronominal enclitics. For Morin (1986: 190), French pronominal vocalic-initial enclitics *en* and *y* are allomorphic forms of the same clitic (*en/zen, y/zy*). In fact, in French imperative the subject is always a speech-act participant (2S, 1P or 2P) and maybe this oral production does not differ from postverbal *liaisons* but occur in an unusual place that could be due to a reanalysis of the linguistic input.

(4) apprends-moi   à   parler

 aprɑ̃ = mwɑ   za   pɑl-e

 learn.PRS.IMP.2.SG = OBJ.1.SG DAT   speak-NFIN

 'Teach me how to speak'

Last but not least, a brief diachronic overview shows that in Latin, the final graphemes ‹ s › and ‹ t › on verbs were pronounced. Final ‹ s › were to be found on verbs conjugated at the second person singular, first and second persons plural while final ‹ t › were to be found on verbs conjugated at the third persons singular and plural. French verbs in first person singular actually present a final graphic ‹ s › in what is traditionally called second and third groups and orally in our corpus some postverbal *liaison* can be produced for these groups or a *pataquès* after verbs from the first group (cf (3)). Our goal here is not to show a preference for an explanation of expansion of ‹ s › or [z] to verbs conjugated in first person singular by graphical

analogy or by oral analogy from Latin to French. Instead, we would rather like to emphasize that verbs conjugated at the first person singular did not present any ‹ s › in Latin, contrary to some French verbs.

Abbreviations

=: clitic boundary	F: feminine	INDEF: indefinite	PRS: present
1, 2, 3: 1st, 2d, 3d person	FUT: future	M: masculine	SBJ: subject
ASS: associative	IMP: imperative	NFIN: non-finite	SG: singular
DAT: dative	IMPF: imperfect	OBJ: object	SUP: superlative
DEF: definite	IND: indicative	PL: plural	

References

Andrieux, Nelly, et Emmanuèle Baumgartner. 1983. Manuel du français du Moyen Age. 3. Systèmes morphologiques de l'ancien français. A. Le verbe.

Bonami, Olivier, et Gilles Boyé. 2003. « La nature morphologique des allomorphies conditionnées: les formes de liaison des adjectifs en français ». Fradin et al.

Bürki, Audrey, et Marina Laganaro. 2014. « Représentation et encodage de la liaison : approches psycholinguistiques et neuropsycholinguistiques ». In La liaison, approches contemporaines, 110:169-85. Sciences pour la communication. Bern: Peter Lang.

Côté, Marie-Hélène. 2005. « Le statut lexical des consonnes de liaison ». Langages, 2: 66-78.

Durand, Jacques, Bernard Laks, Basilio Calderone, et Atanas Tchobanov. 2011. « Que savons-nous de la liaison aujourd'hui ? » *Langue française* 169(1): 103.

Koch, Peter, et Wulf Oesterreicher. 2001. « Langage parlé et langage écrit ». Lexikon der romanistischen Linguistik 1 (2): 584-627.

Lehmann, Christian. 2004. « Interlinear morphemic glossing ». Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung 2: 1834-57.

MacWhinney, Brian. 2014. The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs. Psychology Press.

Mallet, Géraldine-M. 2008. « La liaison en français : descriptions et analyses dans le corpus PFC ». Université Paris Ouest Nanterre La Défense.

Morin, Yves Charles. 2005. « La liaison relève-t-elle d'une tendance à éviter les hiatus? Réflexions sur son évolution historique ». Langages, 2: 8-23.

Morin, Yves-Charles. 1986. « On the morphologization of word-final consonant deletion in French ». Sandhi phenomena in the languages of Europe, 167-210.

Morin, Yves-Charles, et Michèle Bonin. 1992. « Les -S analogiques des 1sg au XVI siècle : les témoignages de Meigret et Lanoue ». Revue québécoise de linguistique 21 (2): 33.

Morin, Yves-Charles, et Jonathan D. Kaye. 1982. « The Syntactic Bases for French Liaison ». Journal of Linguistics 18 (02): 291.

Plénat, Marc, et Camille Plénat. 2011. « La liaison de l'adjectif sur le nom en français: morphologie, syntaxe, phonologie ». Linguistica 51 (1): 299-315.

Siewierska, Anna. 2004. Person. Cambridge University Press.

Tranel, Bernard. 1987. *The sounds of French: An introduction*. Cambridge University press.

Integration of comparative degree into the adjective paradigm.
A case of simple affixation with complex alternations.

Magdalena Derwojedowa

University of Warsaw, Institute of Polish Language

In grammatical tradition, qualitative adjectives are gradable. They share this feature with adverbs. However, it is not always clear, if gradability is fully inflectional or essentially derivational—grammars tend to concern it as lexical derivation, pointing at its irregularity, and semantic, rather than formal, limitations (Szober, 1923; Heinz, 1961; Saloni & Świdziński, 2001; Grzegorczykowa et al., 1998). In consequence, comparatives and superlatives are linked to positive degree adjectives as separate entries in dictionaries or listed as quasi-suppletive forms (ISJP; SGJP3).On the other hand, adverbs are traditionally regarded as a different, inflexible class of lexemes, although they are regularly derived from adjectives and despite the fact that they do generate degree forms. Some grammarians regard them as neutral (to the case, number, and gender) form of the adjectives (Tokarski, 1973; Saloni, 1976; Saloni & Świdziński, 2001; SGJP3).

We aim to present, how to include comparative and superlative degrees into formally defined adjective patterns. We also demonstrate that such paradigms can be extended to adverbs.

Polish has singular and plural, seven cases and no less than five genders (Mańczak, 1956; Corbett, 1983; Saloni & Świdziński, 2001). Adjective syncretisms resulting in the split of the masculine in virile, animate and inanimate, is crucial in syntax, but not in paradigmatic inflection, because accusative is always syncretic with a form already present in the grid. In the plural, the virile is opposite to all other (i.e., non-virile) genders. It has unique nominative and maintains accusative-genitive syncretism. So there are three sets of forms in the singular: masculine, feminine and neuter, and two in the plural. Due to the syncretism of masculine(s) and neuter forms in the singular and virile and non-virile in oblique cases in the plural, only 11 forms fill in all 70 (7·2·5) slots in the paradigm, cf. Tab. 1 (Saloni, 1992; Blevins, 2003).

	vir	manim	m -anim	f	n	vir	-vir
N	ciepły (1)			ciepła (2)	ciepłe (3)	ciepli (9)	=(3)
G	ciepłego (4)			ciepłej (5)	=4	ciepłych (10)	
D	ciepłemu (6)			=(5)	=(6)	=(8)	
A	=(4)		=1	ciepłą =(7)	=(3)	=(10)	=(3)
I	ciepłym (8)			=7	=(8)	ciepłymi (11)	
L	=(8)			=(3)	=(8)	=(11)	
V	=(1)			=(2)	=(3)	=(9)	=(3)

Table 1: Characteristic forms of adjective.

Ten of these forms are attached by one variant of the stem (S1), only one (form 9)—by the other (S2). Endings' variants depend on the final consonant of the stem (soft, hardened, hard, velar stop), resulting in the inflectional types showed in Tab. 2.

Positive degree is the base for the comparative, that is made up either by affixation (ex. 1) or periphrastically (ex. 2-3). The choice between the two depends mainly on factors outside morphology. However, if an adjective (or an adverb) allows suffixation, it is possible to make a periphrastic comparative (sometimes not entirely plausible, but frequent in everyday speech). The scope of periphrastic comparison includes degree forms of (adjectivized) participles or even relative adjectives (otherwise not gradable, ex. 3-4; cf. Grzegorczykowa et al. 1998) and also

type	stem1										stem2
	1	2	3	4	5	6	7	8	10	11	9
I: soft	-i	-ia	-ie	-iego	-iej	-iemu	-ią	-im	-ich	-imi	-i
II: hardened	-y	-a	-e	-ego	-ej	-emu	-ą	-ym	-ych	-ymi	-i
III: k, g	-i	-a	-ie	-iego	-iej	-iemu	-ą	-im	-ich	-imi	-y
IV: hard	-y	-a	-e	-ego	-ej	-emu	-ą	-ym	-ych	-ymi	-y

Table 2: Types of adjective paradigms.

enables negative comparison (ex. 3, cf. Tokarski, 1973). However, this kind of comparison is not about wordform construction, so we put it away in the current discussion.

(1) Jan   jest   wyższy   od   brata.
John.NOM.SG be.PRES.3.SG tall.NOM.SG.COMP than brother.GEN.SG
'John is taller than his brother'.

(2) Jan   jest   bardziej / mniej stanowczy   od   brata.
JOHN be.PRES.3.SG more   / less   resolute.NOM.SG.POS than brother.GEN.SG
'John is more / less resolute than his brother'.

(3) Jan   jest   bardziej / mniej wykształcony   od   brata.
JOHN be.PRES.3.SG more   / less   educated.NOM.SG.POS than brother.GEN.SG
'John is more / less resolute/educated than his brother'.

(4) Lektor   ma   bardziej drewniany   głos   niż
announcer.NOM.SG have.PRES.3.SG more   wooden.ACC.SG timbre.ACC.SG than
lektorka.
announcer.FEM.NOM.SG
'The announcer's timbre is more hollow than those of the other one (female)'.

The comparative suffix has two variants: short (*-sz-*, e.g., *gruby → grubszy* 'fat, fatter'), mainly for single consonant ending stems, and vocalized (*-ejsz- ładny → ładniejszy* 'pretty, prettier'), for stems with a consonant group. The suffix triggers alternations (e.g., *długi → dłuższy* 'long(er)') or reductions (e.g., *krótk–i → krót–szy* 'short(er)'). Some adjectives apply both (e.g., *miękki → miększy / miękciejszy* 'soft(er)' (Rothstein, 1993; Grzegorczykowa et al., 1998). As final consonant of the stem determines endings, all comparative and superlative forms belong to inflectional Type II (cf. Tab. 2). In consequence, the paradigm of the gradable adjective typically consists of four adjective stems and two sets of endings, as presented in Tab. 3.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
S1	wysok-	-i	-a	-ie	-iego	-iej	-iemu	-ą	-im		-ich	imi
S2	wysoc-									-y		
S3	wyższ-	-y	-a	-e	-ego	-ej	-emu	-ą	-ym		-ych	-ymi
S4	wyżs-									-i		

Table 3: Characteristic forms of *wysoki* 'high' in positive and comparative degrees.

Adverbs are regularly derived from adjectives with the use of one of the two suffixes (from lexical derivational position)/endings (from the inflectional stand): *-o* and *-e.* In some cases any of them can be attached; the resulting adverbs differ in meaning (ex. 5 and 6).

(5) Pilnie   /*pilno   potrzebuję   nowego   samochodu.
    urgently.ADV /*in a hurry.ADV need.PRES.1.SG new.GEN.SG car.GEN.SG

    'I need a new car urgently'.

(6) Pilno   /*pilnie   mi   do   domu.
    In a hurry.ADV /*urgently.ADV I.DAT to.PREP home.GEN.SG

    'I am in a hurry to go back home'.

Comparative degree exponent of the adverb is *-ej* (e.g., *głośno → głośniej* 'loud(er).ADV'). Both positive and comparative endings can trigger alternations (e.g., *mądry → mądrze* 'smart.ADJ → .ADV', *rozmaity → rozmaicie* 'various.ADJ → .ADV', *gęsto → gęściej* 'thick(er).ADV', *śmiało → śmielej* 'brave(r).ADV'), and comparison also reductions (e.g., *wysoko → wyżej* 'high(er).ADV'). Adverbs with *-o* ending have S1 stem, with *-e* have S2, so they fit as form 12 in the relevant column of Tab. 1, but the comparative form needs a separate stem (S5 in Tab. 4).

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
S1	wysok-	-i	-a	-ie	-iego	-iej	-iemu	-ą	-im		-ich	imi	-o
S2	wysoc-									-y			-e
S3	wyższ-	-y	-a	-e	-ego	-ej	-emu	-ą	-ym		-ych	-ymi	
S4	wyżs-									-i			
S5	wyż-												-ej

Table 4: Modified paradigm of *wysoki* 'high'. Adverb is form (12).

Morphological superlative results from the prefixation of comparative (Rothstein, 1993; Grzegorczykowa et al., 1998). The prefixation does not affect stem (cf. ex. 1 and 2), so the rule is general: *naj-* prefix +comparative form (e.g., *najlepszy ← lepszy* 'best, better', *najzieleńszy ← zieleńszy* 'greenest, greener', *najgłośniej ← głośniej* 'loudest, louder.ADV').

We presume that any adjective form is tripartite (cf. design of SGJP database in Woliński, 2009). We distinguish the longest string common to all wordforms of a lexeme, that is a *root*, and the altering part, that is a *root extension*. The root and the root extension form together the stem. For example, *wysoki*'s 'high' root is *wy~*, its extensions are *~sok-*, *~soc-*, *~ższ-* and *~ż-*, (cf. Tab. 4); *zły* 'bad' has empty (∅) root (i.e. no string is common to all forms of a lexeme), *zł-* (*zły, zła, złe…,*, Type IV in Tab. 2) and *źl-* (*źli*, adverb *źle*) as extensions of positive degree and *gorsz-* (*gorszy, gorsza…*), *gors-* (*gorsi*), *gorz-* (*gorzej*) of comparative degree. Tab. 5 presents complete patterns for *mały* 'little', *twardy* 'hard' and *suchy* 'dry'. The first one of these is individual, but the others are models for their respective types, e.g., *twardy, hardy* 'haughty', *młody* 'young', *chudy* 'thin'; *suchy* 'dry', *kruchy* 'fragile, frail, brittle', *cichy* 'quiet'.

Because similar alternations (in smaller or larger extent) also occur in other Slavic languages (cf. BCS *gust → gušči* 'thick', *skup → skuplij* 'expensive', Belorussian *багаты → багацей* 'rich' *высокі → вышэй* 'tall', Czech *hebký →hebčí* 'supple', *suchý → sušší* 'dry'), we suppose the model may also be valid for these languages.

References

Blevins, James P. 2003. Stems and paradigms. *Language* 79(4). 737–767.

Corbett, Greville G. 1983. The number of genders in Polish. *Papers and Studies in Contrastive Linguistics* VXI. 83–89. `http://epubs.surrey.ac.uk/2224/`.

Dressler, W. 2016. Prototypical differences between inflection and derivation (42(1)). 3–10. doi:10.1515/stuf-1989-0102.

	root	extension	endings
mały 'small'	m~	~ał- ~al-	-y, -a, -e, -ego,… -ym, -ych, -ymi, -o -i
naj- (SUP) / ∅		~niejsz- ~niejs- ~ni-	-y, -a, -e, -ego,… -ym, -ych, -ymi -i -ej
twardy 'hard'	tward~	~∅ ~z-	-y, -a, -e, -ego,… -ym, -ych, -ymi, -o -i
naj- (SUP) / ∅		~sz- ~s- ~zi-	-y, -a, -e, -ego,… -ym, -ych, -ymi -i -ej
suchy 'dry'	su~	~ch- ~s-	-y, -a, -e, -ego,… -ym, -ych, -ymi, -o -i
naj- (SUP) / ∅		~chsz- ~chs- ~sz-	-y, -a, -e, -ego,… -ym, -ych, -ymi -i -ej

Table 5: Paradigms with tripartite forms of adjectives.

Grzegorczykowa, R., R. Laskowski & H. Wróbel (eds.). 1998. *Morfologia (Morphology)*, vol. II Gramatyka współczesnego języka polskiego (Grammar of contemporary Polish language). Warszawa: Wydawnictwo Naukowe PWN 2nd edn.

Heinz, A. 1961. Fleksja a derywacja (Inflection and derivation). *Język Polski* XLI.

ISJP (ed.). 2000. Bańko, M. *Inny słownik języka polskiego (Different dictionary of Polish)*. Warszawa: Wydawnictwo Naukowe PWN.

Mańczak, W. 1956. Ile jest rodzajów w języku polskim? (How many genders is in Polish?). *Język Polski* XXXIV.

Rothstein, A.R. 1993. *The Slavonic languages* chap. Polish, 686–758. Routledge.

Saloni, Z. 1976. *Cechy składniowe polskiego czasownika (Syntactic features of Polish verb)*. Wrocław.

Saloni, Z. 1992. *Rygorystyczny opis polskiej deklinacji przymiotnikowej (Rigorous description of Polish adjective declination)* (Uniwersytet Gdański. Prace Językoznawcze 16). Gdańsk.

Saloni, Z. & M. Świdziński. 2001. *Składnia współczesnego języka polskiego (Syntax of contemporary Polish)*. Warszawa: Wydawnictwo Naukowe PWN.

Schenker, A. M. 1964. *Polish declension: a descriptive analysis*. The Hague: Mouton.

SGJP3. 2015. Saloni, Z., Woliński, M. and Wołosz, R. and Gruszczyński, W. and Skowrońska, D., *Słownik gramatyczny języka polskiego (Grammatical dictionary of polish)*. online, ed. III, access 01/21/2019. Access: 01/25/2019. `http://sgjp.pl/`.

Swann, O.E. 1983. *A concise grammar of Polish*. Lanham, Md.: University Press of America.

Szober, S. 1923. *Gramatyka języka polskiego (Grammar of Polish language)*. Lwów—Warszawa.

Tokarski, J. 1973. *Fleksja polska (Polish inflection)*. Warszawa.

Woliński, M. 2009. A relational model of Polish inflection in *Grammatical Dictionary of Polish*. In Z. Vetulani & H. Uszkoreit (eds.), *Human language technology: Challenges of the information society* (Lecture Notes in Artificial Intelligence 5603), 96–106. Berlin: Springer-Verlag.

Modelling diasystemic inflexion:
Verb morphology in the *Croissant linguistique*

Maximilien Guérin
CNRS – LLACAN
(UMR 8135)

Louise Esher
CNRS - CLLE-ERSS
(UMR 5263)

Jean Léo Léonard
Université Montpellier 3
& Dipralang (EA 739)

Sylvain Loiseau
Université Paris 13
& LACITO (UMR 7107)

1   Introduction

Our paper explores how tools developed for the formal modelling of individual inflexional systems may be fruitfully applied to the description of inflexion in a dialect continuum or DIASYSTEM. We take as our case study the area of central France termed the CROISSANT LINGUISTIQUE (literally, 'Linguistic Crescent'; Tourtoulon & Bringuier 1876, Brun-Trigaud 1990), usually characterized as a transitional zone between northern Gallo-Romance (Oïl) varieties and southern Gallo-Romance (Occitan) varieties.

For each of six survey points within the Croissant area, and a seventh, Occitan, survey point as a comparator, we analyse the inflexional morphology of the verb, using two principal formalizations: PARADIGM FUNCTION MORPHOLOGY (Stump 2001, 2016, Stump & Finkel 2013, Bonami & Stump 2016) and STEM SPACES (Bonami & Boyé 2002, 2003, 2014, Boyé 2011). By these means, we obtain schemas highlighting the key principles of paradigm organization in each individual variety. Comparison of the schemas reveals structural continuities and discontinuities within the Croissant dialect continuum itself, and between the Croissant dialect continuum and neighbouring varieties.

2 Data

Seven localities, shown in Figure 1, were chosen to represent the study area: Dompierre-les-Eglises (Haute-Vienne), Cellefrouin (Charente), Bonnat (Creuse), Luchapt (Vienne), La Châtre-Langlin (Indre), Naves (Allier) and Châteauponsac (Haute-Vienne).

For each locality, a fieldwork questionnaire was used to elicit inflexional paradigms for 22 verb lexemes, the cognates of French *chanter* 'sing', *lier* 'bind', *couver* 'incubate', *acheter* 'buy', *aller* 'go', *blanchir* 'whiten', *couvrir* 'cover', *partir* 'leave', *vendre* 'sell', *avoir* 'have', *être* 'be', *pouvoir* 'be able', *vouloir* 'want', *savoir* 'know', *devoir* 'have to', *faire* 'do', *venir* 'come', *tenir* 'hold', *dire* 'say', *croire* 'believe', *prendre* 'take' and *voir* 'see'. These lexemes were selected to illustrate the range of conjugational types expected to be present in central Gallo-Romance.
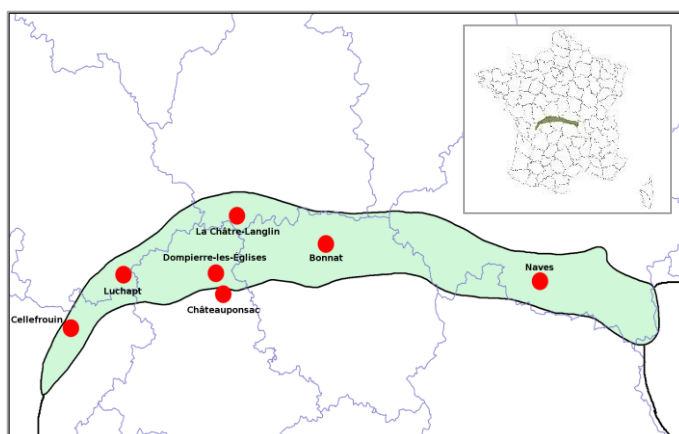
Figure 1. Survey points within the Croissant.
(Inset: Croissant area within France).

For the purposes of the analysis, non-finite forms were excluded from consideration, as the morphological behaviour of these items is liable to show idiosyncratic divergence from that of finite forms (see e.g. Bach & Esher 2013).

3 Analysis

3.1 Method

We identified stems, exponents associated with TAM values, and exponents associated with person/number values, and the paradigmatic distribution of each of these items. Based on this analysis, we formulated stem selection rules and realizational rules within the PFM framework. We then proceeded to model our results for the paradigmatic distribution of stem material, using stem spaces and stem dependency relations.

3.2 Exponents of person/number

A salient property of several inflexional systems in the Croissant area is the uniformity of personal desinences (see e.g. Table 1). For a given TAM category, the series of personal desinences identified showed no variation across lexemes. For a given person/number value, variation across TAM categories is extremely limited: in the singular, the westernmost survey points show a two-way contrast and more central survey points show a three-way contrast, while in the plural no contrast was found. The Croissant systems additionally present a high incidence of SYNCRETISM (Baerman et al. 2005, Baerman 2007) between personal desinences. Some patterns of syncretism are shared with neighbouring Occitan varieties, but not all: notably, the syncretism of second person singular and second person plural, characteristic of north-western Occitan varieties (see e.g. Lavalade 1987), is absent from the Croissant. We hypothesize that the extent of syncretism with respect to person and number values, and the widespread absence of distinctive desinences in singular forms, is linked to the obligatory nature of subject pronouns in varieties of the Croissant.

	IPF.SBJV	PRT	FUT	COND	PRS.IND	IPF.IND	PRS.SBJV
1SG	–	–	e	–	–	–	–
2SG	ɑ	ɑ	ɑ	ɑ	–	ɑ	–
3SG	–	–	–	–	–	–	–
1PL	ɛ̃	ɛ̃	ɛ̃	ɛ̃	ɛ̃	ɛ̃	ɛ̃
2PL	e	e	e	e	e	e	e
3PL	ɛ̃	ɛ̃	ɛ̃	ɛ̃	ɛ̃	ɛ̃	ɛ̃

	α	β	γ
1SG	e	–	–
2SG	ɑ	ɑ	–
3SG	–	–	–
1PL	ɛ̃	ɛ̃	ɛ̃
2PL	e	e	e
3PL	ɛ̃	ɛ̃	ɛ̃

Table 1. Personal desinences in the variety of Bonnat: full and schematic series.

The uniformity of personal desinences across lexemes argues for the inflexional class system in these varieties being based principally or solely on stem distribution (as proposed by Stump & Finkel 2013 for French; compare also Martinet 1958, Dubois 1958) and we therefore propose a classification in these terms.

3.3 Stem distributions

Comparison of the stem spaces obtained reveals a tendency for the distribution of root allomorphy in the varieties of the Croissant to diverge in two principal dimensions from the distributions attested in other Romance languages, such as French (Table 2). Firstly,

inherited patterns of interpredictability between individual TAM categories are split along category lines (future differentiated from conditional; preterite differentiated from imperfect subjunctive). Secondly, where there is differentiation between cells belonging to the same TAM category, such differentiation is aligned with the morphosyntactic opposition of singular and plural. In the varieties of the Croissant, the distribution of root allomorphy is thus becoming increasingly systematized as an exponent of TAM and number.

The stem space analysis also brings internal variation within the Croissant area into focus: for example, singular/plural differentiation within a given TAM category is more extensive in northern than southern or central survey points; differentiation of future and conditional is established in most varieties, but confined to the plural in varieties such as that of Luchapt, and absent from some central survey points.

French	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	3					2
IPF.IND	1					
PRT	11					
IPF.SBJV						
PRS.SBJV	7			8		7
FUT	10					
COND						
IMP	—	5	—	6		

Luchapt	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	3			2		
IPF.IND	1			4		
PRT	—					
IPF.SBJV	9					
PRS.SBJV	5			6		
FUT	7					
COND					8	

Table 2. Stem space for finite forms in French (Boyé 2011:42) and in the variety of Luchapt.

The combination of the stem spaces themselves, and the dependency relations which we identify between stems in individual varieties, thus provide a practical means of comparing the distribution of inflexional exponents across our survey points. Based on these analyses, we identify salient features of the inflexional systems constituting the dialect continuum: shared features and points of contrast among the varieties of the Croissant and the areas adjoining it.

4 Conclusions

The formal descriptors of inflexional systems which we draw on in this study were, in general, developed as models of individual systems. Yet, as argued by Stump & Finkel (2013), such descriptors are of empirical value for the typological and comparative study of speech varieties. Our study of inflexion in the Croissant dialect continuum offers a practical illustration in support of Stump & Finkel's view, also demonstrating the applicability of such formal descriptors to minority languages for which more limited data are available.

Our approach departs from traditional dialectological practice in that we are concerned primarily with the distribution of inflexional exponents (i.e. the structure of the inflexional paradigm itself) as opposed to the phonological realization of those exponents (compare Weinreich 1954 for a similar approach to phonological contrasts in a dialect continuum). For this reason, we select measures developed expressly in order to formalize the description of inflexional systems *qua* systems. The study results illustrate how such formalization facilitates transparent and objective comparisons between the different varieties, allowing points of similarity and divergence between linguistic varieties to be readily identified, and

providing a genuinely morphological perspective on the internal structure of the dialect continuum.

References

Bach, Xavier & Louise Esher. 2013. Morphological evidence for the paradigmatic status of infinitives in French and Occitan. In Dag Haug (ed), *Historical Linguistics*, 135–154. Amsterdam: Benjamins.

Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The Syntax-Morphology Interface. A Study of Syncretism. Cambridge Studies in Linguistics 109.* Cambridge: Cambridge University Press.

Baerman, Matthew. 2007. Syncretism. *Language and Linguistics Compass* 1(5). 539–551.

Bonami, Olivier & Gilles Boyé. 2002. Suppletion and dependency in inflectional morphology. In Frank van Eynde, Lars Hellan et Dorothee Beermann (eds.), *Proceedings of the 8th International HPSG Conference, Norwegian University of Science and Technology (3-5 August 2001)*.

Bonami, Olivier & Gilles Boyé. 2003. Supplétion et classes flexionnelles. *Langages* 37.152. 102–126.

Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. In Florence Villoing, Sarah Leroy & Sophie David (eds.), *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux*, 17–45. Paris: Presses Universitaires de Paris-Ouest.

Bonami, Olivier & Gregory T. Stump. 2016. Paradigm Function Morphology. In Andrew Hippisley & Gregory T. Stump (eds.), *Cambridge Handbook of Morphology*, 449–481. Cambridge: Cambridge University Press.

Boyé, Gilles. 2011. Régularités et classes flexionnelles dans la conjugaison du français. In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (eds.), *Des unités morphologiques au lexique*, 41–68. Paris: Hermes Science Publishing/Lavoisier.

Brun-Trigaud, Guylaine. 1990. *Le Croissant : le concept et le mot*. Doctoral thesis, Université Lyon III.

Dubois, Jean. 1967. *Grammaire structurale du français : le verbe*. Paris, Larousse.

Martinet, André. 1958. De l'économie des formes du verbe en français parlé. In Anna Granville Hatcher & Karl-Ludwig Selig (eds.), *Studia Philologica et Literaria in honorem L. Spitzer*, 309–326. Berne: Francke.

Stump, Gregory T. 2001. *Inflectional Morphology. Cambridge Studies in Linguistics 93.* Cambridge: Cambridge University Press.

Stump, Gregory T. & Raphael A. Finkel. 2013. *Morphological Typology, From Word to Paradigm. Cambridge Studies in Linguistics 138.* Cambridge: Cambridge University Press.

Stump, Gregory T. 2016. *Inflectional Paradigms. Cambridge Studies in Linguistics 149.* Cambridge: Cambridge University Press.

Tourtoulon, Charles de & Olivier Bringuier. 1876. *Limite géographique de la langue d'oc et de la langue d'oil*. Paris: Imprimerie nationale.

Weinreich, Uriel. 1954. Is a Structural Dialectology Possible? *Word* 10.2–3. 388–400.

Predicting cells in word-formation paradigms – a case study

Petr Kos and Jana Kozubíková Šandová

University of South Bohemia in České Budějovice

Czech Republic

1 Introduction

Paradigmatic approaches to inflectional morphology can successfully predict the form of cells in inflectional paradigms (see Ackerman & Malouf, 2013). In derivation, Bonami and Strnadová (2018) have shown that the content of cells can also be predicted in derivational paradigms on the basis of paradigmatic systems. More broadly in word-formation, however, it is not clear what the abstract cells correspond to and how their content can be predicted especially in paradigms which do not seem to belong to any paradigmatic system. This paper aims to suggest how the cells in word-formation paradigms, not only in derivational paradigms, can be defined and their content predicted in a case study based on a complementary application of an onomasiological analysis of coining new naming units (Dokulil, 1962, 1986; Grzega, 2007; Štekauer, 1998) with a semasiological analysis of word-formation paradigms (Boyé & Schalchli, 2016; Bauer, 2017; Bonami & Strnadová, 2018; Fradin, 2018, amongst others). We thus follow the genesis of a new lexeme (onomasiological perspective) and show how the abstractions over the existing lexemes in the mental lexicon of the coiner (semasiological perspective) determine the final form of the new one.

The structuring of abstract cells in inflectional and word-formation paradigms is fundamentally different because it is based on different functional and communicational requirements. In inflection, the structuring of cells is basically dictated by the needs of syntax and highly abstract morphological categories: if a speaker needs to form an inflectional form, they have a specific lexeme as their starting point, and syntax requires that the lexeme should be realized in a form which corresponds to one specific cell in a paradigm in compliance with the necessary grammatical categories. The analysis of inflection is thus a purely linguistic matter. However, the situation in word-formation is different: a speaker needs to name an extra-linguistic reality for which they cannot find an adequate lexeme in their mental lexicon. This initiates the process of naming, the starting point of which is not purely linguistic and should therefore be described from an onomasiological perspective.

2 Structural and lexical meanings

The key to the paradigmatic description of word-formation is the mutual relationship between the lexical and structural meanings (cf. *novelist* whose structural meaning could be described as *someone who is somehow associated with novels*, the lexical meaning being *a person who writes novels professionally*). The identity and structure of cells in word-formation paradigms are given by the structural meanings, which are abstractions over the lexical meanings of the existing lexemes. The creation of lexical meanings, nevertheless, begins in the very process of naming by mapping a specific onomasiological structure on some of the possible structural meanings, which are more general. Consequently, the lexical meaning should not be understood as a secondary idiosyncratic shift of the structural meaning, but it is a direct reflection of the onomasiological structure. Moreover, the existing lexical meanings are a source from which the structural meaning is abstracted.

3 Onomasiological categories

Dokulil (1962) distinguishes three different onomasiological categories, namely the modificational, the transpositional, and the mutational. Within the modificational onomasiological category the coiner merely adds a semantic feature to the existing lexical meaning of the word, e.g. the feature of diminutiveness or gender. Within the transpositional category, the creation of naming units is dictated by the needs of syntax. Within this category it is the phenomenal category, i.e. the word-class, that changes (e.g. *pale – paleness, to drink – drinking, nice – nicely*). The formation of naming units within these two onomasiological categories resembles inflection in that the existence of cells is given intra-linguistically and their form is highly predictable. However, within the mutational category, naming in the narrow sense, the existence of cells and their forms is not as clearly given.

4 Onomasiological process of naming

In the mutational category, the naming process starts with a concept to be named. The extra-linguistic referent is analysed and "both the more general, 'global' features and the more specific, 'local' features of a concept are processed" (Grzega 2005: 77).

The perception of salient features of the concept gives rise to an onomasiological structure. The local feature(s) become(s) an onomasiological mark and the global features become the onomasiological base. The local feature can either be static (a salient physical feature) or dynamic (an activity or relation to another entity), and in either case the local feature has a complex internal structure. This complexity of the local feature becomes reflected in the onomasiological structure which is the basis for the actual act of naming.

The onomasiological structure of the static local feature is

ASPECT (QUALITY) / PART FOR THE WHOLE

onomasiological mark   onomasiological base
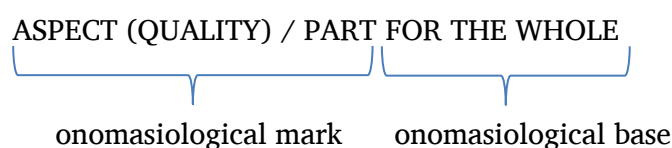
This, in fact, is double metonymy. The salient feature refers to one of the possible aspects of the referent – ASPECT FOR THE WHOLE, e.g. shape, colour, size, and at the same time this aspect refers to a part only or the referent as a whole - PART FOR THE WHOLE. The third part of this structure is the quality itself, e.g. what shape, what colour, or what size. So, for example, the onomasiological structure underlying the name redbreast is COLOUR (RED) / BREAST FOR THE BIRD.

The onomasiological structure of the dynamic local feature comprises, based on Dokulil (1962) and Štekauer (1998), is

DETERMINING / DETERMINED FOR THE WHOLE

onomasiological mark   onomasiological base
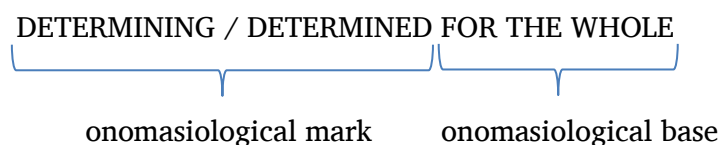
The determining constituent is an entity in a metonymical relation to the referent, and the determined constituent (a verb) expresses the type of the relation (an onomasiological connective) or an activity. For example, the underlying onomasiological structure for *bee-eater* is BEE / EAT FOR THE BIRD.

When searching for a linguistic form of this structure, the coiner first linguistically expresses the most salient member(s) of the structure, the initial salient expression. For the future morphological form of the newly coined naming unit it is decisive which part of the structure is expressed and which word-class the initial salient expression belongs to. For instance, the same colour may be expressed literally by an adjective (*black*) or metaphorically by a noun (e.g. *devil* or *soot*).

5 The role of paradigms in the onomasiological process

For this initial linguistic expression, the coiner searches for a suitable structural meaning – referred to as 'word-formation cell' – by scanning the available structural meanings, which are abstractions over the existing lexicon and its paradigms (semasiological perspective). The onomasiological structure, however, does not match one ideal cell in a paradigm, as in inflection, but the coiner has a range of choices from a number of cells in different word-formation paradigms. The table below shows an example of the possible structural meanings and its paradigms for a static onomasiological structure expressing colour of an organism:

Onomasiological structure (static):		COLOR / PART / ORGANISM	
Structural meaning:	B has the colour	B has the colour like N	B whose part has the colour
Paradigms: (an example for each)	*white agaric*	*devil*	*black cap*
		cornflower boletus	*orange-cap boletus*
		snowy inkcap	*yellow-stemmed mycena*
		clouded agaric	
		brick-colored galera	

The second table shows an example of the possible structural meanings and its paradigms for a dynamic onomasiological structure based on sound produced by an organism:

Onomasiological structure (dynamic):		SOUND / PRODUCE / ORGANISM	
Structural meaning:	B producing sound	B whose sound is like N's	B whose sound is imitated
Paradigms: (an example for each)	*screech*	*lamb*	*whew*
	screecher	*reeler*	*whewer*
	whistling duck	*cat gull*	*jar bird*
	shriek owl	*bell ringer*	

The choice of the structural meaning is determined by the choice of the initial salient expression, and for the choice of the paradigm, in our conception, it is its size in the mental lexicon of the coiner that plays the decisive role.

6 A case study

Bonami and Strnadová (2018) based the predictability of the form of cells in derivational paradigms on the analysis of paradigmatic systems they enter. However, the paradigms employed in naming natural organisms do not seem to belong to any paradigmatic system, so we assume that the form of the cell can be predicted, as stated above, from the frequency of the occurrence of the given paradigm in the mental lexicon of the coiner. As the mental lexicon varies in people with different experience, we chose synchronically transparent English names for mushrooms coined by mycologists and names for birds coined by the common folk. From these we selected those that are motivated by one salient feature only.

In these names we identify all the possible structural meanings that express the same onomasiological structures, namely those for colour (static feature) and sound and food (dynamic features), the former within the names for both mushrooms and birds, and the latter two within the names for birds only, and then calculate the frequency of the occurrence of these structural meanings and corresponding paradigms within the corpus. The corpus comprises approximately 1,000 names for mushrooms and 1,000 names for birds.

The results show that the structural meanings, i.e. the cells, deriving from "literal" initial salient expressions, e.g. colour expressed by a colour adjective or sound by onomatopoeia, are more frequent than those expressed by a metaphor (i.e. the noun). Also, in all structural meanings deriving from the linguistic salient expression realised by a noun, the most frequently occurring structure is N + N – in names motivated by colour being in competition with N-y + N, N-ed + N, N-colored + N structures and in names motivated by food being in competition with N + V-er (synthetic compound) structure.

We believe that in this way we are able to suggest the level of predictability of the form of the word-formation cell for the given ontological type of concepts and offer a model for measuring the predictability of word-formation cells in a language as a whole.

References

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: the low conditional entropy conjecture. *Language*, 89, 429–464.

Bauer, Laurie. 2017. *Notions of paradigm and their value in word-formation.* Paper presented at ParadigMo 2017, First Workshop on Paradigmatic Word Formation Modeling, Toulouse.

Bonami, Olivier & Jana Strnadová. 2018. Paradigm structure and predictability in derivational morphology. *Morphology*, Springer, 1 –31.

Boyé, Gilles & Gauvain Schalchli. 2016. The status of paradigms. In: A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (206–234). Cambridge: Cambridge University Press.

Dokulil, Miloš. 1962. *Tvoření slov v češtině 1, teorie odvozování slov*. Praha: ČSAV.

Dokulil, Miloš. 1986. III Tvoření slov. In: Petr, J. et al., *Mluvnice češtiny 1*. Praha: Academia.

Fradin, Bernard. 2018. Paradigms and the role of series in derivational morphology. *Lingue e linguaggio*, XVII. 2, Bologna: Mulino, 155-171.

Grzega, Joachim. 2005. Onomasiologial approach to word-formation. A comment on Pavol Štekauer: Onomasiological approach to word-formation. *SKASE Journal of Theoretical Linguistics* 2(2), 76–81.

Grzega, Joachim. 2007. Summary, Supplement and Index for Grzega, Bezeichnungswandel, 2004, *Onomasiology Online 8* < http://www1.ku-eichstaett.de/SLF/EngluVglSW/OnOn-Total.pdf >

Štekauer, Pavol. 1998. *An onomasiological theory of English word-formation*. Amsterdam: John Benjamins.

The prefixal template of Umo^{n}ho^n

Julie Marsault

Mondes Iranien et Indien; CRLAO; ED 268

Umo^{n}ho^n (Omaha) is a native american language from the Siouan family spoken in eastern Nebraska, United States. As a head-marking language, grammatical relations are encoded by verbal morphology (with split-S indexation), while case marking is non-existent. The verbal prefixal chain presents templatic features (Bickel & Nichols, 2007) such as, for example, an arbitrary ordering of prefixes, or the encoding of the same category (here, person markers) in different slots. Moreover, Umo^{n}ho^n verbal morphology is particularly complex because of the frequent merging of different prefixes into opaque surface forms.

Table 1 summarizes the verbal template of Umo^{n}ho^n verbs, focusing on prefixes and preverbal particules. Personal inflection appears in red : it corresponds to agentive and patientive indexation markers. Up to two of them can appear, on top of the derivational prefixes from the same slot. For the sake of simplicity, only the prefixes from the regular paradigm are listed. Table 1 thus presents a template somewhat simplified, because several tables would be necessary to describe the exact ordering of prefixes in all their possible combinations (see Marsault, 2016, 47).

Table 1 highlights several features of Umo^{n}ho^n verbal system that are cross-linguistically rare:

1. Person markers often appear closer to the root than derivational morphemes, challenging the Split Morphology hypothesis.

2. Person markers can appear in different slots (1, -3, -5 or -6) according to which derivational morpheme is present in the verb form.

3. There is sometimes a multiple exponence (Harris, 2017) of agentive person markers.

4. Most of the derivational prefixes presented in Table 1 are found lexicalized with some roots, often creating discontinuous stems.

All of these are characteristic features of the Siouan family, and have been surveyed at family level (Rankin et al., 2003) or documented in particular languages (Helmbrecht & Lehman, 2008; Kasak, 2019). In my presentation, I focus on the so called 'dative' prefix ***gí-*** (slot -5). This prefix functions productively as an applicative marker introducing a beneficiary, a maleficiary, or a recipient. It is also lexicalized in a few cases. I will illustrate the features listed above with data including ***gí-***.

1 Split Morphology Hypothesis

We see in Table 1 that personal inflection can appear closer to the stem than derivational prefixes or 'preverbs' (the latter can be considered as the first part of a discontinuous stem). An example of a verbal form displaying such characteristics is presented in (2) below. The insertion of inflectional morphemes closer to the root than derivational morphemes is unexpected (Bybee, 1985, 33).

Table 1: Template of Umoⁿhoⁿ verbal prefixes

Indef. -7	Oblique -6	Dative -5	Outer Instr. and preverb -4	Person markers -3	Reflexive Possessive -2	Inner Instr. -1	Root 1 0	Root 2 1
wa-	*í-*	*gí*	*má-*	*a-*	*gi-*	*tha-*	Root 1	*the*
	á-	+ person	*mú-*	*oⁿ(g)-*	*ki(g)-*	*thi-*		*tha*
	u-	markers	*ná-*	*tha-*		*ba-*		*thiⁿ*
	ithá-		preverb	*oⁿ-*		*bi-*		*thiⁿtha*
	uthú-			*wa-*		*ga-*		+ agentive
	+ person			*thi-*		*noⁿ-*		person
	markers			*wi-*				markers

2 Changes of slot

First of all, example (1) illustrates the respective order of the dative prefix and the 'outer instrumental' prefixes (Rankin, unp.). The instrumental prefixes (slots -1 and -4) specify by which means a process takes place, like 'by fire', 'by shooting', etc. As expected from Table 1, the prefix *gí-* (in blue) stands on the left side of the instrumental.

(1)   **thiⁿge**   **ná-thiⁿge**   **gí-na-thiⁿge**
        be.gone   INSTR.fire-be.gone   BEN-INSTR.fire-be.gone
        it is gone   it was burnt to nothing   it was burnt to nothing on him (Dorsey, 1890)

Examples (2) and (3) illustrate two possible loci of person markers. When a stem comprises a root and any derivational prefix between slot -1 to slot -4, the person markers are inserted on slot -3, i.e. on the right edge of outer instrumentals, as in Example (2).

(2)   **té**   **ma**   **mú-wa-hega-b-azhi**
        buffalo ANIM.PL INSTR.shoot-P3PL-(be.a.)few-PL-NEG
        They shot down many of the buffaloes. (Dorsey, 1890, 350.6 / Big Elk)

On the other hand, the locus of indexation is realized in the same slot as the prefix *gí.* Underlyingly, the indexation takes place before *gí-*, as shown in example (3), and the prefixes merge in the following way : the initial <*g*> of the dative prefix lenites, then the two contiguous vowels undergo monophthongization. The morphophonological merging, together with other features presented in the following section, are the reason why the indexation markers and the prefix *gí-* are integrated in the same slot.

(3)   **gíboⁿ**   **wé-boⁿ**
        gí-boⁿ   ***wa-gí-boⁿ***
        BEN-call   P3PL-BEN-call
        he called him   he called them (Saunsoci & Eschenberg, 2016, 77 / Alice Saunsoci)

In Example (4), it can be seen that the prefix *gí-* has priority over outer instrumentals in determining the locus of inflection. Again, *gí-* and the indexation marker merge into an opaque surface form, and this form appears to the left side of the instrumental.

(4)   **wénathiⁿgai**
        wa-gí-na-thiⁿge-í
        P1PL-BEN-INSTR.heat-be.gone-PL
        it was burnt to nothing for us (Dorsey, 1890, 498.4 / Maxpíya-xága)

In this manner, the locus of inflection has a conditional positioning (Bonami, 2014), conditioned by the presence or absence of the dative prefix. Unlike the examples presented by Bonami (2014), though, the prefix *gí-* is not purely inflectional and cannot be considered as a syntactic feature.

3   Multiple exponence of agentive prefix

Some verbs derived with *gí-* show multiple exponence of the agentive prefix, as a direct result of the reanalysis of merged forms as a specific conjugation paradigm. **Baxú** 'to write it' is called a 'syncopating verb' (Koontz, 2001) because the indexation of the 1st person singular and 2nd person agentive arguments is realized through the alteration of the initial consonant of its stem, as we see in (5a). When this verb is derived with *gí-*, the agentive argument is realized both by the initial consonant alteration and by the merged form **thé-** (< *tha-gí-*). This corresponds to an 'accidental multiple exponence' (with exponents in a subset relationship) according to Harris (2017)'s classification. This double exponence could be an instance of hybrid forms described by Haspelmath (1993) as an intermediate step towards externalization of inflection.

(5)   a.   **baxú** 'write it'   →   **shpáxu** 'you write it'
           write   A2.write

   b.   **gí-baxu**   'write it to him' → **iⁿ-thé-shpaxu**   'you write it to me'
        BEN-write   P1SG.BEN-A2.BEN-A2.write

Inflection or derivation?   The merging of the dative prefix with indexation markers, and the double exponence it entails, question the status of the dative prefix as a derivational morpheme. Exemple 6 shows further support for analyzing *gí-* as an inflection marker: When a verb with the prefix *gí-* indexes both A and P arguments, each of the corresponding indexation markers is realized in its merged form with *gí-*, unlike what is expected.

(6)   underlying form   *__a-wa-gí-boⁿ__
                                 A1SG-P3PL-BEN-call
      expected surface form   †__a-wé-boⁿ__ (with margin of **wa-gí-**)
      attested surface form   **e-wé-boⁿ**

      I called them (Saunsoci & Eschenberg, 2016, 77 / Alice Saunsoci)

Thus, forms like **é-** (< *a-gí-*), **wé-** (< *wa-gí-*), **íⁿ-** (< *oⁿ-gí-*), etc, have been reanalysed as a conjugation paradigm (Koontz, 1989).

4   Lexicalization

Finally, the prefix *gí-* is sometimes lexicalized with the stem on which it occurs, and thus forms a verb which is not analysable as a complex word.

(7)   **gíthe** to be glad   (from **gí-** and ***the**)
      uíkoⁿ to help him/her (< ***u-gí-koⁿ**, from **gí-** and ***u-koⁿ**)

In such examples, *gí-* retains the morphophonological characteristics of the derivational prefix, but otherwise cannot be analysed as such, because the root does not exist without it. As Helmbrecht & Lehman (2008) remark, this feature is common to all verbal derivational prefixes, and makes the morphological analysis of Siouan verbs particularly complex.

All the morphological specificities described here in relation to the prefix *gí-* applies to other prefixes as well ; they are pervasive in Umoⁿhoⁿ verbal morphology. In particular, the 'oblique' prefixes listed in slot -6 display similar features: they are involved in conditional positioning, they entail misaligned positioning (Bonami, 2014) of person markers and double exponence of agentive markers, and they have undergone more lexicalization processes than the prefix *gí-*.

Gloss

1, 2, 3 first, second, third person ; A Agentive argument ; ANIM animate ; BEN benefactive ; INSTR.fire instrumental prefix denoting action of fire ; INSTR.shoot instrumental prefix denoting action done by shooting ; NEG negation ; P Patientive argument ; PL plural ; SG singular.

References

Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language Typology and Syntactic Description, Volume III: Grammatical Categories and the Lexicon*, 169 – 240. Cambridge: Cambridge University Press.

Bonami, Olivier. 2014. La structure fine des paradigmes de flexion : Études de morphologie descriptive,théorique et formelle. Habilitation à diriger des recherches. Université Paris 7 - Denis Diderot.

Bybee, Joan. 1985. *Morphology: a study of the relation between meaning and form* Typological Series in Language. John Benjamin Publishing Company.

Dorsey, James Owen. 1890. *The Ȼegiha Language*, vol. Contributions to North American Ethnology, Vol 6. Washington, D. C., U.S. Government. 6.

Harris, Alice C. 2017. *Multiple exponence*. Oxford University Press.

Haspelmath, Martin. 1993. The diachronic externalization of inflection. *Linguistics* 31. 279–309.

Helmbrecht, Johannes & Christian Lehman. 2008. Hocank's challenge to morphological theory. In *Lessons from Documented Endangered Languages*, vol. 78, 271–316. John Benjamin Publishing Company.

Kasak, Ryan. 2019. *Affix ordering and templatic morphology in Mandan*: Yale University dissertation.

Koontz, John. 1989. Ordering of morphological rules in Omaha-Ponca. American Anthropological Association.

Koontz, John. 2001. Verb paradigms. Manuscript.

Marsault, Julie. 2016. *Le système verbal du omaha : Formes et fonctions des préfixes verbaux*. Université de Paris 3 - Sorbonne Nouvelle MA thesis.

Rankin, Robert. unp. Instrumental prefixes in the Siouan Languages. Manuscript.

Rankin, Robert, John Boyle, Randolph Graczyk & John Koontz. 2003. A synchronic and diachronic perspective on "word" in siouan. In Robert Dixon & Alexandra Aikhenvald (eds.), *Word : A cross-linguistic typology*, 180–204. Cambridge: Cambridge University Press.

Saunsoci, Alice & Ardis Eschenberg. 2016. *500 + Verbs in umóⁿhoⁿ (Omaha) - Doing things in the umóⁿhoⁿ way*.

Romance compounding and language contact: the origin and spread of the pattern *vert bouteille* 'bottle-green'

Sara Matrisciano

WU Vienna

Franz Rainer

WU Vienna

1. Introduction

While a great deal of continuity exists between Latin and Romance in the realm of suffixation, compounding is generally believed to have been created anew in Romance through reanalysis of syntagmatic patterns. There can be no doubt that this view is correct, at least to a large extent. Bold claims for Latin-Romance continuity have been put forward in the 1990s for three patterns of compounding – VN (Bork 1990), NN (de Dardel 1999), and N-i-A (the *pettirosso* type, de Dardel & Zamboni 1999) – but they have not met with general approval. In the light of the fact that Romance patterns of compounding have arisen anew through reanalysis, a process that necessarily takes place in one speech community only, it must come as a surprise that the compounding patterns of present-day Romance languages are so highly similar. It suffices to read the various contributions to the special issue of *Probus* "On Romance compounds" (Scalise & Masini 2012) to get a picture of the high degree of uniformity.

2. Research Gap & State of the Art

The abundant literature on Romance compounds has largely failed to inquire why this is so. A priori, two possibilities come to mind. One could argue that the similarities among Romance patterns of compounding are due to similarities concerning the syntagmatic patterns that gave rise to the compounds, the mechanisms of reanalysis, as well as the conceptual needs of the different speech communities. Alternatively, it could be argued that the similarities are the result of language contact, of a continuous exchange of compounds among the Romance languages. Since these two explanations are not mutually exclusive, a combination is also a possible option. What is less relevant in this case, in contrast to prefixation and suffixation, is re-Latinization, since the compounding patterns of present-day Romance languages did not exist in Classical Latin.[1]

3. The pattern *vert bouteille*

In our poster, we will concentrate on one specific pattern, viz. colour compounds of the type *vert bouteille*. The grammatical status of these expressions has given rise to controversial claims in the literature (cf. García-Page 2009). While most Romance scholars have always treated these formations as combinations of two nouns, Masini & Scalise (2012) treat them as AN compounds. One problematic fact in particular speaks against this interpretation: if the first member of the pattern were an adjectival head, we should expect it to agree with the head of

---

[1] Note, however, that the uniformity of A-A patterns (e.g. the *physico-mathématique* type) is indeed the result of re-Latinization, albeit on the basis of a Neo-Latin pattern of compounding that did not yet exist in Classical Latin (cf. Grossmann & Rainer 2009).

the noun phrase, which is not the case (e.g. *chemise vert bouteille / *verte bouteille*).[2] More controversial is the question whether we are dealing with compounds or phrases. Both options have been defended by numerous scholars in the literature. García-Page (2009: 61-62) shows that in fact they have traits in common with both compounds and phrases, but eventually seems to prefer a phrasal account. He surmises that *chemise vert bouteille* is a kind of elliptical shortening *chemise coleur vert bouteille*, just like *chaise rococo* is short for *chaise style rococo*. Whatever the merits of this account in synchronic terms, we will show that this is not the way in which the *vert bouteille* pattern has arisen in diachrony. The issue of the delimitation of syntax and compounding in Romance NN constructions is, of course, an important one that has ramifications far beyond colour terms. In addition, it is a highly theory-dependent question that it does not make sense discussing outside specific frameworks of syntax and morphology. Construction morphologists might even argue that it is a pseudo-issue, since its constructions all the way down... A poster is not the place to take up these questions in all their complexity. Therefore, we will simply assume that we are dealing with compounds here.

4. Preliminary results

The main goal of our poster is to show that massive borrowing was indeed the decisive factor for uniformity among Romance languages with respect to the *vert bouteille* pattern. We will try to identify the oldest of such examples for the major Romance languages, from Romanian to Portugese (the pattern, tellingly, seems to be absent from the dialects). At the present stage of our investigation, the oldest example is French *jaune paille* in Oudin's dictionary from 1607. French examples, however, only became more abundant in the course of the 18th century. The other Romance languages seem to have borrowed the pattern from French: Italian early on, the Ibero-Romance languages somewhat later, and Romanian only quite recently (after World War II). The picture is not without complications, however. Italian dictionaries, for example, feature some examples of *verdeterra* 'verditer' from around 1400 until the 16th century, which on closer inspection, however, turns out to have been an adaptation of Middle French *verd de terre* (just like English *verditer*, by the way).[3] It is not the leader word of this pattern.

5. Outlook

The *vert bouteille* pattern is not an isolated case. Other patterns are certainly amenable to a similar explanation via inter-Romance borrowing, but this remains to be determined by detailed analyses. Most Romance works on compounding, unfortunately, tend to concentrate on one language only, a notable exception being Ciobanu and Hasan's monograph on Romanian, which acknowledges the deep influence that foreign languages have played in the development of compounding in this language. Much work remains to be done to pin down for each pattern and sub-pattern of compounding the exact part that the common starting point, cross-Romance contact and re-Latinization have had in creating the degree of uniformity that can be found.

---

[2] In substandard varieties, one can occasionally find such combinations (e.g. Sp. *chicas rubias platino* 'platinum blonde girls'). They seem to be a relatively recent phenomenon.

[3] The term that thrived in Italian was *terra verde*. Note that earth is not a prototypical substance for green colour or some shade of green, as straw or a lemon are for the colour yellow, or blood or a cherry for the colour red.

6. References

Bork, Hans Dieter. 1990. *Die lateinisch-romanischen Zusammensetzungen Nomen + Verb und der Ursprung der romanischen Verb-Ergänzung-Komposita*. Bonn: Romanistischer Verlag.

Ciobanu, Fulvia & Hasan, Finuţa. 1970. *Formarea cuvintelor în limba română*. Vol. 1: *Compunerea*. Bucureşti: Ed. Academiei Române.

de Dardel, Robert. 1999. Composés rectionnels nominaux nom + nom en protoroman. In *Probus* 11. 177–208.

de Dardel, Robert & Zamboni, Alberto. 1999. L'interfixe –I– dans les composés protoromans. Une hypothèse de travail. In *Revue de Linguistique Romane* 63. 439–469.

García-Page, Mario. 2009. Los nombres de color del español: de su composición y sus propiedades gramaticales. *Revue Romane* 44(1). 47–66.

Grossmann, Maria & Rainer, Franz. 2009. Italian adjective-adjective compounds: between morphology and syntax. *Italian Journal of Linguistics* 21(1). 71-96.

Masini, Francesca & Scalise, Sergio. 2012. Italian compounds. In *Probus* 24(1). 61–91.

Scalise, Sergio & Masini, Francesca (eds.). 2012. *On Romance compounds*. Special issue of *Probus*, 24(1). Berlin/Boston: de Gruyter.

When a causative could hide a plural marker:
A quest for the origins of the causative in Andi
(Nakh-Daghestanian)

Neige Rochant
Université Sorbonne Nouvelle

This paper is a first attempt at reconstructing a Proto-Andi(c) causative marker. Andi is an underdocumented language from Russia belonging to the Avar-Andic-Tsezic branch of the Nakh-Daghestanian family. Three of its nine dialects have been described in grammar sketches (Dirr, 1906; Sulejmanov, 1957; Cercvadze, 1965; Salimov, [1968] 2010), while a corpus in two different dialects amounting to 32,141 words is available. However, there is no Andi dictionary nor any work dedicated to the history of any aspect of Andi. The present study is based on my fieldwork data of two previously undescribed dialects of Andi, aforementioned references and works on related languages.

Nakh-Daghestanian causatives are often auxiliaries meaning 'make' or suffixes that can be traced back to that verb (Authier, 2018), which is typologically common (Kulikov, 2001, 895). Andic and Tsezic languages, however, all display morphologically opaque causative suffixes, which points to their archaicity and raises the question of their origins. The Andi causative suffix can derive bivalent or trivalent transitive verbs from any verb root, or be applied to some verbs without increasing their valency to alter the semantics of the sentence.

1 A morphological alternation in need of an explanation: what could be the historical underlying consonant of the causative?

Whereas most TAME markers attach either to a past or a nonpast thematic suffix (separated from the root by a dot), the Andi causative is a suffix attached directly to the verbal root. It is subject to a morphological alternation between [-*ol*] and [-*oɫ*], following a slightly different distribution form one dialect to another. In the Gagatl dialect, it is in [-*ol*] when followed by the nonpast thematic suffix /d/ (used in habitual, infinitive and future)[1] and in [-*oɫ*] in all other forms[2] (cf. table 1).

To explain this alternation, Dirr (1906, 54) proposed that the causative suffix is historically composed of a morpheme -*ol* and a grammaticalized form of the verb 'do', whose paradigm is heteroclite in the Andi dialect: in forms using the nonpast thematic suffix /d/, its root is in *i*- (INF *i.d-u*, HAB *i.d-o*, FUT *i.d-ja*). In all other forms, it is in *ih*- (AOR *ih.i*, IMP *ih-o*). After morphological reduction, the causative forms composed of -*ol* + -*ih*- and those composed of -*ol* + *id*- were realized as -*oɫ*- and -*ol.l*- respectively[3]. Support to this hypothesis is provided by a comparison with the realizations of denominative stems formed by the inchoative suffix /-*ɫ*/ < **ɫ* (e.g. *tamaʃa-ɫ*- surprise-INCH- 'be surprised'). Indeed, in all Andi dialects, these realizations differ systematically from those of the causative forms despite identical phonotactic properties, which substantiates a reconstruction of the historical underlying form of the causative conso-

---

[1] The nonpast thematic suffix /d/ gets assimilated by a preceding sonorant.

[2] The perfect form in CL-*iqχ-ol-dːu* easily reads as a realization of **CL-iqχ-oɫ-dːu* (< **CL-iqχ-oɫ.i-dːu*), owing to the phonological rule prohibiting CVC syllables other than CVL and CV*b* (Moroz, 2017).

[3] Following this hypothesis, the transitive imperative form in Zilo would have resulted from an analogical replacement of the causative consonantic element *ɫ* by *ll*.

Table 1: Simple and causative paradigms in the Gagatl dialect (Salimov, [1968] 2010, 212-213, 219-237)

Value	CL-*iqχ*- 'cut'	
	[- CAUS]	[+ CAUS]
AOR	CL-*iqχ.i*	CL-*iqχ-oɬ.i*
PF	CL-*iqχ.i-d:u*	CL-*iqχ-ol-d:u* < *CL-*iqχ-oɬd:u* < *CL-*oɬ.i-d:u*
PROG	CL-*iqχ.i-rado*	CL-*iqχ-oɬ.i-rado*
HAB	CL-*iqχ.id-o*	CL-*iqχ-ol.l-o*
INF	CL-*iqχ.id-u*	CL-*iqχ-ol.l-u*
FUT	CL-*iqχ.id-ja*	CL-*iqχ-ol-ja* < *CL-*iqχ-oɬ.l-ja*
IMP(TR)[1]	CL-*iqχ-o*	CL-*iqχ-oɬ-o*

[1] There are two imperative suffixes, whose distribution is conditioned by the transitivity of the construction.

nant as different from the inchoative, hence **l* rather than **ɬ*[4].

2 Infixation and vowel deletion: was the vocalic element of the causative really part of it?

The vocalic element of the causative is also problematic. Maisak (2016, 1) noticed that the progressive causative form was in **root**-*orallo* in the Rikwani dialect instead of the expected ****root**-*oɬ.i-rado* (table 1). Following Maisak (2016, 1), **root**-*orallo* can be segmented into **root**-*o-ra-l-/d/o*, suggesting that the progressive morpheme is historically split into *-ra* and *-do*. This analysis, supported by the existence of a habitual suffix in *-o* which attaches to the thematic /d/ (table 1) and a present suffix *-ra* in the Muni dialect (personal fielnotes), suggests that the causative morpheme is also split, since the segment *-ra* intervenes between the vowel and the consonant of the causative.

Drawing on a semantic and morphological discussion of forms using the suffix *-ra* in different dialects, I put forward the hypothesis that forms in **root**-*orallo* result from **root**.PST-*rallo* by analogical replacement of the PST vowel by *o* (associated with causative forms). Following this, the only element historically expressing causativity in Rikwani progressive causative forms in **root**-*o-ra-l.l-o* would be the first *-l*. Consequently, the vocalic element *o* of the modern causative suffix *-ol* is either dropped after *-ra* in **root**-*o-ra-l.l-o* or has never existed in these forms. I will describe contexts where comparable morphemes exhibit the same reduction behaviour.

First, in the Muni dialect of Andi (mutually inintelligible with presently discussed dialects), the causative suffix /-ot/ is reduced to its consonant when applied to vowel-ending roots (which no longer exist in other dialects). Indeed, *ts'a*- 'drink', which will be argued to be vowel-ending based on an analysis of its paradigms in Muni and other Andi dialects, forms a causative in *ts'a-t-*, to be contrasted with causative forms formed upon a consonantal root like *k'am-ot*- (eat-CAUS). The second noteworthy morpheme exhibiting the same behaviour is the transitive imperative suffix, which is the same vowel as the causative /-o/ and also the only marker attaching directly to the verb root like the causative. Interestingly, this vowel disappears when followed by the attenuative suffix *-ej* in the Andi dialect (cf. corpus data of Magomedova (2010)), and when

---

[4]At the same time, this comparison dismisses a tempting phonological interpretation of the distribution posing an underlying /oɬ/ whose fricative would become sonorant when followed by a consonant, to avoid a CVC structure.

following a vowel-ending root in the Muni dialect[5]. By suggesting that the transitive imperative suffix */-o/* is identical to the *-o* of the causative from a historical morphological perspective, these cases also add up to examples of reduction of this morpheme.

Then, I will show how data from the the Muni dialect suggest that this vowel */-o/* is rather non-existent than dropped in those contexts, and will present comparative data from the Tsezic languages which support an analysis of this vowel as historically epenthetic. The absence of the causative or imperative vowel *o* could be interpreted as an elision in contact with a preceding or following vowel if instances of a similar elision were attested in Andi. However, no such elision has been found so far, including in a context in the Muni dialect where it could and should occur for this hypothesis to be convincing: when attached to a vowel-ending root, the Muni infinitive suffix *-u* is desyllabified to *-w* (e.g. *ts'a* 'drink' INF *ts'a-w*) rather than being elided (which could be counterbalanced by adding an auxiliary in the infinitive for the form to be recognized as infinitive in synchrony, as for the imperative of the same verb). The fact that this vowel close to *o* is not elided, though in the same phonotactic context as the causative and imperative, discredits the hypothesis of an elision for these morphemes. Moreover, the Tsezic languages display causative suffixes in *-l* or *-r* (Alekseev, 1988, 167) which are very likely to be cognates of the Andi causative *-ol* judging from sound correspondences (Bokarev, 1959, 279; Gudava, 1964, 163). The fact that these cognates are strictly consonantal substantiates a reconstruction of the Proto-Andi causative morpheme as strictly consonantal as well, hence suggesting that the vowel *o* present in the modern causative suffix is historically epenthetic. I propose that it was originally added only when needed to prevent forbidden CVC structures (i.e. all other than CV*l* and CV*b*), before being generalized to all phonetic contexts (except aforementioned cases, considered as instances of fossilization).

3   Could the causative be related to the plural marker?

The reconstructed form **-l* of the first component of the Proto-Andi causative suffix matches the plural marker. The possibility of a relationship between the Andi causative and plural markers was evoked (but not argued for) by Cercvadze (1965, 268). I will show that this hypothesis is relevant both for Andi and a typological perspective.

The modern Andi plural suffix *-l* is mostly used on nominals (e.g. And. *boʃi* PL *boʃi-l* 'young of an animal'), but it also optionally attaches to a few forms of the finite verb paradigm as an agreement marker indicating a plural S/P argument[6], including the transitive imperative form in the Andi dialect (Dirr, 1906, 52), e.g. *ts'ad-o*! 'drink (this)!' PL *ts'ad-o-l*! 'drink (these)!'. Building on the hypothesis that the transitive imperative suffix */-o/* is identical to the *-o* of the causative from a historical morphological perspective, the fact that this specific form can form a plural with the exact same pattern as the first component of the historical causative (*ts'ad-o-l*! 'drink (these)!' vs. **ts'ad-o-l i.d-u* 'to make sme drink') is striking. It suggests that the first component of the historical causative could historically correspond to the plural marker *-l*, (which dates back from Proto-Avar-Andic-Tsezic, cf. Alekseev (1988, 183)).

In typology, several non-related languages show homonymy between causative forms and plural participant verb forms, e.g. Tiipay (Yuman, Miller (2011, 109–110)). Among the two causative morphemes reconstructed by Voeltz (1977) for Proto-Niger-Congo, one also marks agent plurality. It is noteworthy that both of these markers also have intensive and frequentative semantics, which are tightly connected to causativity both cross-linguistically (Aikhenvald,

---

[5]These forms, which coincide with the bare root, are no longer recognised by the Muni speakers as imperative and are hence augmented by the imperative verb *hiʔó!* 'come!', e.g. vowel-ending root *ts'a* 'drink' forms a periphrastic imperative in *ts'a hiʔó!* 'drink!' (lit. 'come drink!').

[6]Number agreement with S/P is otherwise indicated in the prefixal area of some verbs.

2011; Kulikov, 2001, 894) and in Andi (where causativization of some verbs adds an intensive meaning rather than increasing their valency). Intensive meanings could thus be the intermediate link on the chain of the semantic shift from plurality to causativity. This hypothesis finds support in the later suffixation of the verb 'do' to the marker -(o)l, which could have been motivated by a need of disambiguation at a stage when -(o)l had become ambiguous between intensive and causative meanings. This scenario substantiates the hypothesis of a semantic shift from plurality to causativity. I will conclude by reminding that the first component of the composite Proto-Andi suffix *-(o)l + 'do' is shared by other Andic-Tsezic languages and propose a Proto-Avar-Andic-Tsezic causative suffix *-l related to the plural marker.

References

Aikhenvald, A. Y. 2011. Causatives which do not cause: non-valency-increasing effects of a valency-increasing derivation. In A. Y. Aikhenvald & R. M. W. Dixon (eds.), *Language at large: Essays on syntax and semantics*, 86–142. Leiden: Brill.

Alekseev, M. E. 1988. *Sravnitel'no-istoričeskaja morfologija avaro-andijskix jazykov [Historical comparative morphology of Avar-Andic languages]*. Moscow: Nauka.

Authier, G. 2018. Not so stable: valence changing strategies in east caucasian over time. Talk at the 51st Annual Meeting of the Societas Linguistica Europaea.

Bokarev, Evgenij Alekseevič. 1959. *Cezskie (didojskie) jazyki dagestana [The Tsezic (Didoic) languages of Daghestan]*. Moscow: AN SSSR.

Cercvadze, I. I. 1965. *Andiuri ena [the Andi language]*. Tbilisi: Metsniereba.

Dirr, A. M. 1906. *Kratkij grammatičeskij očerk andijskago jazyka [A short grammar sketch of Andi]* (Sbornik materialov dlja opisanija mestonstej i plemën Kavkaza 32). Tbilisi: Upravlenie Kavkaskogo Učebnogo Okruga.

Gudava, T. E. 1964. *Konsonantizm andijskix jazykov. Istoriko-sravnitel'nyj analiz [The consonant system of the Andic languages. Comparative-historical analysis]*. Tbilisi IJaz AN GSSR edn.

Kulikov, L. I. 2001. Causatives. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds.), *Language typology and language universals: An international handbook*, vol. 2, 886–898. Berlin, New York: Walter de Gruyter.

Magomedova, P. A. (ed.). 2010. *Dunjalla baχunnirsːi χarɢiltɬol ɢwanab mitsʼːillasːol muχal [Fairytales of peoples of the World in the Andi language]*. Makhachkala: Nurul' iršad. Translation by M. A. Alisultanova.

Maisak, T. A. 2016. Glagol'naja sistema rikvaninskogo govora andijskogo jazyka : formoobrazovanie [The verbal system of the Rikwani dialect of Andi: inflectional morphology]. Handout for the 1st Andi seminar. IJaz RAN. Moscow.

Miller, A. 2011. *A Grammar of Jamul Tiipay* (Mouton Grammar Library 23). Berlin, New York: Walter de Gruyter.

Moroz, G. A. 2017. Segmentnaja sistema i slogovaja struktura zilovskogo andijskogo [The segmental system and syllabic structure of Zilo Andi]. Handout for the 16th Andi Seminar. IJaz RAN. Moscow.

Salimov, X. S. [1968] 2010. *Gagatlinskij govor andijskogo jazyka [The Gagatl dialect of Andi]*. Makhachkala IJaLi RAN edn.

Sulejmanov, Ja. G. 1957. *Grammatičeskij očerk andijskogo jazyka (po dannym govora s. Rikvani) [A grammar sketch of Andi (dialect of Rikwani)]*. URSS: IJaz AN SSSR dissertation.

Voeltz, E. F. K. 1977. *Proto Niger-Congo Verb Extensions*. Los Angeles: University of California dissertation.

On the influence of creativity upon the formation of complex words

Pavol Stekauer, Livia Kortvelyessy, Pavol Kacmar

Pavol Joze Šafárik University, Košice, Slovakia

Despite extensive morphological research into complex-word formation in recent decades and despite its obvious relevance, language-users' creativity, as a psychological variable, has not been studied as a word-formation factor yet. This paper aims to fill this gap and to address how language users of unequal creativity characteristics coin new complex words.

The paper reviews basic theoretical principles, describes the method and presents some of the preliminary results. The sample consisted of 309 undergraduate students (61% were females). The study itself consisted of two stages, Creativity stage, and Word formation stage.

I) In the first stage, the informants undertook the creativity assessment. Widely used test, Torrance Test of Creative Thinking (TTCT), figural form, was used (e.g. Kim, 2006, 2017). In TTCT, the informants are asked to draw or complete pictures and name them in three tasks. Based on their creations, five indicators are analyzed: originality, fluency, flexibility, elaboration, as well as additional category, creative strengths. Specifically, originality indicates the deviation of the answer from the norm (how special the answer is); fluency reflects the number of relevant answers (how many products are created); flexibility reflects the ability of the informant to switch between different categories (how different the products are); and elaboration reflects the number of details included in an answer (how elaborated the product is). Additionally, creative strengths represent specific category reflecting 13 aspects, such as emotional expressiveness, humour, or the richness of imagery.

II) In the second phase, the respondents were tested for the formation of potential complex words based on a word-formation experiment consisting of three sets of tasks aiming at the formalization of the category Agents:
a) multiple choice from among applicable word-formation rules
b) naming based on a description of a person's unusual activity; and
c) naming based on drawings of people in unusual situations.

The above ultimately targets the resolution of competition in natural languages (e.g. Aronoff 2013, Lindsay & Aronoff 2013, Mac Whinney et al. 2014, Štekauer 2017). In word-formation, it manifests itself as, inter alia, competition between various naming strategies reflecting the conflict between economy of expression and semantic transparency. Unlike Dressler (2005) or Libben (2014: 19) who lay emphasis on the semantic correspondence of the meaning of constituents of a complex word to the original free word counterparts, we view semantic transparency as a degree to which the constituents of a new complex word enable a language user to predict its meaning, as in the following example:

i) novel writer is semantically transparent because it predicts the meaning as 'an Agent who performs an Action the Result of which is a novel', but its form is not economical;

ii) a different potential word denoting the same object of extra-linguistic reality, novelist, is more economical but, at the same time, less transparent, because it may
mean, for example, 'an Agent who performs an Action of writing/reading/stealing/selling/buying/collecting/…. books', where book may be either the Result of Action or an Object of Action.

iii) another option of naming the same object is writer, once again more economical than the first option but les transparent. Here the Result of Action may be not only a book but also a letter, a poem, a newspaper article, an essay, and many other things;

iv) Finally, the most economical and the least transparent option is the noun write (converted from to write). This option offers dozens of possible interpretations. The preference for any of these options reflects a language user's naming strategy, i.e. the preference for a more transparent or a more economical naming strategy.

While this preference can be affected by various factors, both linguistic and extra-linguistic (e.g. the productivity of word-formation rules, fashionable trends in word-formation; age, education, language background, etc.), main aim here is to examine the relationship between economy, transparency and creativity, as psychological factor captured by various indicators of Torrance Test of Creative Thinking mentioned above.

Regarding word formation tasks, each novel complex word was evaluated by means of an onomasiological type (OT). An OT results from the interaction between the onomasiological structure and its morphological representation. As such, it reflects the competition between the tendency towards semantic transparency and the tendency towards the economy of expression. Next, based on OT, ordinal level variable capturing the general level of transparency and economy across all three tasks was calculated. For this, median as a specific measure of central tendency in the sample data was used. For instance, across all three tasks (and/or in task one, two, or three separately), the participant could prefer a higher or lower degree of economy of expressions.

Regarding creativity, scores for originality, fluency, flexibility, elaboration, and creative strengths were computed. Scores were computed as a sum of the points that each informant gained in tasks (for instance, each additional detail in the picture that participant draw is evaluated by 1 point, thus, the picture with more details has a higher score in elaboration). Based on this scoring system, in TTCT, the participant could produce more or less numerous (fluency), unique (creativity), diverse (flexibility), and elaborated (elaboration) creative products.

The preliminary exploratory results indicated that creativity was related to word-formation. As indicated by Spearman's Rank-Order Correlation coefficient, it was shown, for instance, that economy, but not transparency was related to originality, elaboration and creative strengths. Specifically, the more original the informant was and the more the product was elaborated and more creative strengths were used (e.g. humour, emotional expressiveness or colourfulness of imagery and fantasy), the less economical a new complex word. Despite some inherent limitations, this pattern of results indicates that psychological factors, such as creativity, could play some role in word formation and encourages for future investigation.

References

Aronoff, M. 2013. Competition and the Lexicon. To appear in Proceedings of the Annual Meeting of La Società di Linguistica Italiana. Prepublication paper.

Dressler, W.U. 2005. Word-formation in natural morphology. In Pavol Štekauer and Rochelle Lieber (eds.), Handbook of word-formation. Dordrecht: Springer.; 267–284.

Kim, K. H. (2006). Can We Trust Creativity Tests? A Review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, 18(1), 3-14.

Kim, K. (2017). The Torrance Tests of Creative Thinking - Figural or Verbal: Which One Should We Use?, *Creativity. Theories – Research - Applications*, 4(2), 302-321.

Libben, G. 2014. The nature of compounds: a psychocentric perspective. Manuscript.

Lindsay, M. and M. Aronoff. 2013. Natural Selection in Self-Organizing Morphological Systems. In F. Montermini, G. Boyé and J. Tseng (eds.), Morphology in Toulouse: Selected Proceedings of Décembrettes 7, Munich: Lincom Europa.

MacWhinney, B., Malchukov, A. and Moravcsik, E. (eds.), 2014. Competing Motivations in Grammar and Usage. Oxford: Oxford University Press.

Štekauer, P. 2017. Competition in natural languages. In J. Santana-Lario and S. Valera (eds.), Competing Patterns in English Affixation. Bern: Peter Lang; 15-32.

Event/entity polysemy and head identification in deverbal compounds

Yoko Sugioka

Keio University (JAPAN)

1 Introduction: polysemy in deverbal nominals and compounds

Deverbal nominalization exhibits polysemy, in particular of the event/entity ('result') interpretation (Chomsky 1970, Grimshaw 1990, Lieber 2016 inter alia): e.g. *construction, painting* (event/product). This paper takes up hitherto little-discussed instances of polysemy in Japanese deverbal compounds: one type in N-V compounds with exocentric vs. endocentric structures (Section 2), and another type of polysemy in V-N compounds with different heads in the semantic structure and the morphological structure (Section 3).

2 Two competing structures for deverbal N-V compounds

Two different internal structures have been proposed for deverbal (synthetic) compounds in English, one with a nominalized N-V complex, the other as a N-N compound with the deverbal noun head (Ackema & Neelman 2004, Lieber 1983, inter alia).

(1) a. $[[\text{truck drive}]_V \text{ -er}]_N$
    b. $[\text{ truck }]_N [\text{ drive}_V \text{ -er}]_N$

Although (1b) is a more canonical morphological structure (N-N) for a compound, (1a) can attribute the semantic compositionality and productivity of synthetic compounds to the argument-verb relation of the embedded N-V complex. In fact, Wiese (2008) postulates (1a) as morphosyntactic and (1b) as morphophonological structure of German synthetic compound (e.g. *Appetit.hemmer* 'appetite blocker') in the parallel representation model of Jackendoff (2002). Booij (2009), on the other hand, proposes a construction-based analysis for Dutch deverbal compounds, where the two templates $[N\ V]_V$ and $[V\ er]_N$ are conflated.

The importance of the dichotomy in (1a/1b) becomes even clearer when we look at crosslinguistic data. The two structures (1a/1b) have been argued to manifest in two different types of deverbal compounds (N+V-infinitive form) in Japanese, namely, argument and adjunct compounds (Sugioka 1996, 2002). Argument compounds as in (2) are N with the (1a)-type structure ($[N\text{-}V]_N$), an exocentric structure due to the lack of nominalizing affix on the V.

(2) a. event /act   atena-kaki   'address writing';   ame-huri   'rain fall'
    b. agent / instr.   hana-uri   'flower vendor';   tume-kiri   'nail cutter'
    c. property   uso-tuki   'lie teller, liar'   kane-moti   'money-having, rich'

In contrast, adjunct compounds are used as nominal predicates (3), and unlike their English counterparts, can violate First-Sister Principle (i.e. adjunct N-Vt as in 3a).

(3) a. event/act   pen-gaki (-suru) ' (do) pen-writing (writing with pen) '
                         haya-gui   'fast-devouring'
    b. result state   usu-giri (-da)   ' (be) thin-slicing (thin-sliced) '
                         isi-zukuri   'stone-making (stone-made)'

These have (1b)-type endocentric structure with a deverbal N head: $[\ N\ [V]\ _N\ ]_N$, as evidenced by Rendaku (voicing of the initial consonant of the second element, cf. (3a) pen-gaki < kaki), which generally applies to the head of N-N compounds in Japanese (note there is no Rendaku for (2)).

To these two types we can add another type of N-V compounds, product compounds, an 'argument' compounds with (1b)-type structure. This addition can yield the following minimal pair of compounds:

(4) a. [[atena$_N$ kaki$_V$]$_V$]$_N$ 'address writing' (event = 2a)
     b. [atena]$_N$ [gaki$_V$]$_N$]$_N$ 'address writing (entity 'written address')

The first element *atena*$_N$ 'address' in (4a) is the internal argument of the verb *kaki* 'write', whereas it is in (4b) a modifier to the deverbal nominal *kaki* 'writing' and specifies the type of writing by its content, hence the difference in structure and Rendaku. (4a) denotes an event, while (4b) denotes a product. As expected from the difference in productivity between argument compounds (2) and adjunct compounds (3), the first element in (4a) can be freely replaced with nouns selected by the verb *kak* 'write' (i.e. *genkoo* 'manuscript', *tegami* 'letter', *repooto* 'report', *syoosetsu* 'novel', *namae* 'name' etc.), but it is restricted to only a handful of fixed expressions in (4b) (i.e. *memo* 'memo', *tyuui* 'caution', *ninsoo* 'profile'). The product compounds are unproductive and have lexicalized meanings: e.g. *tamago-yaki* '(lit.) egg-fry, specific egg dish', *ume-bosi* '(lit.)plum-dry, pickled plum', *isi-gumi* 'stone-arrangement'.

Exocentric structure similar to (4a) with different word order has been proposed for Romance V-N compounds (Di Sciullo & Williams 1987), and now-obsolete English ones.

(5) French: essui-glace 'window-wiper' ; Spanish: anza-cohetes 'rocket launcher'
     Italian: apri-porta 'door opener' ; Portuguese: afia-lápis 'pencil sharpener'
     cf. English: pick-pocket, scarecrow, killjoy, coverall

These are all argument compounds, and the exocentric structure [[essui$_V$-glace$_N$]$_V$]$_N$ reflects naming the act of 'V-ing N', on which instrument and agent interpretations presumably are based (cf. (2b) in Japanese) . Hence, postulating a V + N complex in exocentric structure (1a) can capture the commonalities found in the productive deverbal compounds denoting event (and metonymic extensions) across typologically different languages, e.g. Germanic and Romance languages, and Japanese. In contrast, compounds of (3/4b) are more idiosyncratic and lexicalized in Japanese, and not productive in Romance languages and English.

In sum, the dichotomy in structures (1a/1b) for deverbal compounds can account for polysemy as seen in (4) as well as those in typologically different languages.

3 Polysemy and head identification in Japanese V-N compounds

Compounds in Japanese with V-N word order in many cases denote an entity (6a), but there are some instances that exhibit entity/event polysemy (6b).

(6) a. tabe-mono 'eat-thing, food'; nomi-mizu 'drinking water'; nagare-bosi 'shooting star'
     b. uti-mizu 'spray-water'; yomi-mono 'read-material'; hari-gami 'put.up-paper';
        taki-bi 'burn-fire (bonfire)'; kakusi-goto 'hide-thing, secret'
     Thus, the following V-N compounds clearly denote an entity.

(7) a. itadaki-mono o   tabe-ru   'eat a gift'
       receive-thing ACC eat-PRES
    b. oki-gasa   o   kari-ru   'borrow (someone's) spare umbrella'
       leave-umbrella ACC borrow-PRES
    c. negai-goto   o   kak-u   'write (one's) wish'
       wish-matter ACC write-PRES

They can also denote action as an argument of the verbs *suru* 'do', hazime-ru 'begin, etc.

(8) a. kyaku kara   itadaki-mono o   su-ru   'receive a gift from a guest'
       guest from   receive-thing ACC do-PRES
    b. kaisya ni   oki-gasa   o   su-ru   'leave a spare umbrella at the office'
       office LOC leave-umbrella ACC do-PRES

    c. nagare-bosi ni   negai-goto o   su-ru   'make a wish to a shooting star'
       shoot-star   DAT wish-matter ACC do-PRES

It should be noted that the compounds in (8) are event nouns (N), rather than verbal nouns (VN) (i.e. Sino-Japanese verbs such as *hookoku* 'report', san-sui 'spray water') that take light verb *-suru* without Accusative case (ACC): *san-sui-suru* vs.*uti-mizu-suru 'spray water'. Hence they are endocentric compounds with a N head: [uti$_V$-mizu$_N$]$_N$.

    Nevertheless, there is evidence showing that the left-hand V is indeed responsible for the event interpretation of the compounds in (8). First, we can argue that the Source (8a), Locative (8b), and Goal (8c) arguments are selected by the leht-hand V in the compounds. This is because, if they were to modify the compound as a whole, they would have to take an adnominal Genitive marker *no* 'of', but that would be unacceptable.

    (9) *[Kyaku kara-no   itadaki-mono] o   suru. 'receive a [gift from a guest]' (cf. 8a)
         guest   from GEN   receive-thing ACC do-PRES

    Second, the aspectual feature of V determines that of the compound. While *taki* 'burn' (10a) denotes activity, *otosi* 'lose' (10b) denotes punctual event, yielding the contrasts below:

    (10) a. 3 zikan taki-bi   o   si-ta.   'burned a bonfire for 3 hours'
                 hour burn-fire ACC do-PAST

         b. *3 zikan otosi-mono o   si-ta. 'lost (something) for 3 hours'
                 hour lose-thing   ACC do-PAST

    (11)a. taki-bi-tyuu   ni   'while burning a bonfire'
            burn-fire-during at

         b. *otosi-mono-tyuu   ni   'while losing (something)'
            lose-thing-during   at

    (12)a.*taki-bi   ga   takusan at-ta   'there was much burning fire'
            burn-fire NOM much   be-PAST

         b. otosi-mono ga   takusan at-ta   'there were many instances of losing items'
            lose-thing NOM much   be-PAST

Time adverbial *3 zikan* 'for 3 hours' and aspectual suffix *-tyuu* 'during' can modify durative activity (10a, 11a) but not punctual event (10b, 111b), while the verb *aru* 'be' can be used with a punctual event nominal to denote its happening (12b), but not a durative event (12a).

    Consequently, we can say that while the head in the morphological structure of these V-N compounds is always the right-hand N so that the whole compound is N rather than VN, the head in the semantic structure can vary depending on the context: it is N when an entity reading suits the context as in (7), but the left-hand V is identified as the head when an event interpretation is called for as in (8). This is a type of structural polysemy in the sense of Pustejovsky (1995), where contextual coercion can force one of the multiple interpretations afforded by the qualia structure of the word (cf. Ono 2013). The following are (due to space limit) partial qualia specifications for these V-N compounds.

    (13)a. otosi-mono 'lose-thing'   Formal: entity (y).event (e)   Agentive: lose (e, x, y)
         b. taki-bi 'burn-fire'   Formal: entity (y).event (e)   Agentive: burn (e, x, y)
         c. yomi-mono 'read-thing'   Formal: entity (y).event (e)   Telic: read (e, x, y)

    When a word has dual Formal qualia as in (13), the multiple senses can sometimes coexist in a sentence (e.g. *She came in through the broken **window*** (physical object/aperture) cf. Kageyama 1999:43). In fact, V-N compounds can form this type of zeugma as well.

    (14)a. Tan-zikan de nagai kaki-mono o   si-ta   'did a long writing in a short time'
            short-time in long   write-thing ACC do-PAST

         b. Kooka-na itadaki-mono o   si-ta   'received an expensive gift'
            expensive receive-thing ACC do-PAST

    c. Tiisa-na taki-bi   o   hazime-ta.   'begin a small bonfire'
        small   burn-fire ACC begin-PAST

Crucially, *suru* 'do' (14a,b), *hazime* 'begin'(14c) and the time adverbial 'in a short time' (14a) call for event meaning, while adjectives *nagai* 'long' (14a), *kooka-na* 'expensive' (14b), and *tiisa-na* 'small' (14c) select entity reading of the V-N compound in the same sentence.

4 Summary and implications

Two types of event/entity polysemy in Japanese deverbal compounds stem from 1) exocentric and endocentric internal structures for N-V compounds, and 2) semantic coercion by the context for V-N compounds. First, nominalization of N-V complex as opposed to V to N conversion yield the contrast in meaning, as well as semantic compositionality and productivity. This contrast is revealing in face of the common assumption that exocentric structure (1a) is non-canonical in morphological structure. On the other hand, event/entity polysemy in V-N compounds is activated by semantic coercion, and can be analyzed by splitting the head in the semantic structure from that in the morphological structure. Ramifications of postulating different heads in different components must be further developed, e.g. in a modular approach to morphology (e.g. Jackendoff 2002, Sadock 2012).

References

Ackema, Peter & Ad Neeleman. 2004. *Beyond morphology*. Oxford: Oxford University Press.

Booij, Geert E. 2009. Compounding and construction morphology. In Lieber, R. and P. Stekauer (eds.) *Oxford handbook of compounding*, 201-216. Oxford: Oxford University Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Jacobs R.A. and P.S. Rosenbaum (eds.) *Readings in English transformational grammar*. Waltham, MA; Ginn, 184-221.

Di Sciullo, Anna-Maria & Edwin Williams. 1987. *On the definition of word*. Cambridge, MA: MIT Press.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. *Foundations of language*. Oxford: Oxford University Press.

Kageyama, Taro. 1993. *Bunpoo to gokeisei* [Syntax and word formation].Tokyo: Hituzi Syobo.

Kageyama, Taro. 1999. *Keitairon to imi* [Morphology and semantics]. Tokyo: Kurosio.

Lieber, Rochelle. 1983. Argument linking and compounds in English. *Linguistic Inquiry* 14: 251-285.

Lieber, Rochelle. 2016. *English nouns*. Cambridge: Cambridge University Press.

Ono, Naoyuki. 2013. *N-o suru* koobun ni okeru koo-sentaku to kyoosei [Argument selection and coercion in *N-o suru* construction]. In Kishimoto, H. & Y. Yumoto (eds.) *Hukuzatu-zyutugo kenkyuu no genzai* [Current studies on complex verbs], Tokyo: Hituzi-shobo. 27-40.

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, MA: MIT Press.

Sadock, Jerrold. 2012. *The modular architecture of grammar*. Cambridge: Cambridge University Press.

Sugioka, Yoko. 1986. *Interaction of derivational morphology and syntax in English and Japanese*. New York: Garland. (Reprinted 2019, London: Routledge)

Sugioka, Yoko. 1996. Regularity in inflection and derivation: Rule and analogy in Japanese deverbal compound formation. *Acta Linguistica* 45: 231-253.

Sugioka, Yoko. 2002. Incorporation vs. modification in Japanese deverbal compounds. *Japanese/Korean Linguistics 10*, CSLI Publications. 496-509.

Wiese, Richard. 2008. Two levels of morphological structure. *Journal of Germanic Linguistics*. 20-3. 243-274.

Poster session B:
Thu 25 September 15:00–16:00

Cross-linguistic research into derivational networks

Alexandra Bagasheva
Sofia University "St. Kliment Ohridski"

Lívia Körtvélyessy
P.J. Šafárik University, Košice

Pavol Štekauer
P.J. Šafárik University, Košice

Salvador Valera
University of Granada

Jan Genci
TUKE, Technical University Kosice

Until relatively recently, derivational networks[1] have received little theoretical attention (exceptions of various theoretical frameworks include, e.g. Dokulil 1962, Pounder 2000, Beecher 2004, Ševčíková & Žabokrtský 2014, Bonami & Strnadová 2016). Largely as a result of this, no major empirical, let alone cross-linguistic, research has been implemented yet. The derivational network is defined as the intersection of the *paradigmatic capacity*, i.e. the system of all direct derivatives from a simple underived word-formation base, and the *order of derivation*, i.e. all the linear derivatives from the word-formation base. This does not necessarily assume linear derivation.

The poster presents the fundamental theoretical and methodological principles as well as selected results of a research project covering 40 European languages.[2] The point of departure is the choice of 10 basic words per three major word-classes: nouns, adjectives and verbs. Starting from a larger list, the informants compiled a list of 30 words identical for the sample languages, excluding any derived basic words. The initial longer lists of words were selected from Swadesh's core vocabulary. Since these words serve as a starting point for the development of cross-linguistically comparable derivational networks, all of them must be simple underived words in all sample languages.

The research introduces a semantic classification of all the possible derivatives (as attested in corpora and/or dictionaries, in this order as available for each language) in individual orders of derivation by means of a list of semantic categories that was compiled for this purpose. This means that each derivational step (order of derivation) away from the base is defined by means of a specific semantic category. This approach allows to identify:

    i)   correlations between the occurrence of individual semantic categories and the order of derivation,

    ii)   semantic categories that systematically block any further derivation, and

    iii)   the maximum number of derivatives per each semantic category within a given order of derivation.

We speak of the paradigmatic capacity of the word-formation base represented by the number of derivatives from the word-formation base. The paradigmatic capacity and the orders of derivation establish the *derivational network*, that is a network of derivatives derived from the same word-formation base (simple underived word) with the aim of

---

[1] Derivational is used here in the narrow sense of affixation. No other processes are taken into account. Conversion being excluded for methodological reasons, no pure transposition was recorded, so word-class change was always found to be associated with an intervening, additional semantic category.

[2] Basque, Bulgarian, Catalan, Chechen, Croatian, Czech, Danish, Dargwa, Dutch, English, Estonian, Finnish, French, Frisian, Galician, Georgian, German, Greek (Modern), Hungarian, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Russian, Saami (North), Serbian, Slovak, Slovene, Spanish, Swedish, Tatar, Turkish, Ukrainian, Welsh.

formally representing specific semantic categories. Thus, the sum total of maximum derivations for all ten words of a given word-class identifies the *maximum derivation network*. This variable serves as a basis for the calculation of the *saturation value* for individual words of a given word-class as well as for the word-class as a whole. The saturation value is calculated as a proportion between the number of actual derivatives in a particular derivational network and the maximum derivational network. The saturation value identifies the degree to which a particular word makes use of the derivational capacity of the derivational network. If the saturation values for all ten words are consistent in a given order of derivation, i.e. if they do not deviate statistically significantly, the predictability of derivation within a given word-class is high. This has been evaluated by means of the parameter of *standard deviation*. In practice, the above allows to evaluate language genera by orders of derivation and word-classes.

Additional findings concern the identification of:

i)   the average number and the maximum number of orders of derivation within the examined sample of languages,

ii)   the maximum and the average paradigmatic capacity for three groups of sample words (nouns, verbs and adjectives),

iii) the total number of derivatives per order of derivation for each basic word and the average number of derivatives per order of derivation per word-class, and

iv)   systematic combinations of semantic categories characteristic of the specific sample/language.

All in all, the poster presents a new perspective for studying word-formation combinability which has been restricted to the study of affix combinations so far. This is particularly relevant, considering semantic combinability in its diverse manifestations in word-formation has not been explored yet. The poster presents the following main conclusions:

i)   Languages differ considerably in their derivational capacity, i.e. in the number of derivatives in derivational networks.

ii)   The derivational potential of simple underived nouns and adjectives is very similar. Verbs have the highest maximum derivational network value in every order of derivation.

iii) The richness of derivational networks is sensitive to the word-class of the basic word and to the order of derivation.

iv)   There is a tendency for languages to actualize 20%-30% of the derivational potential of a word-class.

v)   Some languages keep high saturation values across all three word-classes.

vi)   There is a tendency for saturation values to fall gradually with the rising order of derivation in all three word-classes.

vii) There is no geographically homogeneous territory on which the languages of topmost saturation values are spoken.

viii) Derivational networks are most predictable in the 1st order.

ix)   The maximum number of orders of derivation is five for all three word-classes.

x)   Semantic categories with blocking effects are reported in all the languages of the sample except in Welsh.

xi)   Typical combinations of semantic categories are reported for most languages, also regularly for each word-class.

xii) Recursiveness is relatively frequent in the sample, but it is comparatively higher in Slavic languages.

xiii)   Differences can be found between language types as regards multiple occurrence of semantic categories.

Bibliography

Beecher, Henry. 2004. Derivational paradigm in word formation. Available at: http://pdfcast.org/pdf/research-paperi-derivational-paradigm-in-word-formation.

Bonami, Olivier & Jana Strnadová. 2016. *Derivational paradigms: pushing the analogy*. Paper presented at the *49th Annual Meeting of the Societas Linguistica Europaea (SLE)*, Naples, Italy, 31 August–3 September 2016.

Dokulil, Miloš. 1962. *Tvoření slov v češtině I. Teorie odvozování slov.* [Word-formation in Czech, I. Theory of derivation.] Prague: Nakladatelství Československé akademie věd.

Pounder, Amanda. 2000. *Processes and Paradigms in Word-Formation Morphology*. Berlin/New York: Mouton de Gruyter.

Ševčíková, Magda & Zdeněk Žabokrtský. 2014. Word-formation network for Czech. In Nicoletta Calzolari et al. (eds.), *Proceedings of the 9th International Language Resources and Evaluation Conference (LREC 2014)*, 1087–1093. Paris: ELRA.

Stem spaces in abstractive morphology:
A look at defectiveness in French conjugation

Gilles Boyé

Université Bordeaux-Montaigne & UMR5263

1 Introduction

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 & Fraser 1993, Brown & Hippisley 2012 or Natural Morphology: Kilani-Schoch & 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 & Finkel (2013), or the *joint predictiveness* of Bonami & Beniamine (2016)).

In this paper, we revisit French conjugation and the analysis of defectiveness as suppletion of a null stem proposed by Boyé (2000) and Boyé & Cabredo Hoffherr (2010) in the light of implicative morphology.

2 Defectiveness as suppletion of a null stem

The analysis proposed by Boyé & Cabredo Hoffherr relied on a partition of the conjugation paradigm into morphomic zones systematically sharing the same stem. Figure 1 illustrates the French conjugation stem space of Bonami & Boyé (2002). The colors indicate the cells sharing the same stem, the Grace convention (Rajman et al., 1997) is used for the labels.[1]
In this case as in the case of Spanish conjugation, the shape of the stem space corresponded to the frontiers of defectiveness. French verb CLORE, for example, lacks indicative imperfective, simple past and subjunctive imperfective entirely but it also has gaps for indicative present 1PL and 2PL following the outline of the gray and the mid green stem space.

The proposal was straightforward: a null stem blocked the derivation of inflectional forms based on it. For a syntagmatic account of inflection cast, for example, in Paradigm Function Morphology (Stump, 2001), the selection of a null stem would prevent the production of an inflected form.

However, in an abstractive word-based approach to inflectional morphology, this type of analysis is not possible anymore because stems do not have a primary place in the inflectional system.

---

[1] For finite forms, the label is composed of three parts: one character for the mode, on character for the tense, and two the person. For example, pi1S stands for present indicative 1st person singular.

pi1S	pi2S	pi3S	pi1P	pi2P	pi3P
ii1S	ii2S	ii3S	ii1P	ii2P	ii3P
fi1S	fi2S	fi3S	fi1P	fi2P	fi3P
pc1S	pc2S	pc3S	pc1P	pc2P	pc3P
ps1S	ps2S	ps3S	ps1P	ps2P	ps3P
ai1S	ai2S	ai3S	ai1P	ai2P	ai3P
is1S	is2S	is3S	is1P	is2P	is3P
—	pI2S	—	pI1P	pI2P	—
inf	pP	ppMS	ppMP	ppFS	ppFP

Figure 1: The stem space of French conjugation according to Bonami & Boyé (2002)

3 Defectiveness as the remainder of predictability

In this new context, predictiveness cannot be used to propagate defectiveness because most lexemes appear with many gaps in any given corpus. As noted by Bonami & Beniamine (2016) and Boyé & Schalchli (to appear), the forms found even in large samples cover only a part of the grammatically defined paradigms, and the problem of defectiveness seems not to be how to predict it but rather how it is avoided by speakers even in the face of very sparsely populated paradigms with many missing forms: the PCFP for all verbs.

CLORE:

klo	klo				kloz
klora					
kloz					
—	klo	—			—
klor		klo	klo	kloz	kloz

CLÔTURER:

klotyr					klotyr
klotyrɛ					klotyrɛ
klotyrəra					
klotyre		klotyra			
—		—			—
klotyre		klotyre		kloture	

Figure 2: The forms of CLORE ('to close') and CLÔTURER ('to close') found in Lexique3

To capture both the filling strategy and the defectiveness phenomena we propose an analysis based on two steps.

First, we use SWIM (Boyé, 2017) to fill gaps by evaluating the converging predictions made by existing co-forms for every pair of cells, searching for the largest cliques of concurring predictions to fill the paradigm of each and every lexeme. This allows to generalize inflection classes without exemplary paradigms but with sufficient partial paradigms to cover the whole system. Because of its cliquing mechanism, SWIM does not allow to generalise inflectional classes that do not possess a complete cover.

For CLÔTURER, the sample does not contain an exemplary paradigm but every pair of cells in its inflection class is documented and SWIM can infer content of all the missing cells and find a clique of related forms that almost fills the whole paradigm (Fig. 3). But for CLORE, there is no support for an inflectional class that would be congruent with the known forms and the generalisations possible will be limited to partial generalisations over a very small subset of cells (Fig. 4). This leads to a low recall for this type of verbs.

To fix this caveat, we propose, in a second step, to use zero-entropy rules to predict the remaining missing forms. To avoid marginal rules, we filter zero-entropy rules based on their scopes, i.e. the number of co-pairs involved in establishing the rules in question (100 co-forms

klotyr	klotyr	klotyr	klotyrɔ̃	klotyre	klotyr
klotyrɛ	klotyrɛ	klotyrɛ	klotyrjɔ̃	klotyrje	klotyrɛ
klotyrərɛ	klotyrəra	klotyrəra	klotyrərɔ̃	klotyrəre	klotyrərɔ̃
klotyrərɛ	klotyrərɛ	klotyrərɛ	klotyrərjɔ̃	klotyrərje	klotyrərɛ
klotyr	klotyr	klotyr	klotyrjɔ̃	klotyrje	klotyr
klotyrɛ	klotyra	klotyra	klotyram	klotyrat	klotyrɛr
klotyras	klotyras	klotyra	klotyrasjɔ̃		klotyras
—	klotyr	—	klotyrɔ̃	klotyre	—
klotyre	klotyrâ	klotyre	klotyre	klotyre	klotyre

Figure 3: The forms of CLÔTURER ('to close') generated by SWIM

klo		klo			kloz
		klora	klora		
kloz	kloz	kloz			kloz
—	klo	—			—
klor			klo	klo	kloz kloz

Figure 4: The forms of CLORE ('to close') generated by SWIM

at least). At this stage, we use the zero-entropy rules on all available co-forms to get predictions for missing forms. If all zero-entropy predictions for a cell concur, the gap is filled with the common prediction, if the zero-entropy rules diverge, the gap remains.

With this method, gaps are filled if they belong to a zero-entropy zone where at least one form is known. The emerging zones are similar as those calculated by Bonami & Boyé (2014). While these generalisations fill the gaps for "abstractible" stems, they will leave blank the gaps for "unknown" stems leading to defectiveness that will outline a space equivalent to a stem space.

4 Conclusion

The emergence of stem space like zones in abstractive morphology is natural. As Bonami & Boyé (2014) observed, zero-entropy zones are the image of what the authors used to consider stem spaces and what Stump & Finkel (2013) call distillations. In the case of defectiveness however, what we observe is the negative image of stem spaces of distillation: the lack of a base for predictions creates a dark hole in the paradigm that conforms to the shapes of neighbouring distillations leaving an empty space corresponding to a set of distillations.

Working with realistic data, a lot remains to be desired in the processing, on one hand some zero-entropy rules emerge in random places, even after filtering, and create alternate propositions preventing some correct predictions to emerge, on the other hand, some zones that should be identified as zero-entropy zones fail to appear because of a lack of information in the initial sample, but a new analysis of defectiveness is definitely possible along these lines.

In this perspective, it is natural for defectiveness to have gaps that have the same shape as stem spaces not because they are the suppletion of a null stem but rather because they are what is left after all other generalisations have been made. Defectiveness corresponds lexemes

belonging to inflectional patterns that can't be generalised because they possess some forms that block the general cliquing and gaps that correspond incidentally to a stem space.

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar*, 54–82. Oxford: Oxford University Press.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42. 531–573.

Bonami, Olivier & S. Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.

Bonami, Olivier & Gilles Boyé. 2002. Suppletion and dependency in inflectional morphology. In Frank van Eynde, Lars Hellan & Dorothee Beermann (eds.), *Proceedings of the hpsg-2001 conference*, Stanford: CSLI Publications.

Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux* 17–45.

Boyé, Gilles. 2000. *Problèmes de morpho-phonologie verbale en français, espagnol et italien*: Université Paris 7 dissertation.

Boyé, Gilles. 2017. Inflectional morphology without inflectional classes: a french conjugation reflection. Paper presented at ISMo17.

Boyé, Gilles & Patricia Cabredo Hoffherr. 2010. Defectiveness as stem suppletion in French and Spanish verbs. In Matthew Baerman, Greville G. Corbett & Dunstan Brown (eds.), *Defective paradigms: Missing forms and what they tell us* (Proceedings of the British Academy 163), 35–52. The British Academy Oxford: Oxford University Press.

Boyé, Gilles & Gauvain Schalchli. to appear. Paradigms in realistic data: The paradigm cell finding problem. *Morphology* .

Brown, Dunstan & Andrew Hippisley. 2012. *Network Morphology: a defaults based theory of word structure*. Cambridge: Cambridge University Press.

Corbett, Greville G. & Norman M. Fraser. 1993. Network morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics* 29. 113–142.

Kilani-Schoch, Marianne & Wolfgang Dressler. 2005. *Morphologie naturelle et flexion du verbe français*. Tübingen: Gunter Narr Verlag.

Rajman, Martin, Josette Lecomte & Patrick Paroubek. 1997. *Format de description lexicale pour le français. Partie 2: Description morpho-syntaxique*. Tech. Rep. GRACE, GTR 3-2.1 LIMSI.

Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27. 379–423 & 623–656.

Stump, Gregory T. 2001. *Inflectional morphology. a theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. & Raphael Finkel. 2013. *Morphological typology: From word to paradigm*. Cambridge: Cambridge University Press.

Derivational morphemes in Xochistlahuaca Amuzgo

Bien DoBui

1 Introduction

Observations from Xochistlahuaca Amuzgo (XA) spoken in Guerrero state by about 4,000 people will be shown to challenge Smith Stark and Tapia Garcia who describe Oaxacan Amuzgo derivation as nearly "non-existent" (1986:5-6). In this presentation, we will look at i) the derivational operations in which these morphemes are found using a lexicalist X-bar model (inspired by Herrera's 1995 work on Zoque) and ii) how pretonic reduction while resulting in strict CCV monosyllabic roots also helps to make derivation productive, creating proclitic/prefix material from former lexical roots that grammaticalize over time.

2 Derivational operations

Different operations are attested. Compounds may involve change in lexical class (as in below 'money') or not (as 'helicopter'). Class markers may appear obligatorily (as in 'worm') or derive a new word from a lexical root appearing independently elsewhere (as in 'grilled object'). Prefixation is particularly active in verbal derivation (as with the different verbal derivations from the adjective 'grilled'). The enclitic 'TE' or extended theme (or *tema extendido* in following with Amuzgo linguist Jair Apostol), serves to derive verbs from nouns like 'flesh' or adjectives like 'clumsy'.

Composition[1]	$[\text{w}\text{ʔa}^1 + \text{hɲdje}^{3\cdot} + \text{tse}^{1\cdot}]$ house + wind + shell 'helicopter'		$[\text{tshɔʔ}^{2\cdot} + \text{ɲhẽ}^1]$ stone + thin 'money'	
Class marker[2]	$[\text{ka}^{2\cdot} = \text{tsõ}^3]$ ANIM = worm 'worm'		$[\text{ka}^{2\cdot} = \text{tʃẽ}^1]$ ANIM = grilled 'grilled object'	
Prefixation*	$[\text{ma}^2\text{-tsei}^{1\cdot} + \text{ka}^1 = \text{tʃẽ}^1]$ PROG.SG-do + ANIM = grilled		'to grill (s.th.)'	transitive
	$[\text{ha}^2\text{-wi}^1 + \text{tʃẽ}^1]$ PROG-become + grilled		'to grill'	intransitive
	$[\text{ha}^2\text{-wi}^1 + \text{tʃẽ}^1 = \text{ɲẽ}^{1\cdot}]$ PROG-become + grilled = TE		'to become grilled'	inchoative
	$[\text{kwi}^2\text{-wi}^1 + \text{ka}^1 = \text{tʃẽ}^1]$ PROG-become + ANIM = grilled		'to get grilled'	antipassive
	$[\text{kwi}^2\text{-wi}^1 = \text{tʃẽ}^1]$ PROG-become + grilled		'to be grilled'	passive
Cliticisation	$[\text{se}\text{ʔ}^{2\cdot}]$ 'flesh'	$[\text{se}^2\text{ʔ} = \text{ɲe}^{1\cdot}]$ flesh = TE 'be muscular (animals)'	$[\text{seʔ} + \text{mpa}^3 = \text{ɲe}^{1\cdot}]$ flesh + lung = TE 'be clumsy'	

---

1   Data is taken from mid-century SIL archives (Bauernschmidt, 2014 manuscript) as well as original data elicited with the Endangered Language Alliance in New York City (2010-2013) and during fieldwork in 2016. *Verbal data come from Apostol (2014).

2   The ka = class marker appears before nouns and adjectives and is annotated as ANIM to suggest a generic animated meaning, though in fact, most nouns marked by the proclitic are animals and adjectives marked by ka = typically accord for animacy.

In Figure 1 below, the verbs 'be muscular' and 'be clumsy' are both derived from the noun 'flesh', though in the case of 'be clumsy,' the verb deadjectival from 'clumsy' (itself a denominal adjective). The use of X-bar as a model for derivation serves to clearly show multiple cycles below the word-level. It also assumes the headedness of the derivational morpheme TE.

Figure 1. Derivation of 'be muscular' and 'be clumsy'

Word V^0	Word V^0
N V	V^{-1}
[se^{2}ʔ· =ɲe^1·]	N N V
'be muscular (animals)'	[se^{2}ʔ· +mpa^3 =ɲe^1·]
	'be clumsy'
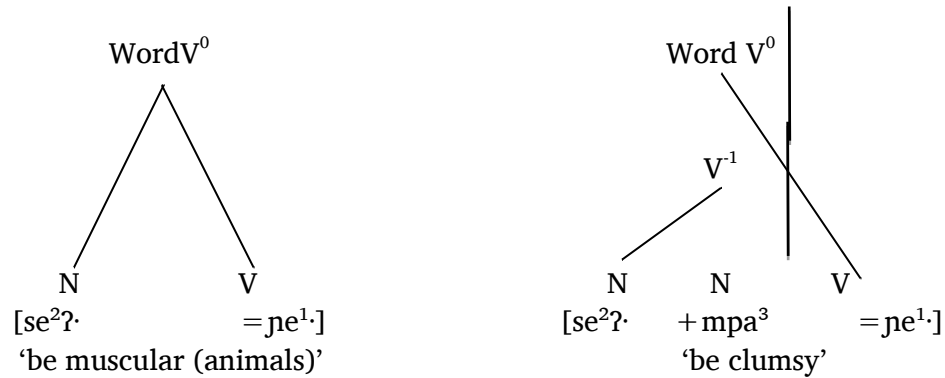
3   Productivity and syllable reduction

Distinguishing between certain compounds, 'class markers' and 'prefixes' depends on prosodic and morphophonological insight, as well as measures of productivity. Below, we see the words for 'house' and 'animal' in independent form. These are also found in composition with other lexical roots that appear independently, as in the compounds for 'car', or 'cow', and in phonologically reduced forms as in the complex words 'church' and 'dog'.

	Independent	**Composition**	**Class marker**		
sg	[wʔa^1] 'house'	[wʔa^1+ʧõ^3·] house+fire 'car'	[wa^1=ts?õ^2·] EDIFICE=heart 'church'		
pl	[l^{12}ʔa^1] PL.house 'houses'	[l^{12}ʔa^1+ɲʔõ^3·] PL.house+PL.fire 'cars'	[wa^1= nʔõ^2] EDIFICE=PL.heart 'churches'	[la^1+ts?õ^2·] PL.house+heart 'churches'	[la^1+nʔõ^2·] PL.house+PL.heart 'churches'
sg	[kioʔ^1] 'animal'	[kioʔ^1+ʃkɛ̃^{2}] animal+female 'cow'	[ka^2·= tsueʔ^{21}] ANIM=dog 'dog'		
pl	[ŋ̍^{2}-kioʔ^1] PL-animal 'animals'	[ŋ̍^{2}-kioʔ^1+ntkɛ̃^{2}] PL-animal+PL.female 'cows'	[ka^2·= lueʔ^{21}] ANIM=PL.dog 'dogs'		

The gradience of prosody in the above ranges from full independent roots (e.g. 'house', 'animal') with lexical tone, phonemic vowel, lexical laryngeal quality to reduced forms of these same roots, with neutralization of laryngeal information and loss of lexical vowel and tone. Different patterns of plural marking show different statuses between lexical roots and so-called class markers, which are not (or are irregularly) inflected. In terms of productivity, compounds with full lexical roots are many but rarely occur with any one root in productive numbers. These observations correspond to grammaticalization in the sense that these word-

initial morphemes are now motivated and can be analyzed as part of a complex whole, though phonological reduction is also often part of this process (Lehmann, 1999).

A set of class markers can be observed in accordance with the above considerations (prosodic and morphophonological insight, measures of productivity). These indicate a generic category of meaning, typologically typical in Mesoamerica (Haspelmath 2001, Palancar 2016). Non-lexical inventories are evidently reductions of historical compositions, an example of a notable phenomenon of pretonic reduction visible at the lexical root level. In Proto-Amuzgo-Mixtecan cognates proposed by Longacre and Millon (1961), the *CVCV couplet is reduced in Amuzgo to C(C)V, and synchronically, in its local group, Amuzgo stands apart typologically for its monosyllable lexical root, as shown below.

Mixtec	XA	
Chalcatongo: *kúkà*	ʃkaʔ.	'comb'
Yosondúa: *nducha*	ṇtá	'water'
Santa Maria: *nūnì*	na^n	'sweet corn'
Copala Trique *yu³we³⁵*	ʦue	'bedroll'

Though reduction is not always a necessary part of grammaticalization (and in fact is not in XA), pretonic reduction phenomena feed a gradient scale of morphemic material that become motivated for derivational purposes.

References

Apostol, Jair. 2014. *Clases flexivas verbales en el amuzgo de Xochistlahuaca, Guerrero*. Mexico: Centro de Investigaciones y Estudios Superiores en Antropología Social. Master thesis.

Bauernschmidt, Amy. manuscript, 2014 version. Diccionario del amuzgo de Xochistlahuaca. SIL.

Lehmann, Christian. 1999. New reflections on grammaticalization and lexicalization. Wischer, Ilse & Diewald, Gabriele (eds.), *New reflections on grammaticalization*. Amsterdam & Philadelphia: J. Benjamins (TSL, 49). Modified 28.04.2005.

Longacre, Robert E. et René Millon. 1961. Proto-Mixtecan and Proto-Amuzgo-Mixtecan Vocabularies: A Preliminary Cultural Analysis. *Anthropological Linguistics*. 3:4. 1-44.

Haspelmath, Martin. 2001. "Word classes/parts of speech." In: Baltes, Paul B. & Smelser, Neil J. (eds.) *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam: Pergamon.

Herrera Z. Esther. 1995. *Palabras, estratos y representaciones: Temas de fonología léxica zoque.* Mexico: El Colegio de México.

Palancar, Enrique. 2016. A typology of tone and inflection: A view from the Oto-Manguean languages of Mexico. In Enrique Palancar et Jean-Léonard Léonard (Eds.). *Tone and Inflection: New facts under new perspectives*. 199-224. Berlin: De Gruyter Mouton.

Smith Stark, Thomas C. and Fermin Tapia Garcia. 1986. The active-stative nature of Amuzgo (Otomanguean). Talk given at the LSA meeting in New York, New York.

French Denominal Verbs: from countability to aspect

Edwige Dugas[1], Pauline Haas[2],
Rafael Marín[1]

[1]UMR 8163 (CNRS) & Université de Lille

[2]UMR 8094 (CNRS, P3) & P13

1   Aim of the study

This paper focuses on French denominal verbs (henceforth dNVs) with the affixes *a-* (*aborder* 'approach, address' < *bord* 'edge'), *dé-* (*défricher* 'clear' < *friche* 'wasteland'), *é-* (*écrémer* 'skim' < *crème* 'cream'), *en/em-* (*embouteiller* 'bottle' < *bouteille* 'bottle'), *-ifier* (*glorifier* 'glorify' < *gloire* 'glory'), and *-iser* (*étatiser* 'nationalise' < *état* 'state'). Our aim is to empirically analyze the mass/count properties of the base nouns (bNs) and the aspectual properties of the derived verbs. To our knowledge, there has been no work on that topic in French. In English, most denominal verbs are converted verbs (*saddle, kennel*), and they have been studied in quite some depth (Clark and Clark 1979, Hale and Keyser 1993, Plag 1999, Harley 1999 and others). French has also quite a few converted denominal verbs (*tapisser* 'cover', *saler* 'salt'), which have been studied in detail (Hirschbühler & Labelle 2008, Tribout 2011 and references therein). As in English, most studies focus on the Locatum/Location distinction and on manner incorporation. Note that French *-ifier* and *-iser* denominal verbs (Willems 1979, Dal & Namer 2000, Sagot & Fort 2009, Namer 2013) are often studied together with deadjectival verbs.

In her analysis of English converted verbs, Harley (2005) shows that denominal verbs formed from count bNs are telic while those formed from mass bNs are compatible with an atelic interpretation, although they can also be given a telic interpretation (e.g. *butter the bread* can be atelic, because *butter* is mass, or telic if the bread is interpreted as a bounded/count incremental theme) (see also Rimell 2012). Our goal in this paper will be to examine whether this relationship can also be observed in the case of French affixed dNVs. Our hypothesis is that there is a relationship between the mass/count properties of the bNs and the aspectual properties of the dNVs: as Harley (1999: 4) puts it (after Talmy 1978, Bach 1986, among others): "the mass/count distinction in the spatial dimension, as exhibited by things, is analogous to the bounded/unbounded distinction in the temporal dimension, as exhibited by events".

2   Data collection and annotation

The corpus comprises of dNVs listed in the TLFi entries of the affixes *a-, dé-, é-, en/em-, -ifier* and *-iser;* this set was supplemented by verbs identified as denominal in TreeLex (Kupść 2009). A total of 313 verbs have been collected: 48 [*a*-N]$_V$, 40 [*dé*-N]$_V$, 41 [*é*-N]$_V$, 81 [*en/em*-N]$_V$, 26 [N-*ifier*]$_V$, 77 [N-*iser*]$_V$. The semi-automatic retrieval of dNVs was followed by a manual filtering. The lexemes which have been deleted fall in the following categories:

- bases corresponding to proper nouns (*enversailler* 'to put in Versailles' < *Versailles, américaniser*), including lexicalised dNVs (*pasteuriser* 'pasteurize' < *Pasteur*), as the N-*iser* construction can take almost any proper noun as a base;
- dNVs whose use is rare today or belongs to specialised vocabulary (*affruitier, déliter*);
- [N-*aliser*]$_V$ most probably built on an adjectival base (*familiariser* 'familiarize', *finaliser* 'complete', *libéraliser* 'liberalize'). See also Lignon (2010) and Namer (2013) who point out the high number of ambiguous *-iser* lexemes.

The derived verbs have been annotated according to lexical aspect and their corresponding base nouns according to countability.

2.1 Verbal aspect

From an aspectual point of view, verbs have been traditionally described as states (STATE), activities (ACT), accomplishments (ACC) or achievements (ACH) (Vendler 1967). States and activities are atelic, i.e. unbounded situations, while accomplishments and achievements are telic, i.e. bounded situations. For this study, the annotation consisted in a double manual annotation made by two experts followed by manual adjudication. The annotators, who had to choose between four tags corresponding to the four Vendlerian classes, agreed in 80.45% of cases. Tags were given after the usual tests presented in the literature on verb lexical aspect (see Garey 1957, Kenny 1963/[1994], Vendler 1967, Dowty 1979, Rothstein 2004, among others): progressive form; duration complements *in x time* and *for x time*, etc.

State	Activity	Accomplishment	Achievement
11 (3.5%)	38 (12.1%)	112 (35.8%)	152 (48.5%)
Atelic		Telic	
49 (15.6%)		264 (84.4%)	

Table 1. Aspectual properties of the derived verbs

As can be seen, telic verbs are much more frequent than atelic ones: 84.4% vs. 15.6%. Other studies (cf. Balvet et al. 2018) also show that telic verbs are more frequent than atelic ones, but not to such a large extent: 72.4% vs. 27.6%. In our corpus, the achievement class is the most represented lexical aspect for all affixation patterns, then come accomplishments, activities and states.

2.2 Count/mass distinction

The count/mass distinction is primarily a grammatical distinction, yet it does to a certain extent have ontological properties. It is complex and gradable (Joosten 2003, Chierchia 2010, Rothstein 2010, Kleiber 2014, Timotheus & Lauwers 2015 among others). Modification by a numeral is generally taken to be a diagnosis for countability (1), while *un peu de* N 'a little bit of N' is a diagnosis for non-countability in French (2):

(1) *deux fourrages/laits/butanes/mousses
(2) un peu de fourrage/lait/butane/mousse

The two tests have been used for this study. The annotation for the count/mass distinction comprised of two steps: (i), annotation of the bN regardless of the context (in order to take into account cases of polysemy); (ii) annotation of the base N as a base for a deverbal noun.

From a total number of 313 bNs, 227 (72.5%) are count, while 86 (27.5%) are mass nouns. All affixation patterns have more count bNs than mass bNs (always more than 70%), although the [N-*iser*]$_V$ pattern has only 55.8% of count bNs.

3 Analysis

The results obtained from the annotation task verify to a large extent our initial hypothesis, since there is a clear correlation between the aspectual denotation of dNVs and the

countability properties of the corresponding bNs: telic dNVs mostly derive from count bNs, while atelic dNVs mostly derive from mass dNVs. However, as shown in Table 2, these results are clearer for telic dNVs than for atelic dNVs: the former are built on count nouns in 77% of cases, while the latter are built on count nouns in 46% of cases. In other words, the countability/aspect relationship seems stronger in the case of a bounded (count/telic) semantics than in the case of an unbounded (mass/atelic) semantics.

	Telic	Atelic
Count	204 (77%)	23 (46%)
Mass	60 (23%)	26 (54%)
Total	264	49

Table 2. Aspect and countability

According to our hypothesis, based on Harley's, which states that dNVs formed from mass bNs are compatible with both an atelic and a telic interpretation, the only counter-examples found in our results are, strictly speaking, atelic dNVs formed from count bNs. As we will show in depth, most of these 23 cases, should be treated as marginal cases for different reasons: (i) the semantic relation between the dNV and the bN is very weak, as in *s'adonner* (< *don*); (ii) the (count) bN must be interpreted as plural, as in *embrasser* (< *bras*).

Nonetheless, we find some true counter-examples, as *dériver* (< *rive*) or *favoriser* (< *faveur*), although they are very rare and thus cannot be used to deny the strong correlation we have found between aspect and countability. In the case of converted verbs, Rimell (2012: 114) also found some counter-examples of count nouns that allow the formation of atelic verbs, as *braid*. On the other hand, the relationship between the countability of the bN and the aspect of the dNV and type of affix is illustrated by Table 3.

	Telic		Atelic	
	Count	Mass	Count	Mass
a-	32 (84%)	6 (16%)	5 (45%)	5 (55%)
dé-	31 (82%)	7 (18%)	2 (100%)	0 (0%)
é-	29 (85%)	5 (15%)	4 (57%)	3 (43%)
en-	56 (81%)	13 (19%)	6 (50%)	6 (50%)
-ifier	18 (75%)	5 (25%)	1 (33%)	2 (67%)
-iser	38 (61%)	24 (29%)	5 (33%)	10 (67%)
total	204	60	23	26

Table 3. Countability-aspect relationship and affixation pattern

In the case of telic verbs, the four prefixed dNVs show a very similar behavior: they are always built on count nouns in a range between 81% and 85%. Suffixed dNVs are slightly different: they are built on count nouns in a range between 61% (-*iser*) and 75% (-*ifier*). As for atelic verbs, we do not have enough cases of any affix to draw clear conclusions. In any case, there is a difference between prefixed and suffixed dNVs that should be explained.

4   Conclusions and perspectives

The hypothesis presented in the introduction is by and large confirmed by our data: most telic dNVs are built on a count bN and most atelic dNVs are built on a mass bNs – yet some dNVs do not behave this way, as is the case for English converted verbs.

This study is part of a broader line of research dealing with the transmission and inheritance of semantic features across parts of speech. We assume that cross-categorial derivation is an ideal testing ground for the investigation of the semantic and conceptual features of grammatical categories and that it gives us more insight into the ontology of natural language.

References

Bach, E. 1986. The algebra of events. *Linguistics and philosophy, 9*(1), 5-16.

A. Balvet, P. Haas, A. Kupsc & Marín, R. (2018). Looking for Syntax/Aspect Mappings: a Case Study on the French Treebank. Grammar and Corpora 2018 (GAC 2018), Université Paris 7.

Bogacki, C. 1988. Les verbes à argument incorporé en français. *Langages,* 89, 7-26.

Chierchia, G. 2010. Mass nouns, vagueness and semantic variation. *Synthese, 174*(1), 99-149.

Clark, H. H. & Clark, E. V.. 1979. When nouns surface as verbs. Language, 55, 767-811.

Dal, G. & Namer, F. 2000. Génération et analyse automatiques de ressources lexicales construites utilisables en recherche d'informations. *Traitement automatique des langues, 41*(2), 423-446.

Dowty, D. 1977/1979. *Word Meaning and Montague Grammar: the Semantics of Verbs and Times in Generative Semantics and Montague's PTQ,* Doredrecht: Reidel.

Garey, H. B. 1957. Verbal aspect in French, *Language,* 33, 91-110.

Hale, K. & Keyser, S. 1993. On argument structure and the lexical expression of syntactic relations. In K. Hale and S. Keyser, *The view from Building 20: essays in linguistics in honor of Sylvian Bromberger,* 53–109. Cambridge: MIT Press.

Harley, H. 1999. Denominal verbs and aktionsart. *MIT Working Papers in Linguistics,* 35, 73–85.

Harley, H. 2005. How do verbs get their names? Denominal verbs, manner incorporation, and the ontology of verb roots in English. The syntax of aspect, *1*(9), 42–65.

Hirschbuhler, P. & Labelle, M. 2008. Alternance locative et Effet d'Incomplétude : différences interlinguistiques. *Linguista Sum, Mélanges offerts à Marc Dominicy à l'occasion de son soixantième anniversaire,* 251-264.

Joosten, F. 2003. Accounts of the Count-Mass Distinction: A Critical Survey, *Nordlyd, 31*(1), 216-229.

Kenny, A. 1963/[1994]. *Action, Emotion and Will,* Bristol: Thoemmes Press.

Kleiber, G. 2014. Massif/comptable: d'une problématique à l'autre. *Langue française,* 3, 3-24.

Lignon, S. 2010. –iser and–ifier suffixations in French: Verify data to verize hypotheses? *Décembrettes 7.* Lincom Europa.

Namer, F. 2013. Adjectival bases of French-aliser and-ariser verbs: syncretism or under-specification. *Selected Proceedings of the 7th Décembrettes: Morphology in Toulouse, München: Lincom Europa,* 185-210.

Plag, I. 1999. *Morphological productivity: Structural constraints in English derivation.* Walter de Gruyter, 2012.

Rimell, L. 2012. *Nominal roots as event predicates in English denominal conversion verbs.* Doctoral dissertation, New York University.

Rothstein, S. 2004. *Structuring Events: A Study in the Semantics of Lexical Aspect*, Blackwell: Oxford.

Rothstein, S. 2010. Counting and the mass/count distinction. *Journal of semantics*, *27*(3), 343-397.

Talmy, L. 1978. The relation of grammar to cognition: A synopsis. In *Proceedings of the 1978 workshop on Theoretical issues in natural language processing,* Association for Computational Linguistics, 14-24.

Timotheus, V., & Lauwers, P. 2015. L'opposition massif/comptable. In *Encyclopédie grammaticale du français,* 1-47.

Tribout, D. 2011. Les conversions de nom à verbe et de verbe à nom en français. Doctoral dissertation, Université Paris Diderot Paris 7.

Sagot, B. & Fort, K. 2009. Description et analyse des verbes désadjectivaux et dénominaux en-ifier et-iser. In *Colloque international sur le lexique et la grammaire,* Vol. 4, 102-109.

Vendler, Z. 1967. *Linguistics in Philosophy,* Ithaca N. Y.: Cornell University.

Willems, D. 1979. Syntaxe, morphosyntaxe et sémantique. Les verbes dérivés. *Cahiers de Lexicologie Paris*, *35*(2), 3-2.

Paradigm migration in the QoTeT verbs of Modern Hebrew

Noam Faust

Université Paris 8/ CNRS SFL

The verbal type QiTeL of Modern Hebrew exhibits two unsuffixed stems (1a). Most verbs in QiTeL involve three different consonants, like (1a), but there are triconsonantal stems with identical 2nd and 3rd consonants. Such verbs appear in one of two patterns, QiTeT and QoTeT (1b,c). At least since Bat-El (1994), a migration from QoTeT to QiTeT has been noted. It is possible to hear speakers use the past <i,e> vocalization for verbs normatively exhibiting <o,e>, as illustrated by the variation in (1d).

(1)	*past stem*	*non-past stem*	*gloss*
a. QiTeL	[diber]	[-daber]	'speak'
b. QiTeT	[dimem]	[-damem]	'shut down (an engine)'
c. QoTeT$_1$	[domem]	[-domem]	'bleed'
d. ??	[roʃeʃ]~[riʃeʃ]	[-roʃeʃ], *[raʃeʃ]	'impoverish'
e. hitQaTeL	[hitχanef]	[-itχanef]	'suck up'
f. hitQaTeT	[hitχanen]	[-itχanen]	'implore'
g. hitQoTeT	[hitkonen], *[hitkanen]	[-itkonen], *[-itkanen]	'prepare'

In terms of vocalization, QiTeT verbs behave like QiTeL verbs. The exceptionality of QoTeT in this respect was claimed by Bat-El (1994) to motivate the migration in (1d). However, this cannot be the entire story, since *only the past stem* exhibits variation. If such entries migrated to the paradigm in (1a,b), one would expect both stems to exhibit variation. Instead, a new paradigm is born, with the <i,e> vocalization in the past and the <o,e> vocalization in the non-past.

Importantly, as shown in (1e-g), no migration is attested in the verbal type hitQaTeL. In this type, as in QiTeL, stems with identical 2nd and 3rd consonants involve either the regular vocalization or an exceptional vocalization <o,e>; yet no migration is attested from hitQoTeT to hitQaTeT. Why is the exceptionality in hitQoTeT more stable than the exceptionality in QoTeT?

I argue that what is exceptional about the QoTeT paradigm is not its vocalization but its pattern of syncretism. Unlike QiTeL/QiTeT, QoTeT verbs do not distinguish between past and non-past stems. For this reason, the past stem, and only this stem, is assimilated to the general pattern; the <o,e> vocalization of the non-past stem does not pose a problem as long as this stem is different from the past stem. This reasoning also motivates the absence of migration from hitQoTeT to hitQaTeT: both paradigms are entirely syncretic to begin with. The pressure in the system is thus one of assimilation in *paradigm shape,* rather than stem shape. This is reminiscent of Maiden's (2004) "morphomic" view, according to which paradigm shape is a morphological object. Interestingly, as in the present case, such pressure can end up reducing syncretism. This move would otherwise be unmotivated (disambiguation cannot be a factor, since no two word-forms in the past and non-past are ever homophonous even if the stem is identical).

This reanalysis raises another question. If the goal is to distinguish past and non-past, why is the past stem assimilated and not the future stem? Why is the new paradigm [riʃeʃ, -roʃeʃ] and not *[roʃeʃ, -raʃeʃ]? There are two possible explanations: **i.** the non-past form is more widespread within the word-shapes of the paradigm, and thus more resistant to change

(McCarthy 2005); **ii.** the vocalization /i/ is a specific marker of [+past] QiTeL, whereas /a/ is a default vowel in the verbal system (Dor 1995).

I concentrate on the second view, as it does not require counting cells in the inflectional paradigm. According to this view, /a/ is inserted in the template whenever there is no specific exponent for the morphological feature bundle requiring realization. Thus, in a pair such as (1a) [diber, -daber], the vowel /e/ in V_2 realizes the verbal class, and the vowel /i/ in V_1 realizes the feature [+past]; but the feature [-past] is not related to a specific realization, and so the non-past V_1 is realized as [a]. QoTeT verbs involve a specific exponent $V_1 = [o]$. In the non-past, this exponent takes the place of the default [a], because it is more specific. In the past form, however, there *is* a non-default vowel /i/ associated with this meaning, whose position is *also* V_1. It is unclear which exponent is more specific, and so either can be selected. For a paradigm of the type (1c), the verb-specific exponent has the upper hand. The variation in paradigms like (1d) nevertheless expresses this equal degree of specification.

Under such an understanding, what are the factors that may influence the choice between two equally-specific realizational statements? The morpheme or canonical paradigm shape is only one possible answer to this question. If it can be shown that there are other factors – for instance, frequency – there might not be a need to appeal to morphemes in this case.

References

Bat-El, Outi. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language & Linguistic Theory* 12/4: 571–596.

Dor, Danny. 1995. Deriving the Verbal Paradigm of Modern Hebrew: A constraint-Based

Approach. In van der Hulst, H. and van de Weijer, J. (eds.), *Leiden in Last: HIL Phonology Papers I*. The Hague: Holland Academic Graphics. 105-144.

Maiden. Martin. 2004. Morphological autonomy and diachrony. *Yearbook of Morphology*. 137–175.

McCarthy, John J. 2005. Optimal paradigms. In Laura Downing, Tracy Alan Hall, and Renate Raffelsiefen *(eds.) Paradigms in Phonological Theory*. Oxford: Oxford University Press. 170-210.

The continuing challenge of the German gender system

Sebastian Fedden   *Greville G. Corbett*

Université Sorbonne Nouvelle   University of Surrey

1 Introduction

We have come a long way since Mark Twain (1880) suggested that German gender had 'no sense or system', a sentiment echoed in Bloomfield (1933: 280). Researchers such as Klaus-Michael Köpcke and David Zubin have demonstrated clear regularities in the assignment of gender to German nouns. However, the German gender system has not yielded all of its secrets; we know many of the parts but the whole still eludes us. It deserves continuing research, as one of the most complex systems, with interacting semantic, morphological and phonological assignment principles. The challenge can be stated simply: given contexts like those in (1a) - (1c), how do speakers of German consistently assign a gender value (masculine, feminine, or neuter) to nouns, and hence produce the agreements?

(1)   a. ein   neu-er   Film
          a[NOM.M/N.SG]   new-NOM.M.SG   film(M)[NOM.SG]
          'a new film'

      b. ein-e   neu-e   Symphonie
          a-NOM.F.SG   new-NOM.F.SG   symphony(F)[NOM.SG]
          'a new symphony'

      c. ein   neu-es   Buch
          a[NOM.N/N.SG]   new-NOM.N.SG   book(N)[NOM.SG]
          'a new book'

The basic semantic assignment rules are straightforward. Sex-differentiable nouns, i.e. nouns which refer to male or female humans or male or female (higher) animals, e.g. *der Mann* 'man', *die Frau* 'woman', *der Bulle* 'bull', *die Kuh* 'cow', are assigned gender on the basis of biological sex. (We adopt the convention of giving nouns with the appropriate definite article (nominative singular) to indicate gender, namely *der* (masculine), *die* (feminine), and *das* (neuter), e.g. *das Buch* [the.N bed(N)] 'the book'.) For most German nouns, however, including the inanimates in (1), gender assignment is formal, that is, it follows their morphology or their phonology. Phonological assignment rules have been investigated in detail (Köpcke 1982; Köpcke & Zubin 1983), and there are studies of the relation between gender and inflection class (Augst 1975; Bittner 1999; Kürschner & Nübling 2011). Yet despite the typological interest of the German system, and the analytical progress on parts of the system, there is no comprehensive analysis of the whole system, as has been done for other languages which rely on a combination of semantic and formal assignment rules, such as Russian (Corbett 1991; Corbett & Fraser 2000).

We will discuss morphological (§3) and phonological (§4) assignment of German nouns. We will pay particular attention to an interesting suggestion, been widely discussed, namely that German has semantic clusters which are associated with a certain gender (§5). These clusters involve smaller numbers of nouns than the semantic assignment rules; for example, most alcoholic drinks are masculine.

2 The overall numbers

Given the complexity of the system, an important step is to put numbers to predictions. Thus when evaluating a suggested morphological or phonological assignment rule, we compare this rule to the overall distribution of nouns over the gender values. This allows us to ascertain whether prediction by the rule is better than chance. Table 1 gives percentages according to noun frequency, based on the CELEX database (Baayen et al. 1995).

Table 1. Overall gender distribution in German (rounded to full percentages)

Most frequent …	MASCULINE	FEMININE	NEUTER
… 100 nouns	34%	35%	31%
… 1,000 nouns	37%	41%	22%
… 10,000 nouns	36%	45%	19%

The difference in proportions depending on frequency-based sample size is revealing. A key factor is that a larger sample contains more derived nouns, and derivation favours the feminine in German. The distribution raises intriguing questions for acquisition, which starts with the least regular part of the lexicon for gender assignment.

3 Morphological assignment

There are two sides to morphological assignment, word formation and inflection. We start with word formation, where overall we find better prediction rates than in inflection.

3.1 Word formation

Morphologically complex German words are governed by the Last Member Principle (*Letzt-Glied-Prinzip*, see Köpcke & Zubin 1984: 28-29, and references there): the gender of the whole word is determined by the gender of the last element. In compounds the last element is a word with its own gender value. For example *der Mutterschutz* 'maternity' consists of the feminine first member *die Mutter* 'mother' and the masculine last member *der Schutz* 'protection'; by the Last Member Principle it is masculine. Derivational affixes are similarly associated with a gender value, which is assigned to the derived word irrespective of the gender of the base (if this is a noun). For example, the suffix *-schaft* derives feminine nouns, e.g. *die Freundschaft* 'friendship' from the masculine noun *Freund* 'friend', or *die Landschaft* 'landscape' from the neuter noun *Land* 'land'.

While assignment for nouns derived with productive suffixes is virtually without exception, e.g. all nouns derived with *-heit* and *-keit* are feminine, less productive derivational affixes are often not as good a predictor. The majority of nouns derived with *-el* are masculine, but we find more of a spread of genders, e.g. in *der Würfel* 'cube, die', *die Klingel* 'bell', *das Kürzel* 'abbreviation, code'. For these nouns derived with *-el*, of 602 nouns, 59% are masculine, 25% are feminine, and 16% are neuter (Augst 1975: 30-32). This is presumably because a decrease in productivity of an affix makes it harder for speakers to associate a single gender value with it. Nevertheless, compared with the figures in Table 1, this generalization has some effect.

3.2 Inflection

Three example paradigms are given in (2). Together they provide evidence for four cases, which are distinguished only poorly in individual inflection classes (ICs). ICs are identified by their principal parts (i.e. genitive singular/nominative plural).

(2) Forms for *Biene* 'bee' (IC –/-(e)n), *Tag* 'day' (IC -(e)s/-e), *Zeuge* 'witness' (IC -(e)n/-(e)n)

	SINGULAR	PLURAL	SINGULAR	PLURAL	SINGULAR	PLURAL
NOMINATIVE	Biene	Bienen	Tag	Tage	Zeuge	Zeugen
ACCUSATIVE	Biene	Bienen	Tag	Tage	Zeugen	Zeugen
GENITIVE	Biene	Bienen	Tag(e)s	Tage	Zeugen	Zeugen
DATIVE	Biene	Bienen	Tag	Tagen	Zeugen	Zeugen

The perspective on German gender and IC used to be one-directional: typically the method was to start from gender to predict IC (see, for example, Augst 1975: 24-36; Bittner 1999). More recently, one also finds approaches where both gender and IC have predictive value (Kürschner & Nübling 2011). Given that there are more than a dozen ICs, though not of equal size (Pavlov 1995: 44), and just three genders, the logic suggests that we are likely to find prediction of gender from IC. We investigate both directions.

Given the gender of a noun what can we say about its inflection? The default IC for feminine nouns is *–/-(e)n*, cf. the paradigm of *die Biene* 'bee' in (2) above. Prediction of IC from gender is not fully reliable, since there are at least four other ICs which contain feminine nouns, all of them extremely small in terms of membership. According to Pavlov (1995: 46), more then 95% of feminine nouns choose the majority pattern *–/-(e)n*. While the fact that a noun is feminine allows us to predict its paradigm in almost all cases, this is not the case for masculine or neuter nouns where we find several ICs with a substantial membership.

We now turn to the other direction. If we know how a noun inflects, how reliably can we predict its gender? For at least four ICs, gender can unambiguously (or nearly unambiguously) be predicted from IC, e.g. all nouns of IC *-/-(e)n*, i.e. the class that *Biene* 'bee' belongs to, are feminine. Then there are several ICs whose nouns cannot be feminine. For instance, we can predict that *der Knauf* 'knob' (a masculine noun following inflection class *-(e)s/*U-*e*, U indicating umlaut) cannot be feminine based on its paradigm. To tell masculine nouns apart from neuter nouns can be less straightforward, but often further phonological cues help (see §4).

4 Phonological assignment

Köpcke (1982) and Köpcke & Zubin (1983) establish a number of phonological rules to account for the gender of monosyllabic nouns. For example, almost all monosyllabic nouns starting with the cluster /kn/ are masculine (93%), e.g. *der Knauf*, 'knob', *der Knick* 'crease', the only exception being the neuter noun *das Knie* 'knee'. The great majority of nouns ending in /et/ are neuter (86%), e.g. *das Duett* 'duet', an exception being *der Anisett* 'anisette'. The majority of nouns which end in the specific clusters /ft/, /xt/ or /çt/ are feminine (64%), e.g. *die Zunft* 'guild', *die Frucht* 'fruit', *die Sicht* 'visibility', an exception being *der Duft* 'smell'. And in general, the more consonants a monosyllabic noun has in its onset or coda, the higher the probability that the noun is masculine. These phonological assignment rules are not exceptionless, but in comparison to the overall numbers in Table 1, they have some value.

5 'Crazy' rules

Semantic non-core rules, termed 'crazy' rules by Enger (2009), add a new dimension to gender assignment. They involve semantic clusters; for example, nouns denoting alcoholic

beverages are masculine, introvert affect nouns are feminine, and basic chemical elements are neuter (Köpcke & Zubin 1984, 1996, Zubin & Köpcke 1984). These are only tendencies, with exceptions, e.g. the neuter alcoholic drink *das Bier* 'beer'. Moreover, in contrast to semantic assignment rules, crazy rules can be overridden by form, e.g. the extrovert affect noun *die Strenge* 'sternness' should be masculine by 'crazy' rule, but is in fact feminine because it belongs to a class of (no longer productive) deadjectival derivations in *-e* (from *streng* 'stern'), which are all feminine.

6 Conclusion

The German gender system involves many interacting regularities, of different types: semantic, phonological, morphological and 'crazy' rules. We suggest that progress can be made by treating the system as a whole and by putting numbers to assignment rules so that we can evaluate their effectiveness. We reaffirm the unique typological interest of the German gender system, and make modest progress in illuminating it.

References

Augst, Gerhard. 1975. *Untersuchungen zum Morpheminventar der deutschen Gegenwartssprache*. Forschungsberichte des Instituts für deutsche Sprache Mannheim 25. Tübingen: Gunter Narr.

Baayen, R. Harald, R. Piepenbrock & L. Gulikers. 1995. The CELEX Lexical Database (CD-ROM), Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Bittner, Dagmar. 1999. Gender classification and the inflectional system of German nouns. In Barbara Unterbeck (ed.), *Gender in grammar and cognition*, Part 1: *Approaches to gender*, 1-23. Berlin: Mouton de Gruyter.

Bloomfield, Leonard 1933. *Language.* New York: Holt, Rinehart and Winston.

Corbett, Greville G. 1991. *Gender.* Cambridge: CUP.

Corbett, Greville G. & Norman M. Fraser. 2000. Gender assignment: a typology and a model. In Gunter Senft (ed.), *Systems of nominal classification*, 293-325. Cambridge: CUP.

Enger, Hans-Olav. 2009. The role of core and non-core semantic rules in gender assignment. *Lingua* 119. 1281-1299.

Köpcke, Klaus-Michael. 1982. *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübingen: Niemeyer.

Köpcke, Klaus-Michael & David A. Zubin. 1983. Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für germanistische Linguistik* 11. 166-182.

Köpcke, Klaus-Michael & David A. Zubin. 1984. Sechs Prinzipien für die Genuszuweisung im Deutschen. Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte* 93. 26-50.

Köpcke, Klaus-Michael & David A. Zubin. 1996. Prinzipien für Genuszuweisung im Deutschen. In Ewald Lang & Gisela Zifonun (eds.), *Deutsch-typologisch*. Jahrbuch des Instituts für Deutsche Sprache 1995, 473-491. Berlin: De Gruyter Mouton.

Kürschner, Sebastian & Damaris Nübling. 2011. The interaction of gender and declension in Germanic languages. *Folia Linguistica* 45(2). 355-388.

Pavlov, Vladimir. 1995. *Die Deklination der deutschen Substantive. Synchronie und Diachronie*. Frankfurt a. M.: Peter Lang.

Zubin, David & Klaus-Michael Köpcke. 1984. Affect classification in the German gender system. *Lingua* 63. 41-96.

Morphology, Semantics, and Pragmatics
of Negative Rhetorical Questions in Megrelian

Rusudan Gersamia   *Alexander Rostovtsev-Popiel*

Ilia State University, Tbilisi, Georgia   Collège de France, Paris, France

This paper discusses a complex set of morphological and semantic features that characterize negative rhetorical questions in Megrelian, a Kartvelian (South Caucasian) language spoken in Western Georgia. These include such diverse linguistic parameters as affirmation, negation (including double negation), verbal aspect, and presupposition, whereby the morphological means involved is circumfixation of complex verb forms that already carry prefixes and suffixes. These parameters have been discussed in the literature on Megrelian severally (Gudava & Gamq'reliʒe 1981/1987; Harris 1991; Reseck 2014), but have never been treated together as parts cumulatively inducing one phenomenon, namely negative rhetorical question constructions. The aim of this contribution is thus to assemble this puzzle and explain its structure.

The Kartvelian verb is well-known for its complexity, and it is Megrelian that stands out from the family in terms of the degree that this complexity reaches: this language has both the highest number of TAM paradigms and the most sophisticated system of verbal inflection, in view of the number of affixes attachable to the root and the rules applying to their interaction. Tables 1 and 2 illustrate, in a simplified manner, the slot structure of the finite verb in Megrelian:

Table 1   Finite verb prefixation

SLOT	−6	−5	−4	−3	−2	−1
MARKER	AFF/FOC/PFV/NEG/PROH	PV(:PFV)	IMPFV	EVID	IO/DO/S	VERS
	ko-, ge-, o-, va-, num-	*mic'o-…*	*tima-…*	*no-*	*v-, g-…*	*i-…*

Table 2   Finite verb suffixation

SLOT	+1	+2	+3	+4	+5	+6	+7	+8
MARKER	R.EXT	AUX/CAUS	ITR	SM	EM	IND/SBJV	IO/DO/S (SG/PL)	COND
	-on…	*-apu…*	*-d*	*-en…*	*-d*	*-i, -a…*	*-s, -u…*	*-k'on*

Note that slot −5 preverbs are involved in lexical derivation and can, under particular circumstances, additionally perfectivize the verb, thus yielding inherently perfective lexical items featuring both finite and nonfinite forms. At the same time, the perfective aspectual value expressed by some of the slot −6 prefixes (which only occur in finite verb forms), viz. *ge-*, *o-*, and, more rarely, *ko-*, is confined to the domain of inflection.

Apart from the affixes presented in Tables 1–2, the finite verb can attach clitics of diverse function, including the question enclitic =*o* that is obligatory on the predicate in polar questions and ruled out in content question constructions; note that =*o* a clitic and therefore not included in the suffixation template, for it can be separated from the verb by other clitics, such as quotative =*a* (cf. ex. 1) that are also attachable by other parts of speech (e.g. nouns, cf. ex. 2):

(1)   *kirsian rekiavo? k'itx.*
        kirsian   r-e-k=i=<u>a</u>=v=**o**   k'itx-[u]
        Christian   be-SM-S2SG=EV=<u>QUOT</u>=EC=**QST**   ask-[S3SG.PST]

'Are you Christian? <u>X asked Y</u>.' (Xubua 1937: 45.29)

(2)   *kumop'idi ate skani lak'via.*
       ko-mo-m-ʔid-i   ate   skan-i   lak'v-i = <u>a</u>
       PFV-PV-IO1-buy-IMP   this   your-NOM   puppy-NOM = <u>QUOT</u>
       'Sell me this puppy of yours, <u>X said to Y</u>.' (Xubua 1937: 7:10)

Together with the negative marker *va=* that is otherwise to be found in slot –6, the question enclitic forms a clitic circumfix, or, rather, a circumclitic, *va=__ =o*, whose function is to create a verb form that can serve as a predicate in a negative rhetorical question construction that implies an affirmative answer:

(3)   *dasuro koɣolə ʒɣabik irpeli, muč'o uc'ɯɯ bošikən, teši, vakiʔɯɯo martalk?*
       dasuro   ko-ɣol-u   ʒɣab-i-k   irpel-i
       indeed   AFF-do-S3SG.PST   girl-B-ERG   everything-NOM
       muč'o   u-c'w-u   boš-i-k = ni   teši
       how   VERS-tell-S3SG.PST   boy-B-ERG = COMP   that way
       va = ko-i-ʔ-u = **o**   martal-k
       NEG = AFF-VERS-be-S3SG.PST = **QST**   true-ERG
       'Indeed, the girl did everything as the boy had told her, (exactly) that way, and (**do you think** everything) **didn't** prove to be true? (i.e. it did.)' (Xubua 1937: 55.13-14)

(4)   *mioǯineso, dasuro vagemc'odirtuo ʒɣabik eše?*
       mi-o-ǯin-es = o   dasuro
       PV-VERS-look-S3PL.PST = COMP   indeed
       va = ge-mic'o-dirt-u = **o**   ʒɣab-i-k   eše
       NEG = PFV-PV-stand.up-S3SG.PST = **QST**   girl-B-ERG   upwards
       'When they looked (at her), **didn't** the girl **really** stand up? (i.e. she did.)' (Xubua 1937: 25.1)

(5)   *do vaoč'k'omɯo?*
       do   **va** = o-č'k'om-u = **o**
       and   **NEG** = PFV-eat-S3SG.PST = **QST**
       'And **didn't** X eat Y? (i.e. X did.)' (Rostovtsev-Popiel 2011: 205)

The formation of such verb forms requires the verb to carry a slot –6 inflectional prefix (either multiuse *ko-* whose primary function is to mark affirmation and focus, and perfective aspect as well—among its secondary functions, ex. 3; dedicated perfectivizer *ge-* used with preverbed verbs, i.e. such verbs that carry a slot –5 preverb, ex. 4; or dedicated perfectivizer *o-* attached by verba simplicia, i.e. those without a slot –5 preverb, ex. 5). Consequently, with a view to the combinability of the negative marker and the affirmative marker *ko-*, this phenomenon was addressed as "combining the uncombinable" (Rostovtsev-Popiel 2011; see also Boeder 2013 for discussion). However, as was maintained later on in (Rostovtsev-Popiel 2012), it was rather the perfective aspectual function of slot –6 prefixes that made it possible to juxtapose the affirmative marker *ko-* and the negative marker *va=* within one verb form (cf. absence of this phenomenon in imperfectives). Furthermore, there is also recent fieldwork evidence of circumfixation, or, rather, circumcliticization, of verb forms, such that already carry the negative marker *va-* (as well as its positional variants, such as

ve-, vo-, v-), i.e. one of further mutually-exclusive slot –6 prefixes. This inflectional operation yields double negation:

(6) *gemi tiʒgura xargeli rdu, šk'a ɣwas išo vavedinc'q'uo?*

gem-i	tiʒgura	xargel-i	r-d-u
ship-NOM	that.kind	loaded-NOM	be-EM-S3SG.PST
šk'a	[z]ɣwa-s	išo	**va** = <u>va</u>-do-i-nc'q'-u = **o**
middle	sea-DAT	thither	**NEG₁** = <u>NEG₂</u>-PV-VERS-sink-S3SG.PST = **QST**

'(Given that) the ship was so loaded, **didn't** she sink to pot in the middle of the sea? (i.e. she did.)'

As the translation suggests, such constructions do not differ semantically from those formed after models *va = ko-__ = o*, *va = ge-__ = o*, and *va = o-__ = o* and convey an affirmative presupposition, whereby the proclitic part of the circumclitic can be dropped, as in e.g. ***vedinc'q'uo?*** '**didn't** she sink? (i.e. she did)', thus eliminating the double negation pattern. It is worth mentioning here that in their pilot work on Megrelian morphophonemics, Gudava and Gamq'reliʒe pointed at the fact that prefixal *va-* and *va =* that occurs in negated rhetorical questions should not be treated as one same expression (Gudava & Gamq'reliʒe 1981/1987: 239); however, the authors, within the framework of their paper, neither expanded on the latter *va =* as part of a circumclitic nor discussed its semantic and pragmatic functions.

Basing upon available corpora, extant textual materials, and our first hand fieldwork data, we aim to provide a comprehensive account of morphological, semantic, and pragmatic properties of the phenomenon in question, both synchronically and diachronically—in particular, against the background of the data found in the sister languages, and to explain how diverse domains of linguistic structure interact in the formation of negative rhetorical questions in Megrelian. Special attention will be drawn to lexically-determined aspectual properties of verb lexemes discussed and their interplay with the inflectional perfective aspect value, as well as to constraints on the appearance and retention of certain combinations of affixes and clitics throughout the Kartvelian family.

Abbreviations

2 – 2nd person; 3 – 3rd person; AFF – affirmative marker; AUX – auxiliary; B – base; CAUS – causative; COMP – complementizer; COND – conditional; DAT – dative; DO – direct object; EC – euphonic consonant; EM – extension marker; ERG – ergative; EV – euphonic vowel; EVID – evidential; FOC – focal marker; IMPFV – imperfectivizer; IND – indicative; IO – indirect object; IMP – imperative; ITR – intransitivizer; NEG – negative marker; NOM – nominatve; PFV – perfectivizer; PL – plural; PROH – prohibitive; PST – past; PV – preverb; QST – question marker; R.EXT – root extension; S – subject; SBJV – subjunctive; SG – singular; SM – series marker; VERS – versionizer.

References

Boeder, Winfried. 2013. A Note on the Pragmatics of Interrogativity in Megrelian. *Enatmecnierebis sak'itxebi*, 97–120.

Gudava, T'ogo & Tamaz Gamq'reliʒe. 1981/1987. Tanxmovant k'omp'leksebi megrulši [Consonant Clusters in Megrelian]. In Elguʒa Xintibiʒe (ed.). *Ak'ak'i Šaniʒe – 100* [Ak'ak'i Šaniʒe's 100th Anniversary Festschrift]. Tbilisi: Tbilisi State University Press: 202–243.

Harris, Alice C. 1991. Mingrelian. In Alice C. Harris (ed.). *The Indigenous Languages of the Caucasus. Vol. 1: The Kartvelian Languages*. Delmar / New York: Caravan: 313–394.

Reseck, Tamar. 2014. *Präverbien im Megrelischen*. (Diversitas Linguarum 37.) Bochum: Brockmeyer.

Rostovtsev-Popiel, Alexander. 2011. Combining the Uncombinable: An Insight from Megrelian. *The III International Symposium on Ibero-Caucasian Languages: Structure, History, Functioning*. Tbilisi: 204–206.

Rostovtsev-Popiel, Alexander. 2012. *Grammaticalized Affirmativity in Kartvelian*. Ph.D. Dissertation, Frankfurt University.

Xubua, Mak'ar. 1937. *Megruli t'ekst'ebi* [Megrelian Texts]. Tbilisi: Georgian Academy of Sciences Press.

Entropy, analogy and paradigm structure

Matías Guzmán Naranjo
Heinrich-Heine Universität Düsseldorf

1 Introduction

Recent work on paradigm organization has focused on the question of how speakers can deduce the complete paradigm of a lexeme given that they only encounter a limited number of inflected forms of that lexeme (Ackerman et al., 2009). This is also known as the paradigm cell filling problem (PCFP). These studies have proposed entropy (Shannon, 1948) as a measurement of paradigm complexity/predictability. The basic idea is that one can calculate the conditional entropy between two cells of a paradigm, which measures the predictability between cells (i.e. how much information does Cell 1 provide about Cell 2 of a paradigm). Bonami & Beniamine (2016) have even expanded this approach to work for multiple cells.

Entropy-based approaches have a serious of limitations, however. First, entropy is not a normalized metric, which makes it unreliable for comparing different systems/languages. Second, many studies have convincingly shown that the inflection class of a lexeme is predictable from its phonology and semantics (Bybee & Slobin, 1982; Skousen, 1992; Eddington, 2002; Matthews, 2005; Blevins et al., 2017), which is information that entropy cannot easily take into account.

Using the Russian nominal inflection system as an example, I will argue that analogical classification (i.e. class assignment on the basis of similarity) offers a convincing solution to the PCFP, and that accuracy metrics are a better measurement of predictability/paradigm complexity than entropy.

1.1 Materials and methodology

From the Grammatical Dictionary of Russian by Zaliznyak (1977), I extracted all nouns (43412) with their complete paradigm (including the prepositional case). I then converted the extracted forms to a phonological transcription using epitran (Mortensen et al., 2018). This phonological transcription is not perfect but it is a reasonable approximation of the Russian system. For the present study I did not consider stress but this feature could be easily included into the models.

Many accounts of Russian nominal inflection have been proposed in the literature, each suggesting a different analysis of the inflection classes found in the Russian system (Fraser & Corbett, 1995, for a well known example). To sidestep these discussions, I extracted the inflection class of each noun automatically with a surface-based method. The method is as follows:

1. find the non-continuous phonological sub-sequence common to all cells in the paradigm of a lexeme (from now on the *stem*),
2. remove this sub-sequence from each cell. In cases of discontinuous sub-sequences add a separation mark (-),
3. the result in each cell is the *marker* for that cell,
4. the inflection class of the lexeme is the set of markers for all cells.

Because this method makes no assumptions about underlying representations, it is very conservative and thus it produces the maximum possible number of inflection classes. As an

Cell	sgular form	marker	plal form	marker
NOM	fʲirma	-a	fʲirmɨ	-ɨ
GEN	fʲirmɨ	-ɨ	fʲirm	-ø
DAT	fʲirm'e	-'e	fʲirmam	-am
ACC	fʲirmu	-u	fʲirmɨ	-ɨ
INS	fʲirmoj; firmou	-oj; -ou	fʲirmami	-am'i
PRE	fʲirm'e	-'e	fʲirmax	-ax

Table 1: Markers for *фирма*.

example we consider the lexeme *фирма* ('firm'). The phonological transcription of *фирма* is *fʲirma*, the longest common substring (stem) is *fʲirm*, and the resulting markers are in Table 1.

the analogical models in the next section consider both phonological and semantic information of the *stem*. To include semantic information I used the pre-trained semantic vectors provided by Kutuzov & Kuzmenko (2017) using word2vec.[1] To match a lexeme to a semantic vector I used the NOM.SG cell. From the dataset I only kept those nouns for which there was a semantic vector.

In order balance the dataset I only considered 1000 nouns for each class and removed all nouns belonging to classes with fewer than 20 nouns. Limiting the maximum number of nouns to 1000 helps the model avoid overestimating a couple of very frequent classes. The final dataset contained 17275 nouns, with 79 different inflection classes. This step also removes certain errors in the inflection class induction, as well ass irregular/suppletive forms.

On the resulting dataset I trained several analogical models using a multilayer perceptron following Guzmán Naranjo (2019).[2] For every cell in the paradigm I trained models predicting that cell from: (i) one other cell, (ii) two other cells, (iii) one cell and stem information,[3] (iv) and two cells and stem information.

For evaluation I used accuracy because this intuitively corresponds to our intuition of what predictability means Using Kappa scores or any other similar metric would also work. The important point is that accuracy metrics are normalized and therefore allow for comparisons across different models (even across different systems and languages).

---

[1] More precisely the `ruwikiruscorpora-func_upos_skipgram_300_5_2019` semantic vector data-set downloaded `http://rusvectores.org/en/models/`, accessed 17.06.2019.

[2] Each model had three hidden layers (with n*4, n*2 and n neurons respectively, where n = number of classes) with tanh activation. For all models, the learning rate was kept at 0.001, the momentum at 0.8 and dropout at 1. The models were trained with an Nvidia Titan Xp donated by the NVIDIA Corporation.

[3] The stem information consisted of the last four segments of the stem plus the semantic information in the semantic vectors (only looking at the nominative singular form).

	Predictor										
Predicted	NOM.SG	GEN.SG	DAT.SG	ACC.SG	INS.SG	PRE.SG	NOM.PL	GEN.PL	DAT.PL	ACC.PL	INS.PL	PRE.PL
NOM.SG	1	0.77	0.77	0.97	0.83	0.47	0.56	0.79	0.46	0.64	0.46	0.46
GEN.SG	0.7	1	0.89	0.7	0.84	0.63	0.63	0.74	0.57	0.73	0.57	0.57
DAT.SG	0.77	0.9	1	0.78	0.9	0.64	0.61	0.71	0.58	0.71	0.58	0.58
ACC.SG	0.82	0.61	0.61	1	0.7	0.41	0.48	0.64	0.4	0.66	0.4	0.4
INS.SG	0.77	0.76	0.77	0.75	1	0.51	0.52	0.74	0.44	0.61	0.44	0.44
PRE.SG	0.62	0.67	0.77	0.6	0.74	1	0.68	0.71	0.55	0.72	0.55	0.55
NOM.PL	0.52	0.67	0.58	0.52	0.57	0.6	1	0.64	0.38	0.89	0.38	0.38
GEN.PL	0.67	0.66	0.62	0.65	0.67	0.5	0.54	1	0.4	0.67	0.4	0.4
DAT.PL	0.97	0.99	1	0.97	0.99	0.99	1	0.96	1	0.97	0.99	0.99
ACC.PL	0.41	0.51	0.42	0.51	0.46	0.42	0.71	0.48	0.24	1	0.24	0.25
INS.PL	0.97	0.99	1	0.97	0.99	0.99	1	0.96	0.99	0.97	1	0.99
PRE.PL	0.97	1	1	0.97	0.99	0.99	1	0.96	0.99	0.97	0.99	1

Table 2: Cell predictability without stem information.

1.2 Results

Table 2 shows the accuracy score for the model predicting each cell using only one other cells as predictor. This result is comparable to the use of entropy to measure implicational relations between cells of a paradigm. This table shows that some cells in the paradimg can perfectly predict other cells. For example, DAT.SG completely predicts DAT.PL. Similarly, these results show that the ACC.SG cell is the best overall predictor of other cells in the paradigm.

However, it is also clear that most cells are not completely predictable from only one other cell. Table 3 shows how the results change once the analogical models also consider the information in the stem of the nouns. The effect is a very clear improvement.

	Predictor										
Predicted	NOM.SG	GEN.SG	DAT.SG	ACC.SG	INS.SG	PRE.SG	NOM.PL	GEN.PL	DAT.PL	ACC.PL	INS.PL	PRE.PL
NOM.SG	1	0.99	0.99	1	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98
GEN.SG	0.98	1	0.99	0.98	0.99	0.98	0.98	0.98	0.97	0.98	0.97	0.97
DAT.SG	0.99	1	1	0.99	1	0.99	0.99	0.99	0.98	0.99	0.98	0.98
ACC.SG	0.96	0.95	0.95	1	0.94	0.94	0.93	0.94	0.93	0.98	0.93	0.93
INS.SG	0.98	0.98	0.98	0.98	1	0.97	0.97	0.97	0.97	0.97	0.97	0.96
PRE.SG	0.99	1	1	0.99	1	1	0.99	0.99	0.99	0.99	0.99	0.99
NOM.PL	0.99	0.99	0.99	0.99	0.99	0.99	1	0.99	0.99	1	0.99	0.99
GEN.PL	0.99	0.98	0.98	0.99	0.99	0.98	0.98	1	0.97	0.99	0.97	0.97
DAT.PL	1	1	1	1	1	1	1	0.99	1	1	1	1
ACC.PL	0.93	0.92	0.92	0.96	0.93	0.9	0.92	0.92	0.9	1	0.9	0.9
INS.PL	1	1	1	1	1	1	1	1	1	1	1	1
PRE.PL	1	1	1	1	1	1	1	0.99	1	1	1	1

Table 3: Cell predictability including stem information.

The first thing to notice is that all cells (except those which were already at 1) in Table 3 have accuracy scores higher than the corresponding cells in Table 2. Seen in absolute terms, we can say that just knowing one form of a Russian noun (including the stem) is enough to give almost perfect predictive accuracy for 7 cells (INS.PL, INS.SG, NOM.PL, NOM.SG, PRE.PL, PRE.SG and DAT.PL), it gives a reasonable certainty for three cells (DAT.SG, GEN.PL and GEN.SG) and it gives a some certainty for the remaining two (ACC.PL and ACC.SG). It is an interesting result that these two final cells, ACC.SG and ACC.PL, are the hardest to predict from the other cells and at the same time ACC.SG is the best predictor of other cells in average.

An important point is that not all cells increased in their predictability by the same amount.

While the predictability of ACC.PL from GEN.PL increased from 0.64 to 0.79, the predictability of ACC.PL from PRE.PL increased from 0.56 to 0.78. Since in both cases we are predicting the same cell (ACC.PL), it is not the case that the stem in one model had more information than in the other model. What this shows is that the interaction between the predictor GEN.PL and the stem carries less information about ACC.PL than the interaction between the stem and PRE.PL.

It is possible that the accuracy metrics are simply restating (with a normalized metric) the same information that the information theoretic approach can already capture. The check this we can explore the correlation between the analogical models and the conditional entropy estimates as shown in Table 5.[4]. The overall correlation values for the three models and the entropy model are shown in Table 4. The entropy model and the analogical model using markers are very close to each other, while the analogical models with stem information are less so. This result is important for two reasons. First, the fact that the entropy model and the marker model capture very similar information means that, if we accept that entropy is a valid metric, accuracy is in fact a valid alternative to quantify paradigm complexity. At the same time, it is clear that adding stem information to the model does greatly change class predictability.

model	correlation
marker model	-0.95
marker + phonology + semantics model	-0.81

Table 4: Correlation with entropy values

	Predictor											
Predicted	NOM.SG	GEN.SG	DAT.SG	ACC.SG	INS.SG	PRE.SG	NOM.PL	GEN.PL	DAT.PL	ACC.PL	INS.PL	PRE.PL
NOM.SG	0.00	1.24	1.21	0.26	1.00	1.87	1.61	1.14	2.15	1.18	2.15	2.15
GEN.SG	0.88	0.00	0.43	0.77	0.68	1.30	0.95	1.02	1.74	0.78	1.74	1.74
DAT.SG	0.65	0.23	0.00	0.58	0.41	0.87	1.03	0.87	1.55	0.76	1.55	1.55
ACC.SG	0.65	1.52	1.53	0.00	1.34	2.19	1.99	1.55	2.50	1.04	2.50	2.50
INS.SG	0.74	0.78	0.71	0.69	0.00	1.40	1.48	0.95	1.93	1.03	1.93	1.93
PRE.SG	1.00	0.79	0.56	0.92	0.78	0.00	0.61	0.93	1.30	0.68	1.30	1.30
NOM.PL	1.24	0.93	1.22	1.22	1.36	1.11	0.00	1.18	1.70	0.36	1.70	1.70
GEN.PL	1.25	1.49	1.54	1.27	1.31	1.91	1.66	0.00	2.29	0.85	2.29	2.29
DAT.PL	0.14	0.09	0.09	0.09	0.18	0.16	0.07	0.17	0.00	0.07	0.00	0.00
ACC.PL	2.37	2.33	2.52	1.84	2.48	2.75	1.93	1.93	3.28	0.00	3.28	3.28
INS.PL	0.14	0.09	0.09	0.09	0.18	0.16	0.07	0.17	0.00	0.07	0.00	0.00
PRE.PL	0.14	0.09	0.09	0.09	0.18	0.16	0.07	0.17	0.00	0.07	0.00	0.00

Table 5: Conditional entropy on the Russian data-set

2 Concluding remarks

These results show that to solve the PCFP it is not enough to look at the information and predictability between markers, nor is it enough to consider the class information hidden in the stems. Both are necessary. I have shown that an analogical classifier based on a perceptron can make use of stem and marker information. With this method we can measure predictability between any number of cells, as well as making use predictors like semantic vectors, which are

---

[4]I calculated these following Ackerman & Malouf (2013)

hard to take into account with entropy-based approaches. Finally, this method allows us to calculate accuracy metrics, which are normalized and allow for easy model comparison.

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–82. Oxford, New York: Oxford University Press.

Ackerman, Farrell & Robert Malouf. 2013. Morphological Organization: The Low Conditional Entropy Conjecture. *Language* 89(3). 429–464. doi:10.1353/lan.2013.0054.

Blevins, James P., Petar Milin & Michael Ramscar. 2017. The Zipfian paradigm cell filling problem. In Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Perspectives on Morphological Organization*, 139–158. Leiden, The Netherlands: Brill.

Bonami, Olivier & Sacha Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.

Bybee, Joan L. & Dan I. Slobin. 1982. Rules and Schemas in the Development and Use of the English past Tense. *Language* 58(2). 265–289.

Eddington, David. 2002. Spanish gender assignment in an analogical framework. *Journal of Quantitative Linguistics* 9(1). 49–75.

Fraser, Norman M & Greville G Corbett. 1995. Gender, animacy, and declensional class assignment: A unified account for Russian. In *Yearbook of Morphology 1994*, 123–150. Springer.

Guzmán Naranjo, Matías. 2019. *Analogical classification in formal grammar* Empirically Oriented Theoretical Morphology and Syntax. Language Science Press.

Kutuzov, Andrey & Elizaveta Kuzmenko. 2017. WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In Dmitry I. Ignatov, Mikhail Yu. Khachay, Valeri G. Labunets, Natalia Loukachevitch, Sergey I. Nikolenko, Alexander Panchenko, Andrey V. Savchenko & Konstantin Vorontsov (eds.), *Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, April 7-9, 2016, Revised Selected Papers*, 155–161. Cham: Springer International Publishing. doi:10.1007/978-3-319-52920-2_15.

Matthews, Clive A. 2005. French Gender Attribution on the Basis of Similarity: A Comparison Between AM and Connectionist Models. *Journal of Quantitative Linguistics* 12. 262–296.

Mortensen, David R, Siddharth Dalmia & Patrick Littell. 2018. Epitran: Precision G2P for many languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, .

Shannon, Claude. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27. 379–423, 623–56.

Skousen, Royal. 1992. *Analogy and structure*. Dordrecht: Springer.

Zaliznyak, Andrey Anatolyevich. 1977. *Grammatical dictionary of the Russian language*. Moscow: Russkij Jazyk.

Competing vowels in feminine formation:
Evidence from Hebrew and Jordanian Arabic

Lior Laks

Bar-Ilan University

1 Introduction

This study examines variation in the formation of feminine forms in Hebrew and Jordanian Arabic, based on three case studies. The study shows that variation (and lack thereof) can be predicted based on systematic guidelines. Such competing forms within a single morphological slot deviate from canonicity (Corbett 2005, 2007) and they pose a challenge for models that aim to explain why speakers select one form and not another. I argue that in the three cases, the competition in not just between rival suffixes, but between vowels. Specifically, I will show that the selection of a suffix with one vowel and not another is correlated with word formation in the lexicon and in the syntax.

2 Hebrew loanwords

Hebrew has three main feminine suffixes: *-a, -it, -et*. Suffix selection is predictable based on properties of the base, but is subject to irregularity (Schwarzwald 1984, 2002, Faust 2013, Asherov & Bat-El 2016). *-a* consists of the unmarked vowel *a* and can be attached to a variety of stems. *-it* is the productive default suffix in acronyms and loanwords. While most loanwords take only *-it*, there is a noticeable variation with some words. The loanword *snob* 'snob' takes *-it* (1a) or *-a* (1b), and both forms are found in similar contexts, while *larj* 'large (generous)' (2) takes only *-it*.

(1) a. *hem xošvim še-ani eyze **snob-it***
      'they think I am some snob'   (http://www.tapuz.co.il/blogs/viewentry/371153)
   b. *ve-hu xošev še-ani eyze **snob-a***
      'and he thinks I am some snob'
      (https://stips.co.il/ask/ %AA-%D7%90%D7%95-%D7%9E%D7%99-%D7%A9%D7%9E%D7%91%D7%99%D7%9F-%D7%AA%D7%A8%D7%90%D7%A9)
(2) *hi mamaš **larj-it** / ***larj-a** be-tšekim*
   'she is really large with checks'   (celebs.walla.co.il/item/2968806)

What triggers this variation and what blocks it?

Most loanwords that can take *-a* in addition to *-it*, are derogative, e.g. *debil* 'stupid', *farš* 'jerk' and *xnun* 'nerd'. Hebrew has feminine loanwords with no masculine base, and the majority of them have derogatory meaning, e.g. *pustema* 'blockhead' (**pustem*), *kunefa* 'ugly' (**kunef*). These words end with *-a*, which is not a suffix in such cases, and this makes this vowel more typical of derogatory words. I argue that as a result of the negative association of this vowel in loanwords, it is attached other loanwords with negative meaning in order to mark words as negative. While gender is considered inflection, which typically applies in the syntax (Perlmutter 1988, Anderson 1992, Corbett 2013, Štekauer 2015, Stump 2016, among others), I argue that marking the derogatory meaning applies in the lexicon because of the accessibility of *-a*. The tendency to select a suffix with *a* in order to lexically mark it with additional meaning is not surprising. In general, *a* has a morpho-lexical status in Hebrew. It is the most frequent vowel in word formation processes (Plada 1959, Bat-El 1994, Bolozky

1999, 2003, Schwarzwald 2002, 2012, Schwarzwald & Cohen-Gross 2000, Bolozky & Becker 2006). I claim that the use of *-a* in feminine formation of loanwords does not only play a grammatical role of gender marking, but it also labels these words as typically negative. It follows that the two suffixes *-it* and *-a*, do not completely overlap in their function. While *-it*, which is the default suffix of loanwords, has only a grammatical role of gender marking, *-a* also has a semantic role. The distinction between the two suffixes resembles, to some extent, the distinction between inherent and contextual inflection (Booij 1996, 2006). In general, inherent inflection is not required by the syntactic context, although it can be relevant to the syntax, e.g. number and gender for nouns. Contextual inflection is dictated by syntax, e.g. agreement markers for verbs and adjectives (Anderson 1992). I argue that the use of *-a* as a feminine marker for loanwords is "more inherent" than the use of *-it* in the sense that it provides a morphological marker of the negative meaning and does not only denote grammatical gender.

3 Hebrew passive participle patterns

Semitic morphology relies highly on non-concatenative morphology, where words are formed in patterns. Focus here is on the participle patterns *meCuCaC* and *muCCaC*, which have two main functions: (i) present tense forms of passive verbs in the *CuCaC* and *huCCaC* patterns respectively; (ii) adjectives (and a few nouns) (Berman 1978, 2017, Schwarzwald 2002, Ravid & Levie 2010, Doron 2013, among others). The same form can be ambiguous. For example, *mesudar* is both the present form of *sudaʁ* 'was tidied up', denoting 'is tidied up', and an adjective 'tidy'. Participle forms typically take the fem. suffix *-et*. However, a web search reveals cases where *-a* is also used. Such variation exists only with adjectives and not verbs. The participle *murkav*, for example, can denote both the adjective 'complicated' and the verb 'is composed'. The adjectival form takes both *-et* (3a) and *-a* (3b), while the verb form takes only *-et* (3c).

(3) a. *lelo safek zo šeela **murkev-et***
‘no   doubt   it’s   a   complicated   question’(http://www.mako.co.il/nexter-weekend/Article-2875f5d77c56b51006.htm)

   b. *zo šeela **murkav-a***
‘it’s a complicated question’ (https://www.fxp.co.il/showthread.php?t=16146268&page=13)

   c. *hi **murkev-et** al-yedey cevet mecumcam*
‘it is composed by a small team’ (https://cars.walla.co.il/item/2833011)

Morphology provides marking for different lexical categories in feminine formation. Similarly to the case of loanwords, it is not surprising that only adjectives can take *-a*. I assume that adjectives are derived and stored in the lexicon, while passive verbs are derived in the syntax (see Horvath & Siloni 2008). The morphological mechanism is sensitive to whether the masculine form 'comes' from the lexicon or is derived only in the syntax. It marks lexically derived forms with a vowel that is more typical to the lexicon.

4 Feminine formation on Jordanian Arabic

Arabic has one feminine suffix with two allomorphs: *-a* and *-e*. The suffix is attached to masculine nouns, adjectives and participle forms (*xabba:z* – *xabba:z-e* 'baker') and it could also be part of inanimate fem. nouns with no masc. base (*madin-e* 'city', **madi:n*) Some dialects

have only *-a* but in most dialects there is some degree of variation. Data in this study is based on a corpus of 25 narrative texts in Jordanian Arabic. The conditions for selection of either allomorph is primarily phonological. *-a* tends to surface after *r (mudi:r-a* 'manager') and low consonants like pharyngeals (*falla:ħ-a* 'farmer'), while *-e* surfaces elsewhere (*mʕallm-e* 'teacher') (Levin 1994, 2011, Shachmon 2011, Al-Wer & Horesh 2017, among others). In addition to the phonological conditions, the data shows that the distinction between the lexicon and the syntax is also relevant to the selection of either vowel. There is greater chance for *-e* to surface in adjectives. This is because feminine formation of adjectives, unlike that of nouns, is contextual inflection, which is more associated with the syntax. As in Hebrew, *-a* is more typical to Arabic word formation processes that apply in the lexicon (Holes 2004). As shown in (4), both the noun *ħufr-a* 'opening' and the adjective *zɣir-a* 'small' have *r* before the suffix and should take *-a*. This is indeed the case in (4a), but in (4b), only the adjective takes *-e* despite the phonological condition. The prediction is that both words can take *-a* or *-e,* but in cases where there is a distinction,   it is only the adjective that takes *–e (*ħufre zɣira* is supposed to be excluded).

(4) a.   *ħa:wal yeʕmel ħufr-a zɣi:r-a* (File-7)
          '(he) tried to make a small opening in the rock'

     b.   *ša:f ħufr-a zɣi:r-e* (File-10)
          '(he) saw a small opening'

Similarly, *-e* tends to surface in nouns that are heads of construct states. For example, *majmu:ʕ-a* 'group' typically surfaces with *a*, as it is preceded by *ʕ*, but when it is the head of a construct, there is some variation, e.g. *majmu:ʕ-at/majmu:ʕ-et kutta:b* 'a group of authors' . In such cases, the head is also marked with the consonant *t*, which indicates its syntactic position. Since this is a syntactic structure, there is tendency to also mark it with *-e*, which is more typical to the syntax. In contrast, when the word is not part of a construct state it tend to retain the vowel *-a*.

5 Conclusions

The three cases discussed above demonstrate competition between two fem. suffixes with different vowels: *a* vs. *e* or *i.* In each case, the selection of the less typical suffix (or allomorph) tends to take place systematically and occurs in order to mark an additional distinction: negative meaning and different lexical categories in Hebrew, and syntactic positions in Jordanian Arabic. These differences are all related to processes that typically apply in the lexicon or in the syntax and the morphological mechanism seems to be sensitive to such distinctions. The study sheds light on the role of gender with respect to morphological change. It adds to previous accounts of morphological variation and change, which result in doublet formation and competition (Thornton 2011, 2012, Aronoff 2016, 2017, among others).

References

Al-Wer, E. & U. Horesh. 2017. Dialect contact and change in an Arabic morpheme: Examining Jordanian and Palestinian dialects. A paper presented at the *North American Research Network in Historical Sociolinguistics (NARNiHS)*.
Anderson, S.R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, M. 2016. Competition and the lexicon. In A. Elia, C. Iacobino & M. Voghera, Livelli di Analisi e fenomeni di interfaccia. Atti del XLVII congresso internazionale della società di linguistica Italiana. Roma: Bulzoni Editore. 39-52.

Aronoff, M. 2017 (submitted). Competitors and alternants. To appear in a volume of *selected papers from IMM 17*, Vienna, 2016.

Asherov, D. & O. Bat-El. 2016. Multiple defaults: Hebrew feminine *–et* and *–a* in the present tense. *Morphology* 26. 399-423.

Bat-El, O. 1994. The optimal acronym word in Hebrew. *Proceedings of the 1994 Annual Conference of the Canadian Linguistic Association*. In Koskinen (ed.), *Toronto Working Papers in Linguistics*. 23-37.

Berman, R.A. 1978. *Modern Hebrew Structure*. Tel-Aviv: University Publishing Projects.

Berman, R. 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* edited by In Valentina Vapnarsky and Edy Veneziano. Amsterdam: John Benjamins.

Bolozky, S. 1994. On the formation of diminutives in Modern Hebrew morphology. *Hebrew Studies* 35, 47–63.

Bolozky, S. 1999. On the Special Status of the Vowels *a* and *e* in Israeli Hebrew. *Hebrew Studies* 40, 233–250.

Bolozky, S. 2003. Phonological and morphological variation in spoken Hebrew. In B. H. Hary (ed.), *Corpus Linguistics and Modern Hebrew: Towards the Compilation of the Corpus of Spoken Hebrew (CoSIH)*, 19–156. Tel Aviv: Tel Aviv University Press.

Bolozky, S. & M. Becker. 2006. Living lexicon of Hebrew nouns. Ms., University of Massachusetts Amherst. Available at http://becker.phonologist.org/LLHN.

Booij, G. 1996. Inherent vs. contextual inflection and the split morphology hypothesis. In G. Booij & J. van Marle (eds.), *Yearbook of Morphology 1995*, 1-16. Dordrecht: Kluwer.

Booij, G. 2006. Inflection and derivation. In: Brown, Keith (Eds.) *Encyclopedia of Language & Linguistics*. New York: Elsevier. 654–661.

Corbett, G.G. 2005. The canonical approach in typology. In Z. Frajzyngier, A. Hodges and D.S. Rood (eds.), *Linguistic Diversity and Language Theories*, 25–49. Amsterdam / Philadelphia: John Benjamins.

Corbett, G.G. 2007. Canonical typology, suppletion, and possible words. *Language* 83.1, 8–41.

Corbett, G.G. 2013. *The Expression of Gender*. De Gruyter Mouton

Doron, E. 2013. Participle. In G. Khan (ed.), *The Encyclopedia of Hebrew Language and Linguistics*. Leiden: Brill.

Faust, Noam. 2013. Decomposing the Feminine suffixes of Modern Hebrew. *Morphology* 23, 4, 409-440.

Holes, Clive. 2004. *Modern Arabic: Structures, Functions and Varieties*. Washington D.C.: Georgetown University Press.

Horvath, J. & T. Siloni. 2008. Active lexicon: adjectival and verbal passives. In S. Armon-Lotem, G. Danon, and S. Rothstein (eds.), *Current Issues in Generative Hebrew Linguistics*. Amsterdam/ Philadelphia: John Benjamins. 105-134.

Levin, Arie. 1994. *A Grammar of the Arabic Dialect of Jerusalem*. Jerusalem: The Hebrew University Magnes Press. (in Hebrew).

Levin, A. 2011. 'Imala. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich & A. Zabroski (eds.), Encyclopedia of Arabic Language and Linguistics, Vol. 2. 311-315. Leiden: Brill.

Perlmutter, David. 1988. The split-morphology hypothesis: evidence from Yiddish. In M. Hammond & M. Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*, 79–99. Orlando: Academic Press.

Plada, N. 1959. Phoneme frequency in Hebrew. *Leshonenu* 23, 235-242. (in Hebrew)

Ravid, D. & R. Levie. 2010. Adjectives in the Development of Text Production: Lexical, Morphological and Syntactic Analyses. *First Language* 30, 27-55.

Schwarzwald, O.R. 1982. Feminine Formation in Modern Hebrew". *Hebrew Annual Review* 6. 153–178.

Schwarzwald, O.R. 2002. *Hebrew Morphology*. 12 Study units. Tel Aviv: The Open University. (In Hebrew)

Schwarzwald, Ora R. 2012. On Vowel Polarity in Hebrew. *Leshonenu* 74, 79-92. (in Hebrew)

Schwarzwald, O.R. & Dalia Cohen-Gross. 2000. The productive noun patterns in Hebrew. In M. Horvits (ed.), *The Language of Contemporary Press: Mina Efron's Memorial Volume*, 148–161. Tel-Aviv: The Mofet Institute. (in Hebrew)

Shachmon. O. 2011. *Pausal final imāla in central Palestinian dialects. Jerusalem Studies in Arabic and Islam (JSAI)* 38, 145-161.

Štekauer, P. 2015. The delimitation of derivation and inflection. In P.O., Müller, I. Ohnheiser, S. Olsen, F. Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe*. 218–234.

Stump, Gregory T. 2016. *Inflectional Paradigms: Content and Form at the Syntax-Morphology Interface*. Cambridge University Press.

Thornton, A.M. 2011. Overabundancy (multiple cells realizing the same cell: a non-canonical phenomenon on Italian verb morphology. In M. Maiden, J.C. Smith, M. Goldblach & M. Hinzeli (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*. Oxford: Oxford University Press. 358-381.

Thornton, A.M. 2012 .Reduction and maintenance of overabundance: A case study on Italian verb paradigms. *Word Structure* 5, 183-207.

Non-linear morphology in Romance: the case of vowel length in Friulian verbs

Nicola Lampitelli   *Paolo Roseano*   *Francesc Torres-Tamarit*
University of Tours   University of Barcelona   CNRS

1   The facts

Friulian is a Rhaeto-Romance (Western, Gallo-Romance) language spoken by approximately 300,000 speakers (source: Ethnologue) in the Friuli region in North-Eastern Italy. Three main dialects exist in the Friulian linguistic area, each of which exhibits specific morpho-phonological features (Roseano 2015a).

Table 1: NE paradigms — PI

	Conj 1 'swim'	Conj 2 'drink'	Conj 3 'understand'
1s	ˈnaːdi	ˈbeːf	kaˈpiʃ
2s	ˈnadas	ˈbeːfs	kaˈpis
3s	ˈnada	ˈbeːf	kaˈpiʃ
1p	naˈdiŋ	beˈviŋ	kaˈpiŋ
2p	naˈdais̭	beˈveis̭	kaˈpiːs
3p	ˈnad-iŋ~aŋ	ˈbeviŋ	kaˈpiʃiŋ

The Northern Friulian dialect spoken in Negrons (NE) displays an intriguing case of vowel length occurring exclusively in conj. 1 verbs. More precisely, vowel length occurs in the 1s of the Present Indicative (PI) (table 1) and in the 1-3s and 3p of the Present Subjunctive (PS) (table 2), as reported in Roseano (2015b). Tables 3 and 4, in turn, show data from Central Friulian as spoken in Çupicje (ÇU), in which no lengthening ever occurs in conj. 1 verbs. New data were collected by the authors in December 2018. 4 speakers were interviewed and recorded in NE, and 5 speakers in ÇU. The survey consisted of a translation task: a question was asked in Italian to trigger an answer in Friulian in which the inflected form of the verb appeared in final position.

In Friulian nouns and verbs, vowel length contrasts occur only in stressed, utterance-final position (e.g. [ˈbruːt] 'broth' vs. [ˈbrut] 'ugly'; [ˈpaːs] 'peace' vs. [ˈpas] 'step'; [ˈfiːs] 'sons' vs. [ˈfis] 'fixed, dense') (Torres-Tamarit 2015). This kind of vowel lengthening is predictable from the underlying laryngeal specification of the following obstruent despite final obstruent devoicing (Vanelli 1979, Finco 2009), which turns the phonological generalization opaque (e.g. [ˈkruːt] 'raw-m-sg' vs. [ˈkrude] 'raw-f-sg'; [ˈfuːk]

Table 2: NE paradigms — PS

	Conj 1	Conj 2	Conj 3
1s	ˈnaːdi	ˈbevi	kaˈpiʃi
2s	ˈnaːdis	ˈbevis	kaˈpiʃis
3s	ˈnaːdi	ˈbevi	kaˈpiʃi
1p	naˈdiŋ	beˈviŋ	kaˈpiŋ
2p	naˈdais̭	beˈveis̭	kaˈpiːs
3p	ˈnaːdiŋ	ˈbeviŋ	kaˈpiʃiŋ

'fire' vs. [fuˈɡut] 'fire-dim'; [ˈnaːs] 'nose' vs. [naˈzut] 'nose-dim'). The same process of vowel lengthening occurs in those verbal forms that are consonant-final and have final stress in PI in both NE and ÇU (see tables 1 and 3, conj. 2 and 3). The vowel length in 2p of conj. 3 verbs is the surface outcome of /i+i/. Interestingly, the source of vowel length observed in NE conj. 1 verbs, however, is of a different kind: it does not appear in final position, but in penultimate position, and is not conditioned by the laryngeal specification of the following consonant.

2   The analysis

Table 3: ÇU paradigms — PI

	Conj 1	Conj 2	Conj 3
1s	ˈnodi	ˈbeːf	kaˈpis
2s	ˈnodis	ˈbɛvis	kaˈpisis
3s	ˈnode	ˈbeːf	kaˈpis
1p	noˈdiŋ	beˈviŋ	kaˈpiŋ
2p	noˈda̰ḭs	beˈvḛḭs	kaˈpiːs
3p	ˈnodiŋ	ˈbɛviŋ	kaˈpisiŋ

At first sight, root allomorphy could be advocated. NE conj. 1 verbs could be lexically associated with two allomorphs (e.g. /ˈnaːd/ and /ˈnad/), and each allomorph would be inserted in a specific morphosyntactic environment. This path raises a question we cannot answer to, namely: why are only conj. 1 verbs specified as root-alternating verbs? The second possibility is to analyze lengthening as a T(ense)/M(ood) morpheme: lengthening occurs in PS. If so, however, why does it also occur in 1s PI and, again, only in conj. 1 verbs? The third possibility, the one we pursue here, is the following: vowel lengthening expresses conj. 1 Th(eme). In other words, conj. 1 Th can spell out as a melodically-empty mora, a prosodic morpheme that is realized as length on the stressed root vowel.

We claim that the conj. 1 Th morpheme in NE has the following lexically listed allomorphs: /ˈa, ˈi, a, i, μ/. Only the vowels subcategorize for specific forms (e.g. stressed vowel allomorphs subcategorize for 1p-2p forms). In the absence of subcategorization, then the default allomorph, /μ/, with any subcategorization frame, is selected. As illustrated in table (5), there is only lengthening when the T/M morpheme is not realized as a vowel.

Table 4: ÇU paradigms — PS

	Conj 1	Conj 2	Conj 3
1s	ˈnodi	ˈbɛvi	kaˈpisi
2s	ˈnodis	ˈbɛvis	kaˈpisis
3s	ˈnodi	ˈbɛvi	kaˈpisi
1p	noˈdini	beˈvini	kaˈpini
2p	noˈdadis	beˈvedis	kaˈpidis
3p	ˈnodiŋ	ˈbɛviŋ	kaˈpisiŋ

Table 5: NE: analysis of PS

	root	Th	T/M	φ-Fs
1s	ˈnad	μ	i	Ø
2s	ˈnad	μ	i	s
3s	ˈnad	μ	i	Ø
1p	nad	ˈi	Ø	ŋ
2p	nad	ˈa	Ø	is
3p	ˈnad	μ	i	ŋ

The fact that Th vowels and lengthening stand in complementary distribution supports the hypothesis that lengthening is in fact one of the possible phonological realizations of the Th morpheme. According to this analysis, the T/M morpheme in PS is realized as /i/, or zero when the Th morpheme is realized as a stressed vowel. In PI, only 1s exhibits vowel length. As illustrated in table (6), 1s is the only form in which Th is not spelled out as a vowel.

This morphological segmentation allows for a uniform analysis of the T/M morpheme as a null morpheme across the PI paradigm in all conjugations. The terminal element /i/ in 1s is analyzed as a φ-feature morph, and crucially cannot express Th.

This is independently supported by (i) /i/ also appears in 1s Imperfect Indicative after the Th vowel /ˈa/ (e.g. [nad-ˈa-v-i]), so it is reasonable not to interpret this /i/ in 1s PI as a Th vowel but as the φ-feature for 1s across some conj. 1 tenses; and (ii) the /i/ in 1s PI is different from the /i/ in 3p PI in that only the latter stands in free variation with /a/ (e.g. [ˈnadiŋ] or [ˈnadaŋ]), the typical Th vowel for conj. 1.

This variation indirectly suggests that /i/ is only

Table 6: NE: analysis of PI

	root	Th	T/M	φ-Fs
1s	ˈnad	μ	Ø	i
2s	ˈnad	a	Ø	s
3s	ˈnad	a	Ø	Ø
1p	nad	ˈi	Ø	ŋ
2p	nad	ˈa	Ø	is
3p	ˈnad	i∼a	Ø	ŋ

Th in 3p PI but not in 1s PI. Note that the /i/ in 3p PI is different from the /i/ in 3p PS. Only in PS /i/ expresses uniformly the T/M morpheme in 1-3s and 3p. This is the reason why there is only vowel length in 3p PS, where /i/ is the T/M morph, but not in 3p PI, where /i/ (or /a/) is a Th vowel (table 7). Dialect ÇU, in turn, never displays vowel length in conj. 1 verbs because Th in this dialect simply does not include any prosodic allomorph.

3 OT formalization

We propose a formal analysis of morphological length in Friulian couched within OT that makes use of internally layered ternary feet (Martínez-Paricio 2013), general well-formedness markedness constraints on the size of syllables and two morphology-phonology interface constraints on compliance with lexical subcategorization frames

Table 7: Analysis of 2p PI vs. PS

	root	Th	**T/M PI**	ϕ-Fs
3p	'nad	i∼a	$\emptyset$	ŋ
	root	Th	**T/M PS**	ϕ-Fs
3p	'nad	μ	i	ŋ

and the realization of morphs, respectively. Below we define the set of constrains used in the analysis:

(1) Respect: Respect idiosyncratic lexical specifications (Bonet et al. 2007). = It enforces compliance with lexical subcategorization requirements.

(2) All-Feet$_{Max}$-Right: Every maximal foot must be right-aligned (no intervening σ between Ft$_{Max}$ and ω). = It prohibits long vowels in stressed antepenultimate position.

(3) *SuperHeavy: Superheavy, trimoraic syllables are banned. = It prohibits long vowels in closed syllables.

(4) Max-Morph: Morphs cannot be deleted. = It enforces morpheme realization.

(5) All-Feet$_{Max}$-Right ≫ Max-Morph: /μ/ is not realized if stress is antepenultimate (e.g. ['canti] 'I sing' cf. *[caːnti]).

(6) *SuperHeavy ≫ Max-Morph: /μ/ is not realized if the stressed syllable of the root is closed (e.g. ['liberi] 'I free' cf. *['liːberi]).

(7) The undominated position of Respect ensures absolute compliance with subcategorization requirements.

4 Conclusion

The present analysis of morphological length in Friulian shows that there is no need for an L-shaped morpheme analysis of the data (Maiden 2004). In our analysis, each morph, including length, spells out a morphosyntactic feature.

5 Selected references

Bonet, Eulàlia, Maria-Rosa Lloret & Joan Mascaró (2007). Allomorph selection and lexical preferences: two case studies. *Lingua* 117: 903–927.

Finco, Franco (2009). Fonetiche e fonologjie dal furlan centrâl. *Gjornâl furlan des siencis* 11: 53–85.

Maiden, Martin (2004). Morphological autonomy and diachrony. *Yearbook of Morphology 2004*:

137–175.

Martínez-Paricio, Violeta (2013). An exploration of minimal and maximal metrical feet. PhD dissertation, University of Tromsø-Arctic University of Norway.

Roseano, Paolo. (2015a). Dialetti. In Sabine Heinemann and Luca Melchior (eds), Manuale di linguistica friulana. Berlin/Boston: De Gruyter.

Roseano, Paolo. (2015b). Morfologia non-lineare in Romanzo: la flessione verbale del friulano gortano. Archivo glottologico italiano C(I): 85–110.

Torres-Tamarit, Francesc. (2015). Length and voicing in Friulian and Milanese: or why rule-free derivations are needed. Natural Language & Linguistic Theory 33: 1351–1386.

Vanelli, Laura (1979). L'allungamento delle vocali in friulano. *Ce fastu?* 55: 66–76

The Status of Affixes and
the New Words by *-ment* in Present-Day English

Ryohei Naya

University of Tsukuba

1 Introduction: The Status of Affixes

Contemporary discourse on the status of derivational affixes is fraught with controversy; are they functional or lexical morphemes? Creemers et al. (2018) summarized recent studies and point out that some scholars have argued that all derivational affixes are functional morphemes while others have argued that they are lexical morphemes (or roots). Another argument has emerged between these two views. Creemers et al. (2018) argue that some affixes are functional and others are lexical. These arguments assume that a given affix is exclusively functional or lexical. In contrast, Emonds' (2000) syntactic theory of morphology, which hypothesizes that the lexicon consists of Syntacticon and Dictionary, states that a single derivational affix may behave as a functional morpheme in some cases and a lexical morpheme in others.

This study aims to examine the status of derivational affixes based on newly found empirical data of *-ment* retrieved from the *Oxford English Dictionary Online* (*OED*). More precisely, this study argues that among the several views on the status of affixes, Emonds' (2000) hypothesis is the most promising. Section 2 presents the results of my *OED*-based survey and details the relevant data. The data are analyzed in Section 3 on the basis of Emonds' (2000) hypothesis. Section 4 summarizes the findings and outlines a conclusion.

2 Problematic Behaviors of *-ment* in Present-Day English

The suffix *-ment* is often described as unproductive in present-day English (PDE) but is still available (Marchand 1969: §4.65.1; Bauer et al. 2013: §10.2.1.1). However, previous studies have not sufficiently analyzed its use and creative aspects in PDE; thus, *-ment* in PDE can be expected to provide a new perspective on the discussion of the status of suffixes.

Using the *OED*'s advanced search function, I retrieved 23 *-ment* nouns that were first recorded after 1900. These nouns can be classified according to their origins of the base words and their syntactic categories, as shown in Table 1. Each example is followed by the date of the first attestation in the *OED*. Nouns in the square and those in the circle in Table 1 are of the greatest interest to this study.

	Romance Base		Non-Romance Base	
Verbal Base	excystment (1928) piercement (1925) perturbment (1901)	bemusement (1907) reforestment (1921) superencipherment (1940)	embrittlement (1921) encirclement (1919) upliftment (1926) underlayment (1956)	weldment (1945) soothment (1900) staggerment (1933)
		centrement (1975) munitionment (1915)	motherment (1914) stakement (1904)	— Converted Words + *-ment*
Non-verbal Base	foolishment (1918)	curettement (1908) encoppicement (1935) memberment (1904) scholarment (1922)	bashment₂ (1996)	
	A + *-ment*		N + *-ment*	

Table 1. Classification of *-ment* Nouns Recorded after 1900

2.1 Problem 1: Categorial Selection

Functional morphemes are generally sensitive to the categorial properties of their selected elements, as shown in the alignments C-T-*v*-V and D-N. In fact, *-ment* primarily selects verbs and changes them into nouns. However, *-ment* in the examples in (1), which are also presented in the square in Table 1, departs from this derivative pattern in that it attaches to adjectives and nouns.

  (1) a. A + *-ment*: *foolishment* 'foolishness'

       b. N + *-ment*: *bashment₂, curettement, encoppicement, memberment, scholarment*

Previous studies have reported that *-ment* can attach to certain nouns (e.g., *devilment, illusionment*) and adjectives (e.g., *merriment, embeddedment, insensiblement*) (Bauer et al. 2013: 198); however, these cases are few and may be regarded as exceptions. The situation is different in PDE; the nouns in (1) cannot be dismissed as trivial because they account for 26% of new *-ment* nouns. Moreover, given that the derivative pattern of *-ment* "seems to have been stabilized after 1450" (Marchand 1969: 331), *-ment* should strictly follow this pattern in PDE. However, it can still attach to adjectives and nouns. Accordingly, the examples in (1) problematize the argument that *-ment* is a functional morpheme.

2.2 Problem 2: Myers' Generalization

The suffix *-ment* attaches to converted (or zero-derived) nouns and adjectives in (2), which are presented in the circle in Table 1.

  (2) a. Denominal Verb + *-ment*: *centrement, motherment, munitionment, stakement*

       b. Deverbal Noun + *-ment*: *bashment₂*

In *motherment*, 'motherly care or supervision, mothering,' for example, the noun *mother* undergoes N-to-V conversion, and the resultant verb is combined with *-ment*, as in (3).

  (3)   *mother*ₙ > N-to-V Conversion > *mother*ᵥ > *-ment* suffixation > *mother-ment*

This type of word-formation process conflicts with Myers' Generalization, which states that "no derivational suffix may be added to a zero-derived word" (Myers 1984: 66). For example, adding *-ant* to denominal converted verbs is not allowed as in *[[[experiment]ₙ]ᵥ -ant]ₙ (see Nagano 2008: 17). Nagano (2008: 16–18) points out that Myer's Generalization has three exceptional suffixes: *-er, -ing,* and *-able,* which are known as highly productive suffixes (e.g., *documenter, documenting, documentable*). However, *-ment* is not considered an exception. Thus, if *-ment* is a derivational suffix, the nouns in (2) are problematic to this generalization.

3 Analysis

The two problems described above seem unrelated, but they follow one assumption that is naturally derived from Emonds' (2000) hypothesis.

3.1 Theoretical Background

Emonds (2000) hypothesizes that affixes exhibit the dual nature in principle; not only can they behave as functional morphemes, but they can be used in the same manner as lexical morphemes as well. This hypothesis successfully captures the fact that suffixes such as *-ation* and *-ment* can form two types of deverbal nominals known as complex event nominals (CENs) and result nominals (RNs) in Grimshaw's (1990) terms. CENs function similarly to verbs in that they inherit argument structures of base verbs as illustrated in (4a), and RNs can be regarded as genuine nouns in that they prototypically refer to physical objects and can be pluralized as shown in (4b).

(4) a. *The assignment of that problem too early in the course always causes problems.*
    b. *The assignments were too long.*

(Grimshaw 1990: 54; underlining mine)

In Emonds' (2000: §4.6) analysis, *-ment* in (4a) is a functional morpheme, whereas in (4b), it behaves in the same manner as nouns. He hypothesizes that functional morphemes—not lexical ones—undergo lexical insertion after syntactic computation and that prior to lexical insertion, functional morphemes are inert (Emonds 2000: 115). In the underlined expression in (4a), the nominal suffix *-ment* is inactive, and the verb *assign* is substantially a head in syntactic computation. Consequently, *assign* can introduce the argument *that problem*. In this case, *-ment* simply transforms a verb into a noun as a functional morpheme. In (4b), *-ment* undergoes lexical insertion at the beginning of syntactic derivation along with lexical morphemes. Accordingly, *-ment* is active as a nominal head throughout the derivation; thus, *assignment* behaves as a genuine noun.

Naya (2016) clarifies the categorial status of *-ment* in (4b), which is inserted at the beginning of syntactic derivation. He argues that the suffix as used in RNs is a lexical morpheme meaning 'thing / entity.' The noun *assignment* in (4b) thus means 'thing that is assigned.' In this example, the verb *assign* modifies the head *-ment*. If *-ment* in (4b) is a lexical morpheme, *assignment* in (4b) is formed by combining two lexical morphemes. This process is equivalent to compounding.

Emonds' (2000) hypothesis indicates the possibility that a single affix can behave as a functional morpheme in some aspects and a lexical morpheme in others. The next subsection demonstrates that this hypothesis is useful in understanding the behaviors of *-ment* in PDE.

3.2 Proposal

I propose that *-ment* in (1) and (2) is a lexical—rather than functional—morpheme. The nouns in (1) and (2) are formed through root compounding, and they have the same status as the RN in (4b). This proposal naturally accounts for the unconventional behaviors of *-ment* in PDE. First, *-ment* attaches to other than verbs in (1). Unlike derivation, compounding does not impose categorial restrictions on its input elements. For example, the noun *man* can be combined with a verb (e.g., *wash man*), noun (e.g., *sandwich man*), and adjective (e.g., *merry man*). Thus, if *-ment* is a lexical morpheme, it can participate in compounding and attach to any word, including nouns and adjectives. Second, *-ment* attaches to converted words in (2). Recall that Myers' Generalization is concerned with *derivational* suffixes. If *-ment* is used as a *lexical* morpheme, it is not subject to this restriction, and it can attach to converted words. Thus, the behaviors of *-ment* in PDE are explicable.

Note that I do not argue that *-ment* always behaves as a lexical morpheme in PDE. As shown in (4a), *-ment* can also behave as a functional morpheme bearing a category-changing function from V to N. That is, it has a dual nature. This property can be naturally captured under Emonds' (2000) view on the status of affixes without extra assumptions. If we assume that *-ment* is always a functional morpheme, its ability to attach to nouns, adjectives, and converted words is not easily explainable. If we consider that *-ment* is exclusively a lexical morpheme, we overlook its category-changing function. The data presented in this study must be investigated with a hypothesis that allows the suffix flexibility as that proposed in Emonds (2000).

3.3 Supporting Evidence

If the nouns in (1) and (2) are compounds, the nouns containing verbal bases should behave in the same way as RNs. Unlike CENs, RNs cannot co-occur with the arguments of the verbs

used therein. For example, the RN *assignment(s)* in (4b) cannot be accompanied by *the problems*, which corresponds to the object of the verb *assign*:

(5) * *The assignments of the problems took a long time.*   (Grimshaw 1990: 54)

This property can be observed in V-N combinations like *tax man*, which cannot co-occur with the argument of *tax*, as in **a tax man of hidden assets* (cf. *to tax hidden assets*) (Roeper 1987: 268). If the proposed analysis is correct, we can predict that the nouns in (2) that contain a verbal base are RNs and cannot co-occur with the objects of the verbs. This prediction is borne out as shown in (6) (I excluded *stakement* because this is a historical technical term).

(6) a. *The computer's {*centrement / centering} of all sentences in the paper took a long time.*
     b. *Her constant {*motherment / mothering} of my children was very helpful.*
     c. *The defense industry's {*munitionment / munitioning} of the forces took a long time.*

For example, *centrement* in (6a) is not compatible with *all sentences in the paper*, which corresponds to the object of the verb *centre* (cf. *to centre all sentences in the paper*). The ungrammaticality of the *-ment* nouns in (6) indicates that the nouns are RNs formed by root compounding, which supports the proposal that *-ment* is a lexical morpheme in (6). In contrast to *-ment,* the highly productive suffix *-ing,* an exception to Meyers' Generalization, can successfully derive impeccable CENs from the relevant verbs. That is, the verbs can transform into CENs through derivation by their nature. This fact also shows that the ungrammaticality in (6) should be attributed to the nature of *-ment* as a lexical morpheme.

4 Summary

Emonds' (2000) approach to affixes differs from other approaches in that his hypothesis allows a single affix to behave as both a functional and lexical morpheme. This hypothesis is useful in analyzing the behaviors of *-ment* in PDE. While the suffix is a purely functional in category changing, it behaves non-canonically in new PDE words: it can attach to adjectives, nouns, and converted words. This property is not surprising if we assume that *-ment* can function as a lexical morpheme and be used in compound formation. Emonds' (2000) view successfully explains the otherwise unexpected behaviors of *-ment,* which demonstrates the dual status of an affix.

References

Bauer, L., Lieber, R. & Plag, I. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Creemers, A., Don, J. & Fenger, P. 2018. Some affixes are roots, others are heads. *Natural Language & Linguistic Theory* 36, 45–84.

Emonds, J. E. 2000. *Lexicon and grammar*. Berlin: Mouton de Gruyter.

Grimshaw, J. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Marchand, H. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach,* 2nd ed. Munich: C. H. Beck.

Myers, S. 1984. Zero-derivation and inflection. In Speas, M., & Sproat, R. (eds.), *MIT Working Papers in Linguistics 7,* 53–69. Department of Linguistics and Philosophy, MIT.

Nagano, A. 2008. *Conversion and back-formation in English*. Tokyo: Kaitakusha.

Naya, R. 2016. Deverbal noun-forming processes in English. *English Linguistics* 33, 36–68.

Roeper, T. 1987. Implicit Arguments and the head-complement relation. *Linguistic Inquiry* 18, 267–310.

Uninflecting and uninflectable lexemes: implications for paradigm structure

Andrew Spencer
University of Essex

1   Uninflecting and uninflectable lexemes

In inflecting languages some (classes of) lexemes are **uninflecting**: *of, the, almost, …*, but some members of otherwise inflecting classes may (unexpectedly) fail to inflect, i.e. they show **uninflectability**. Russian nouns distinguish 6 x 2 = 12 case/number forms but about 3000 are indeclinable (uninflectable): all the forms of their paradigm are identical to the root, e.g. *kenguru* 'kangaroo'. We distinguish such uninflectable lexemes from uninflecting lexemes on the one hand and defective lexemes (e.g. Russian *mečta* 'dream', lacking gen.pl) on the other. Uninflectable lexemes can occur in all the contexts open to inflecting lexemes (unlike defective lexemes), but the form is invariable, e.g. Russian (where UNINFL indicates and uninflecting or uninflectable form):

(1) a. odin   vombat/kenguru
         one.NOM.M.SG wombat[M].NOM.SG/kangaroo[UNINFL]
         'one wombat/kangaroo'

     b. s   ètimi   vombat-**ami**/kenguru
         with these.INSTR.PL wombat[M].INSTR.PL/ kangaroo[UNINFL]
         'with these wombats/kangaroos'

Lexemes can also be partially (un)inflectable: Polish *muzeum* inflects in the plural but is indeclinable in the singular; verbs of the English *hit* class have only *-s* and *-ing* inflections; Macedonian adjectives such as *kasmetlija* 'lucky, sg', *kasmetlii* 'lucky, pl' fail to inflect for [GENDER:{m,f,n}] in the singular.

2   Typology of uninflectedness

Russian KENGURU is uninflectable in all its occurrences — lexical uninflectability. However, English KANGAROO is inflectable except in compounds: *kangaroo(*s) tails* . This is *constructional uninflectability*, in which an otherwise inflecting lexeme has to be realized as a non-inflecting word form, or even as an uninflected bound stem, as in many cases of noun incorporation. A further case of constructional uninflectability is the predicative adjective in German:

(2) a. Ich bin ein **kleines**   Känguru
         I   am a   little.N.SG.NOM/ACC kangaroo
         'I am a little kangaroo'

     b. das Känguru ist **klein**(*-e/*-es/…)
         the kangaroo is little[UNINFL]
         'The kangaroo is little'

A particularly interesting case is that of indeclinable foreign names in Russian (though not necessarily in other Slav languages). Many borrowed or cited foreign words, and especially names,

have a phonological shape which is difficult to accommodate to the Russian morphological system, like *kenguru* (no native nouns have a stem ending in *-u*) or *Si (Czin′pin)* 'Xí (Jìnpíng)'. Other foreign names, however, can easily be inflected as though they were Russian: *Bil* 'Bill', *Bil-a* 'genitive singular' (cf *Kiril, Kiril-a*), or *Klinton* 'Clinton', *Klinton-a* 'gen. sg.' (cf *Solženicyn, Solženicyn-a*). However, female referent names do not have phonological forms such as *Klinton*, and native given names do not end in *-i* for either sex, so that *Xilari Klinton* 'Hillary Clinton' is indeclinable, like *Si* 'Xí':

(3) a. reč′   Xilari   Klinton
         speech Hillary[UNINFL] Clinton[UNINFL]
         'Hillary Clinton's speech'

     b. reč′   Bil-a   Klinton-a
         speech Bill[M]-GEN.SG Clinton[M]-GEN.SG
         'Bill Clinton's speech'

     c. reč′   Si   Czin′pin-a
         speech Xí[M] Jìnpíng[M]-GEN.SG
         'Xí Jìnpíng's speech'

where *Czin′pin*, but not *Si*, inflects. Cases such as these seem to be intermediate between lexical and constructional uninflectability.

3   Theoretical questions

I address two of the several questions that a theoretical model of morphology has to answer in order to accommodate uninflectability, and link these to Question 3.

1. How can an uninflected form of a lexeme be treated by the morphosyntax as though it were fully inflected (and what prevents **all** lexemes from behaving in this way)?

2. In cases of constructional uninflectability with inflectable lexemes, such as the German predicative adjective construction, what is the morphosyntactic description of the uninflected lexical form? How is that form accessed by the grammar?

3. In the PFM2 class of models, what is the interface form between the entry for an uninflecting lexeme and the syntactic terminal it occupies, i.e. what 'word form' of a lexeme such as ALMOST undergoes 'lexical insertion'?

4   A PFM2-class approach

With paradigm-driven approaches, discussion of the morphology-syntax interface seldom asks how a completely uninflecting lexeme, such as an English prepositions, is represented in the syntax. Specifically, how do we interpret the notion 'word form of lexeme $\mathscr{L}$' where $\mathscr{L}$ belongs to a class which has no inflectional paradigm, in a model in which the morphology-syntax interface is supposed to be mediated through the inflectional paradigm, as in PFM2? Sag (2012, 119) is unusual in recognizing and addressing this problem. In SBCG, uninflecting lexemes are shifted to the type *word (form)* by a Zero Inflection Construction. This effectively treats such words (even particles, *oh, y'know, …*) as trivially inflected forms of lexemes with one-celled paradigms. While Sag's proposal permits uninflecting words to enter the syntax it does not directly address the problem of uninflectable lexemes, whether lexically uninflectable or constructionally.

PFM2 distinguishes crucially between CONTENT and FORM paradigms (Π_C, Π_F). Π_C defines all the syntactically accessible inflectional contrasts a lexeme is obliged to make, Π_F defines the morphophonological forms expressing those contrasts, a **_Correspondence_** function **_Corr_** specifies the mapping $\Pi_C \mapsto \Pi_F$, in part defined by the function **_pm_** defined over the feature sets, Σ, T of Π_C, Π_F. In addition, we make explicit the obvious assumption that every lexeme has a lexical (lexemic, dictionary) entry specifying a basic morphophonological form, or 'root' (i.e. a default lexical stem, STEM_0). For a completely uninflecting lexeme (class) we take $\Sigma = \text{T} (= \Pi_C = \Pi_F) = \varnothing$. This applies to, say, English prepositions or to uninflecting coverbs in languages with extensive LVCs. For uninflectable, non-defective members of an otherwise inflecting lexical class we take $\Sigma (= \Pi_C) \neq \varnothing$, T $(= \Pi_F) = \varnothing$, e.g. *kenguru* lexemes.

We now extend the definition of the **_Corr_** function minimally. Consider the application of **_Corr_** to a lexeme with lexemic index/LID λ, that is associated with no CONTENT paradigm features, Σ, at all (i.e. an uninflecting lexeme, say, English ALMOST). Stump's definition of **_Corr_** maps the Π_C cells to corresponding Π_F cells via the Paradigm Function, $\text{PF}(\langle Z, \tau \rangle)$, which maps the root of the lexeme, Z, and the form feature set, τ, to a cell in Π_F, (4).

(4)   Given λ the lexical index of any lexeme/lexemic entry, $\Lambda =$ the complete set of lexemic entries ($=$ the lexicon).
Then $\forall \sigma \in \Sigma$ (possibly null), $\forall \lambda \in \Lambda$, **_Corr_**$(\lambda, \sigma) = \text{PF}(\langle Z, \textbf{\textit{pm}}(\sigma) \rangle)$

However, this mapping will be undefined as things stand for cases where $\Sigma = \text{T} = \varnothing$. We therefore assume mapping (5), the Default Exponence Principle, as the default morphosyntactic expression of all lexemes defined over null feature sets.

(5)   Default Exponence Principle (DEP)
$\text{PF}(\langle Z, \varnothing \rangle) = Z = \text{STEM}_0(\lambda)$

Together, (4, 5) state that the default realization of all lexemes is STEM_0 ($\equiv$ Z, the lexical root). For inflecting lexemes the DEP (5) is overridden by the more specific **_Corr(espondence)_** function. However, for an uninflecting lexeme DEP has the same effect as Sag's Zero Inflection Construction. This means that we do not require uninflectable lexemes to have a non-null inflectional paradigm. The uninflected lexemes whose form is defined by (4, 5) have just a root form but no paradigm, CONTENT or FORM.

For Russian uninflecting KENGURU, KLINTON[F], SI etc, Π_C defines 12 cells, as expected. However, we assume a special application of **_pm_** triggered by a lexical class feature under which the Π_F is undefined (for partially uninflectable lexemes the relevant part of Π_F is undefined). This will mean that the **_Corr_** function will deliver the value $\text{PF}(\langle Z, \varnothing \rangle)$. By (5) this means that the uninflectable lexeme reverts to its root form for all or part of its CONTENT paradigm.

Note that this account correctly characterizes *kenguru*-words as (highly) irregular: the otherwise default mapping to FORM paradigm is overridden and undefined. This has consequences for entropy-based perspectives on paradigm structure. Entropy can only be defined over cells in FORM paradigms (CONTENT paradigms are maximally regular by definition). But for uninflectable lexemes we seem to minimize entropy, giving the false impression that such lexemes are highly regular. On my approach, entropy measures are simply undefined for uninflecting/uninflectable lexemes.

We have accounted for non-inflecting lexemes with exactly the same machinery, and shown how ordinary inflecting words are the result of a specific application of the **_Corr_** function overriding the DEP, answering Question (1) above. At the same time we have begun to address Question (2): the uninflected forms of lexemes that appear in compounds or German predicative adjective constructions are also root forms. The most direct way to handle this is to assume

that such a construction includes an interface specification overriding the normal morphology-syntax interface under which lexical insertion is defined over cells in the Π_C. Instead, such a construction has to be so formulated that it specifies the Π_C of any lexeme corresponding to the appropriate syntactic terminal as 'locally undefined' with respect to that construction. How this is to be implemented formally depends on precisely how the morphology-syntax interface is formalized.[1]

The approach makes the prediction that there can be no defective lexemes such as Russian МЕЧТА. A defective lexeme is one which unexpectedly has an unfillable, undefined Π_C cell, and a fortiori a corresponding gap in Π_F. However, given the DEP, such cells should be treated like (the unique form of) an uninflecting lexeme, and so should revert to the root form, contrary to fact. This means that additional machinery will have to be developed to actively prevent defective forms from getting realized. Given the very peculiar ontological status of defective lexemes/cells, this seems to be the correct conclusion.

References

Sag, Ivan A. 2012. Sign-Based Construction Grammar: an informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 69–202. Stanford, CA: CSLI Publications.

---

[1]Polish *muzeum*-words have to be given special treatment: their uninflectable singular form, *muzeum*, is not presumably the STEM_0 form. Such cases appear to be rather rare.

The influence of 'absence' on Sanskrit morphology – the case of negative compounds

Małgorzata Sulich-Cowley

University of Warsaw, Faculty of Oriental Studies, Chair of South Asian Studies

1 Compounds and negation – general remarks

Sanskrit linguistic tradition distinguishes various types of compounds with different relations between their constituent elements: *avyayībhāva* (indeclinable compounds), *tatpuruṣa* (where the second member is treated as primary), *dvandva* (with two, or more, constituents of equal status) and *bahuvrīhi* (compounds that refer to an outside entity). One of the most important factors which helps us to determine the final meaning of the formation is the primacy assigned to one of the members in a compound. In my presentation I would like to discuss negative compounds in Sanskrit, which are an extremely interesting phenomenon from the perspective of linguistic, as well as philosophical, description as they could be classified in different categories, which would, consequently, influence their morphological behaviour.

There are two important factors to be taken into account while determining the semantic value of a negative compound. Firstly, it is the actual meaning of a negative particle (naÑ). According to the earliest known Indian etymological treatise *Nirukta* (Yāska, c. 5th century BCE), this particle can express both 'absence' as well as 'comparison' (later grammatical tradition makes the semantic domain of negation more detailed). Based on the examples, I would like to show that both these meanings can actually be present at the same time. Secondly, it is the potential meaning carrier properties of particles in general. Sanskrit linguistic and philosophical schools considered parts of speech either independently expressive (*vācaka*), with their own meaning that they bring into the final meaning of a formation, or merely indicative (*dyotaka*), which meant that they helped to reveal the meaning that was inherently present in the word they accompanied. This latter issue bears particular significance for the formation of compounds and the establishment of qualifier-qualified relation between its members as it is closely connected with an issue of the subject of negation.

In my presentation I will refer to the formal system of Sanskrit grammar as well as the related philosophical description of the Indian school of philosophers-grammarians. I will demonstrate the influence of the semantics component on the morphological behaviour of a negative compound, both as an independent formation and as a part of a sentence, where those mutual interactions can be most visibly seen.

2 The analysis of negative compounds

The negative particle naÑ, which Sanskrit grammarians adopt as representing the main notion of negation, stands for both the prefix *a/an* in compounds as well as to the verbal negation *na*. The first Indian grammatical treatise, *Aṣṭādhyāyī*, composed by Pāṇini (4th century BCE) introduces this particle in the section describing *tatpuruṣa* compounds such as *abrāhmaṇa* ('a non-Brahmin'), *adṛṣṭa* ('not seen') or *aneka* ('not one, many'). Pāṇini's approach to negative compounds, therefore, seems to be non-controversial; in such

compounds meaning predominance (*pradhānya*) is awarded to the second member, which is preceded by a qualifier. From the structural point of view, the form *abrāhmaṇa* is identical to *nīlakamala* ('a blue lotus') and as such, the entire compound adopts the gender and number of its last member.

Grammarians and philosophers that followed Pāṇini were not convinced as to the simplicity of such formations. Starting with Patañjali (c. 2nd century BCE) and his *Mahābhaṣhya*, a commentary on the *Aṣṭādhyāyī*, various possibilities of classification were discussed. The particular character of negative formations, which lies in denying the existence of the accompanying word, gives rise to questions regarding the predominant meaning in negative compounds. The options that were given most attention were the primacy of the second member of a compound, resulting in a *tatpuruṣa*, and the initial one with the result of an *avyayībhāva*. This discussion was prompted by the fact that the particle naÑ is, naturally, classified in Pāṇinian system as an indeclinable. This discussion is not purely academic but it holds serious morphological consequences. Morphological behaviour of these constructions is different, both during derivational process as well as at the final outcome. In a *tatpuruṣa* compound both (or more) elements are – at least theoretically – subject to declension and the addition of case endings which are subsequently deleted; despite being a particle, naÑ is also treated in the same way during the process of forming a compound. The gender and number of the final formation is determined based on the qualities of the second member of a compound. If an *avyayībhāva* compound is formed, however, the particle – being an indeclinable – is not subject to declension and, consequently, the entire formation adopts a fixed nominal ending and neuter gender. It would then also serve as an adverb in a sentence. These varying approaches prompted questions regarding the semantic value of the negative particle (as mentioned earlier) and its influence on constructing negative compounds, and as a consequence, their morphological and syntactic behaviour. It has been believed that semantics does not play a major role in the *Aṣṭādhyāyī*; it seems, however, that this role has been underestimated and Pāṇini does acknowledge the importance of the semantic component in forming a meaningful utterance. The thought that it is the whole sentence that is a proper meaning carrier was explored by all the commentators of Pāṇini's work and heavily developed by the later emerged school of philosophers-grammarians. This thought is believed to have originated with the *Aṣṭādhyāyī* itself. This is the reason why the analysis of the semantic value of the particle naÑ was crucial for the understanding of the formation, which influences the meaning of a sentence it was a part of. The discussion regarding the extent to which the meaning of a final construction, such as a compound, is composed of the meanings of its individual parts was heated in Sanskrit linguistics.

3 Research questions

In my presentation I would like to discuss the morphological implications of negative compounds in Sanskrit being assigned to different categories based on the analysis of exerpts from grammatical and philosophical treatises such as the *Mahābhāṣya* of Patañjali (2nd century BCE) or the *Vākyapadīya* by Bhartṛhari (5th century CE). I will show the problems we might face on either morphological or semantic grounds when we analyse compounds such as *aneka* or *abrāhmaṇa*. I would also like to consider whether it is possible for a compound to be assigned two categories, namely if it is possible for a negative formation to be a *tatpuruṣa* and *bahuvrīhi* at the same time. This final option (of a negative formation being a *bahuvrīhi*) was briefly discussed in Sanskrit grammatical literature, and quickly discarded, based on morphological grounds. The exocentric characteristic of *bahuvrīhi* compounds is manifested

in the gender and number they adopt, that is one of the external referent. This very characteristic was the reason for its rejection with reference to negative compounds, which in principle accept the gender and number of the second member of a compound. However, I do believe that certain aspects of *bahuvrīhi* formations should be taken into account while talking about negative compounds. I will show that, due to their semantic uniqueness, they cannot be clearly classified and that the semantic component has to play a crucial role in the compounding process itself. A significant number of negative formations, despite expressing a form of 'absence', are, in fact, positive in their denotation. The meaning 'comparison' (mentioned in 1.1), which the particle naÑ was supposed to express, comes in handy when determining this positive denotation. This is where the domains of *tatpuruṣa* and *bahuvrīhi* meet and this is why their classification can be cross-categorical.

Bibliography
Primary sources

Aṣṭādhyāyī of Pāṇini. Various editions.

Kāśikākavṛtti of Jayāditya-Vāmana along with Commentaries Vivarana-pañcika - Nyasa of Jinendrabuddhi and Padamañjari of Haradatta Mishra. 1985. Śrīnārāyaṇa Miśra (ed.). Varanasi: Ratna Publications.

Mahābhāṣyapradīpavyākhyanāni. 1983. M. S. Narasimhacharya (ed.). Pondichéry: Institut Français d'Indologie.

The Nighaṇṭu and The Nirukta, The Oldest Indian Treatise on Etymology, Philology and Semantics. 1998. Sarup, Lakshman (ed. and tr.). New Delhi: Motilal Banarsidass.

The *Vākyapadīya* of Bhartṛhari with the vṛtti of the Paddhati of Vṛṣabhadeva 1995. Iyer, K. A. Subramania (ed.). Pune.

The *Vākyapadīya* of Bhartṛhari, chapter III, pt. ii. 1974. Iyer, K. A. Subramania (ed. and tr.). New Delhi: Motilal Banarsidass.

The *Vyākaraṇamahābhāṣya* of Patañjali. different editions.

Secondary sources

Bilimoria, Purushottama. 2008. *Śabdapramāṇa = Word and Knowledge as Testimony in Indian Philosophy*. New Delhi: DK Printworld.

Bhate Saroja and Bronkhorst Johannes (eds.). *Bhartrhari - philosopher and grammarian*. New Delhi: Motilal Banarsidass.

Bronkhorst, Johannes. 1983. On the history of Pāṇinian grammar in the early centuries following Patañjali. *Journal of Indian Philosophy* 11. 357-412.

Coward, Harold G. 1980. *The sphota theory of language, A philosophical Analysis*. New Delhi: Motilal Banarsidass.

Mazumdar, Pradip Kumar. 1994. *Negative words, their ways in language. A socio-semantic approach*, Sanskrit. Calcutta: Pustak Bhandar.

Patnaik, Tandra. 2007. *Śabda: a study of Bhartṛhari's philosophy of language*. New Delhi: DK Printworld.

Rath, Gayatri. 2000. *Linguistic philosophy in Vakyapadiya*. Delhi: Bharatiya Vidya Prakashan.

Vergiani, Vincenzo. 1993. The negative *asamarthasamāsas* in the Pāṇinian tradition. *Rivista degli Studi Orientali* 67. 65-81.

Composés N-N et N-A dans la littérature française du 17ᵉ au 20ᵉ siècle : la productivité morphologique

Elena Voskovskaia

Grande Prairie Regional College, Canada

1 Introduction

La productivité des mots composés est un domaine de recherche très peu exploré, surtout en français (Krott, Schreider & Baayen 1999), (Fernandez Dominguez 2007, 2009). La mesure quantitative la plus couramment utilisée est basée sur la notion de l'hapax (Baayen & Lieber 1991), (Baayen 1992). Toutefois, Hay (2003) montre que c'est la fréquence relative plutôt que la fréquence absolue qui a un effet sur la décomposition et la productivité des mots complexes. Ce papier vise à étudier la productivité morphologique des noms composés N-N et N-A dans la littérature française du 17ᵉ au 20ᵉ siècle. Nous examinons une corrélation possible entre la productivité morphologique et la fréquence relative de ces deux types de composés.

Ce travail vise à vérifier une hypothèse suivante : Il existe une corrélation inverse entre la productivité et la fréquence relative des composés N-N et N-A : les composés dont la fréquence relative est plus basse sont plus productifs que ceux dont la fréquence relative est plus élevée. Nous cherchons à vérifier si la notion de la fréquence relative proposée par Hay (2003) en dérivation peut être appliquée aux mots composés.

Mots-clés : Morphologie, productivité morphologique, corpus textuel, diachronie.

2 Corpus et méthodes

La recherche est basée sur le corpus textuel *Frantext* (plus de 170 millions de mots) divisé en quatre périodes selon leur importance dans l'histoire de la langue française : 1606-1694 (17,3 millions de mots) ; 1695-1798 (34,4 millions de mots) ; 1799-1872 (41 millions de mots) et 1873-1920 (28 millions de mots). La liste de 72 composés N-N (*croix-pile, volte-face*) et 39 composés N-A (*arc-boutant, acquit-patent, bec-cornu*) a été créée à partir du Dictionnaire de Littré (1877).

Deux mesures sont utilisées pour évaluer la productivité des composés. La première mesure utilisée pour évaluer le niveau de productivité est la mesure P au sens strict (Baayen & Lieber 1991), (Baayen 1992) calculée comme $P = n_1/N$ (où P = le taux de productivité ; n_1 = le nombre d'hapax legomenon ; N = le nombre total d'occurrences). Selon cette mesure, une catégorie avec un grand nombre de mots de haute fréquence a une grande valeur de N et un degré de productivité P moins élevé.

La deuxième mesure appliquée est celle de la fréquence relative FR (Hay 2003) élaborée pour la composition comme *f*relative = *f*composé/*f*base. Selon cette mesure, les formes dont la fréquence relative est basse sont plus productives.

3 Résultats

Notre étude a démontré que parmi 72 composés N-N, 65 composés (90%) sont moins fréquents que leurs bases, ce qui permet de les envisager comme décomposables et potentiellement productifs. La corrélation entre les mesures P et FR dans les composés N-N est présentée dans le Tableau 1.

Tableau 1 Corrélation entre la productivité et la fréquence relative dans les composés N-N

mesure	composés N-N			
	1606-1694	1695-1798	1799-1872	1873-1920
FR_1	0.0349	0.0455	0.0427	0.0296
FR_2	0.2501	0.2268	0.1506	0.1855
FR_3	0.0099	0.0188	0.0129	0.0118
FR_4	0.0199	0.0376	0.0257	0.0218
P	0.0770	0.0100	0.0014	0.0012

La corrélation inverse entre les mesures FR_3 et FR_4 et la productivité P a été observée pour 73 % des composés N-N. Le niveau de productivité le plus élevé (P1 = 0.0770) correspond à la fréquence relative la plus basse si on considère la mesure FR_3 = 0.0099 pendant la période 1606-1694 et 1799-1872 (P3 = 0.0014 vs FR_3 = 0.0129). Quant à la mesure FR_2, elle ne présente aucune corrélation inverse. Toutefois, les chiffres du Tableau 1 illustrent que la corrélation n'est pas parfaite : le patron inverse apparaît seulement dans les années 1799-1872 (FR_3 = 0.0427 vs P3 = 0.0014).

L'analyse du type N-A a été basée sur 39 composés dont 35 composés (90 %) sont moins fréquents que leurs bases indépendamment de la variante FR utilisée. Ainsi, ils peuvent être envisagés comme décomposables et potentiellement productifs. La corrélation entre les mesures P et FR dans les composés N-A est résumée dans le Tableau 2.

Tableau 2 Corrélation entre la productivité et la fréquence relative dans les composés N-A

mesure	composés N-A			
	1606-1694	1695-1798	1799-1872	1873-1920
FR_1	0.0435	0.1469	0.1894	0.2238
FR_2	0.2138	0.2939	0.3111	0.3366
FR_3	0.0119	0.0141	0.0345	0.0241
FR_4	0.0238	0.0283	0.0689	0.0481
P	0.0348	0.0051	0.0012	-

Les données du Tableau 2 révèlent que le patron de la corrélation inverse fonctionne parfaitement pour les composés N-A si on considère la mesure FR_3. Le taux de productivité le plus élevé (P1 = 0.0348) correspond à la fréquence relative la plus basse si on considère la mesure FR_3 = 0.0119.

3 Conclusion

Nous avons avancé l'hypothèse qu'il existe d'une corrélation inverse entre la productivité et la fréquence relative dans les composés français N-N et N-A. L'étude a révélé une faible présence d'une corrélation inverse entre les mesures FR et P : 34% pour les composés N-N et 50% pour les composés N-A.

La mesure de la fréquence relative peut varier considérablement en fonction du choix de la base du composé. Globalement, les mesures FR_3 et FR_4 correspondent mieux à la corrélation attendue, où la fréquence relative la plus basse correspond au niveau de productivité la plus élevée. Étant donné que la mesure FR_4 est un dérivé de la mesure FR_3, la mesure FR_3 est envisagée comme la meilleure variante pour évaluer la fréquence relative en composition.

Ainsi, à l'opposé de ce qu'on observe dans la dérivation, les données obtenues indiquent qu'en composition, la corrélation inverse entre la fréquence relative et la productivité n'est pas explicitement présentée. Le critère FR$_3$ sert plutôt à établir la présence de la productivité (i.e. la décomposition du composé), mais il ne démontre pas son degré.

Les résultats de cette étude démontrent la nécessité de recherches supplémentaires sur les différents types de composés concernant une corrélation possible entre la productivité morphologique et la fréquence relative.

Références

Baayen, H., Lieber, R. 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29. 801-843.

Baayen, Harald 1992. Quantitative aspects of morphological productivity. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of morphology 1992*, 109-149. Dordrecht: Kluwer.

Baayen, Harald 1993. On frequency, transparency and productivity. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of morphology 1993*, 181-208. Dordrecht: Kluwer.

Bauer, L. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Dal, Georgette 2003. Productivité morphologique : définitions et notions connexes. Dans *Langue française*, v.140, 3-23. Larousse : Paris.

Fernandez-Dominguez, J., Diaz-Negrillo, A. & Štekauer, P. 2007. How is Low Morphological Productivity Measured? *Atlantis* 29.1. 29-54. Revista de la Asociación Española de Estudios Anglo-Norteamericanos.

Fernández-Domínguez, J. 2009. *Productivity in English Word-formation: An approach to N+N compounding*. European University Studies: Peter Lang Publishing.

Hay, Jennifer & Baayen Harald. 2002. Affix Productivity and Base Productivity. Paper presented at the Morphological Productivity Seminar, ESSE 6, Aug 30-Sept 3, Strasbourg.

Hay, J. 2003. *Causes and consequences of word structure*. New York: Routledge.

Krott, A., Schreuder, R. & Baayen, R. H. 1999. Complex words in complex words. *Linguistics* 37-5. 905-926.

Index