Booklet of abstracts

17th International Morphology Meeting

18-21 February 2016, Vienna

Organizing Committee: Franz Rainer, Francesco Gardani, Elisabeth Peters

Program Committee: Franz Rainer, Wolfgang U. Dressler, Francesco Gardani, Hans Christian Luschützky

Scientific Committee: Farrell Ackerman, Willem Adelaar, Kimi Akita, Adam Albright, Peter Arkadiev, Giorgio Francesco Arcodia, Mark Aronoff, Harald Baayen, Matthew Baerman, Eric Bakovic, Hans Basbøll, Laurie Bauer, Melanie Bell, Thomas Berg, Jim Blevins, Jonathan Bobaljik, Olivier Bonami, Kersti Börjars, Dunstan Brown, Sebastian Bücking, Chiara Cappellaro, Chiara Celata, Jeroen Claes, Greville Corbett, Denis Creissels, Wolfgang Dressler, Hans-Olav Enger, Pattie Epps, David Erschler, Bernard Fradin, Nanna Fuhrhop, Livio Gaeta, Christina Gagné, Francesco Gardani, Hélène Giraudo, Stefan Th. Gries, Martin Haspelmath, Alice Harris, Holden Härtl, Bernd Heine, Jarich Hoekstra, Martin Hummel, Claudio Iacobini, Sharon Inkelas, Brian Joseph, Steven Kaye, István Kenesei, Krista Kerge, Emmanuel Keuleers, Ferenc Kiefer, Seppo Kittilä, Katharina Korecky-Kröll, Lívia Körtvélyessy, Johanna Laakso, Gary Libben, Rochelle Lieber, Michele Loporcaro, Ana Luís, Hans Christian Luschützky, Martin Maiden, Andrej Malchukov, Stela Manova, Jörg Meibauer, Chiara Melloni, Marianne Mithun, Karlheinz Mörth, Sylvia Moosmüller, Peter O. Müller, Ingeborg Ohnheiser, Jerome Packard, John Payne, Jacob B. Phillips, Vito Pirrelli, Ingo Plag, Frans Plank, Bernhard Pöll, Franz Rainer, Angela Ralli, Tom Rankin, Christian Rathmann, Dorit Ravid, Davide Ricca, Martine Robbeets, Elke Ronneberger-Sibold, Chris Schaner-Wolles, Barbara Schlücker, Wolfgang Schulze, Thomas Schwaiger, Peter Sells, Andrew Spencer, Pavol Štekauer, Thomas Stolz, Gregory Stump, Renata Szczepaniak, Pius Ten Hacken, Anna M. Thornton, Jochen Trommer, Natsuko Tsujimura, Freek Van de Velde, Maria Voeikova, Andrew Wedel, Martina Werner, Björn Wiemer, Ronnie Wilbur, Amir Zeldes, Inge Zwitserlood.

February 2016
Vienna

The Program Committee

Table of Contents

Invited talks

Talks and Posters

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

Poster - Distribution and productivity of changing past participles in Frisian
Anne Merkuur, Arjen Versloot and Willem Visser

Meaning in word-formation: The frame-semantic hypothesis .
Ingo Plag, Lea Rebecca Kawaletz and Marios Andreou

On borrowing compounds
Angela Ralli

Semantic competition in Hebrew morphological families: A developmental analysis
Dorit Ravid, Orit Ashkenazi, Levie Ronit, Galit Ben Zadok, Tehila Grunwald, Ron Bratslavsky, Efrat Raz and Gillis Steven

The puzzling case of overabundance in Italian comparatives
Enzo Santilli

Issues of competition in evaluative morphology: A corpus-based study of Modern Greek intensifying affixes and compounding components
Paraskevi Savvidou

The memorization of complex constructions from a cross-linguistic perspective
Marcel Schlechtweg and Holden Härtl

Singular nouns looking like plurals cause more agreement attraction than genuine plurals
Natalia Slioussar

Gender agreement attraction in Russian: Different profiles in production and comprehension
Natalia Slioussar and Anton Malko

Multiple affixation in Portuguese: Structural restrictions and processing conditions
Alexandra Soares Rodrigues

Morphomic splits within periphrasis: mixed paradigms in Italo-Romance
Pavel Štichauer

Polysemy and event nominals competition in Italian
Rossella Varvara and Roberto Zamparelli

Lexical competition within the derivational and inflectional paradigm: Some evidence from masked priming with French stimuli
Madeleine Voga and Helene Giraudo

Competition in the German plural system: A product-oriented perspective
Verena Wecker

Chinese adjective-noun combinations in a framework combining Construction Morphology with Realization Optimality Theory
Zheng Xu

Morphological status and acoustic realization: Is there a difference between Bra[d] Pitt and a grille[d] cheese omelet, or between Kate Mo[s] and killer robot[s]?
Julia Zimmermann

Workshops

Discriminative perspectives on morphology
Jim Blevins

Computational methods for descriptive and theoretical morphology
Olivier Bonami and Benoît Sagot

Lexicalization patterns in colour naming: a cross-linguistic perspective
Ida Raffaelli

Morphology and the brain: Online studies of morphological processing
Phaedra Royle and Karsten Steinhauer

Mark Aronoff

Competition and variation

Twentieth-century linguistics was dominated by the search for invariance. Its greatest achievements, the phoneme and the transformation (at least in its earliest incarnations) were tools designed to reduce distributional variants to a more basic entity. Allophones were distributional variants of phonemes, while the earliest transformations served to derive 'equivalent' expressions from their more basic forms, for example, the Passive and the Active forms of a verb from a more abstract underlying representation.

A much older strand, often traced to the Hindu grammarians, treats distribution as competition between forms, in Max Mueller's (1870) words "the struggle for life among words and grammatical forms which is constantly going on in each language." Mueller saw this struggle as the linguistic analogue of natural selection. I will apply this idea of a struggle between linguistic elements to a number of well-known types of cases that have often been thought of in terms of variation: synonymous words, competing affixes, and constraint ordering in both phonology and morphology. In many of these types, competition (framed in terms of Gause's (1934) ecological principle of competitive exclusion), provides a more enlightening account than more standard treatments. It also allows us to see language not solely as an idiosyncratic human device but as a system governed at least in part by much more general properties of systems.

References

Gause, G. F. 1934. *The struggle for existence*. Baltimore: Williams and Wilkins.
Mueller, Max. 1870. Darwinism tested by the Science of Language. *Nature* 1(10). 256–259.

Michael Ramscar

Learning, competition and the nature of morphology

For most of the last century, the study of language has largely assumed an atomistic model in which linguistic signals comprise discrete, minimal form elements which are in turn associated with a discrete, minimal elements of meaning. Accordingly, production has been seen to involve the composition of messages from an inventory of form (i.e., morphological) elements, and comprehension the subsequent decomposition of these messages. Research in linguistics has thus tended to focus on identifying and classifying these elements, and on attempting to formulate lossless processes of composition and decomposition (Bloomfield, 1933; Matthews, 1991). This program has raised as many questions as answers, especially when it comes to specifying the nature of form - meaning associations (Blevins, 2016).

By contrast, across the same period behavioral and neuroscience research based on human and animal models has revealed that "associative learning" is a discriminative process (Ramscar, Dye, & McCauley, 2013). Learners acquire predictive understandings of their environments through competitive mechanisms that tune systems of internal cue representations to eliminate or reduce any uncertainty they promote. Critically, models of this process better fit empirical data when these cue representations do not map discretely onto the aspects of the environment learners come to discriminate (Ramscar & Port, 2015). Seen from this perspective, languages are probabilistic communication systems (Shannon, 1948; Ramscar & Baayen, 2013) that exhibit continuous variation within a multidimensional space of form-meaning contrasts. Discrete descriptions of these systems at either an individual (psychological) or community (linguistic) level are thus necessarily idealizations. Since idealizations inevitably lose information, the different types of idealizations explored in different atomistic models of morphology over the past century can be seen to differ mainly in terms of the kinds of information that they lose.

References

Bloomfield, Leonard (1933). *Language*. University of Chicago Press, Chicago.

Blevins, James P. (2016). *Word and Paradigm Morphology*. Oxford University Press, Oxford.

Matthews, Peter H. (1991). *Morphology*. Cambridge University Press, Cambridge.

Ramscar , Michael, Melody Dye, and Stewart McCauley. (2013) Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language*, 89(4), 760-793.

Ramscar, Michael and R. Harald Baayen. (2013) Production, comprehension and synthesis: A communicative perspective on language. *Frontiers in Language Science*s. 4:233. doi: 10.3389/fpsyg.2013.00233

Ramscar, Michael and Robert Port. (2015). Categorization (without categories). In E. Dawbroska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics*. De Gruyter Mouton.

Shannon, Claude (1948). *A mathematical theory of communication*. The Bell System Technical Journal, 27, 379-423, 623-656

Anna M. Thornton

Overabundance: A Canonical Typology

I define overabundance (henceforth OA) as the situation in which two (or more) inflectional forms are available to realize the same cell in an inflectional paradigm (Thornton 2011).
The term *overabundance* is my English rendering of the Italian term *sovrabbondante*, current in descriptive grammars of Italian throughout the 20th century, and based on Latin *abundantia*, a term used in descriptive grammars of Latin and Greek at least since the early 19th century, to cover several kinds of nouns presenting doublets in their paradigm.
In my talk, building on previous work by Corbett (2005, 2006, 2007a, 2007b) on non-canonical phenomena in morphology, on my previous work on OA (Thornton 2011, 2012a, 2012b, 2012c, 2013) and on work by Bond (2013) and Brown and Chumakina (2013) on Canonical Typology, I will present some elements of a Canonical Typology of Overabundance.
The definition of OA given above is the base of a canonical typology; several criteria of canonicity will be presented and discussed, and the paper will conclude with a definition of canonical overabundance.
Data from Italian and other well-described languages will be discussed.

References

Bond, Oliver. 2013. A base for canonical negation. In Dunstan Brown, Marina Chumakina & Greville G. Corbett (eds.), *Canonical morphology and syntax*, 20-47. Oxford: Oxford University Press.

Brown, Dunstan & Marina Chumakina. 2013. What there is and what there might be: An introduction to Canonical Typology. In Dunstan Brown, Marina Chumakina & Greville G. Corbett (eds.), *Canonical morphology and syntax*, 1-19. Oxford: Oxford University Press.

Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges, and David S. Rood (eds.), *Linguistic diversity and language theories* (Studies in language companion series 72), 25-49. Amsterdam: John Benjamins.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2007a. Canonical typology, suppletion, and possible words. *Language* 83.1, 8-41.

Corbett, Greville G. 2007b. Deponency, Syncretism, and What Lies Between. *Proceedings of the British Academy* 145, 21-43.

Thornton, Anna M. 2011. Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology. In Maiden, Martin, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*, 358-381. Oxford: Oxford University Press.

Thornton, Anna M. 2012a. Overabundance in Italian verb morphology and its interactions with other non-canonical phenomena. In Thomas Stolz, Hitomi Otsuka, Aina Urdze & Johan van der Auwera (eds.), *Irregularity in Morphology (and beyond)*, 251-269. Berlin: Akademie Verlag.

Thornton, Anna M. 2012b. La sovrabbondanza nei paradigmi verbali dell'italiano contemporaneo**.** In Patricia Bianchi, Nicola De Blasi, Chiara de Caprio & Francesco Montuori (eds.), *La variazione nell'italiano e nella sua storia. Varietà e varianti linguistiche e testuali,* Atti dell'XI Congresso SILFI, 445-456. Firenze: Franco Cesati Editore.

Thornton, Anna M. 2012c. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5.183-207.

Thornton, Anna M. 2010-11 [published 2013]. La non canonicità del tipo it. *braccio // braccia / bracci:* Sovrabbondanza, difettività o iperdifferenziazione? *Studi di grammatica italiana* 29-30, [2010-2011, published in 2013], 419-477.

Blake Allen and Gunnar Ólafur Hansson

Multiple bases and empirical priors in paradigm inference: Experimental evidence from Icelandic

Theoretical Context: The task native speakers face when predicting unknown inflected forms of a lexeme—the Paradigm Cell Filling Problem (Ackerman et al. 2009)—is widely acknowledged to require a complex form of inference, and models of this inference often include simplifying hypotheses intended to render the problem tractable. We consider two such hypotheses. The Single Surface Base Hypothesis (Albright 2002) holds that when performing inflectional inference, speakers can rely only on the properties of one single privileged base form. A second hypothesis is that speakers adhere strictly to the system of inflectional classes, and will not produce a form that results in an unattested arrangement of exponents across the paradigm. Based on evidence from an experimental study of Icelandic noun inflection, we reject both hypotheses as untenable. We instead propose a simple, powerful Bayesian model of inflectional inference, the predictions of which are consistent with the experimental findings.

Background: In Icelandic nouns, most case-number combinations have various suffixal exponents. Individual cells differ greatly in their predictiveness; e.g. DatPl takes *-um* for nearly all nouns and is thus uninformative. In descriptive works, GenSg and NomPl are often treated as "principal parts" (along with NomSg), from which other forms can determined. We focus on one aspect of these predictive relations: the four AccPl exponents shown in Table 1 and how these can be inferred from combinations of certain GenSg/NomPl exponents. For each GenSg/NomPl combination, one can predict with near-perfect accuracy the indicated AccPl exponents.

	NomPl: *-ar*	NomPl: *-ir*
GenSg: *-s*	*-a*	*-i*
GenSg: *-ar*	*-ar*	*-ir*

Table 1: Accusative Plural exponents of nouns, as predicted by GenSg and NomPl

Meanwhile, AccPl exponents differ greatly in their frequency across the lexicon. This is partly due to some inflectional classes dwarfing others, but also due to other classes not shown here sharing the same AccPl exponents. For example, AccPl *-a* also characterizes the extremely productive class defined by GenSg *-a*, NomPl *-ar*.

Experiment: We conducted an experiment with 122 Icelandic speakers, using an online *wug test* paradigm (Berko 1958), but manipulating the information provided about other inflectional forms. A set of nonce nouns was created and distributed across the four inflectional classes shown in Table 1. Using a Latin square design, each noun was presented (in syntactic context) in anywhere from one to three forms: {DatPl}, {DatPl, GenSg}, {DatPl, NomPl}, or {DatPl, GenSg, NomPl}. Subjects then filled in a test sentence by selecting their preferred AccPl form from the four options in Table 1.

Mixed-effects logistic regression analysis reveals that subjects combine information from all inflected forms known to them, counter to the Single Surface Base Hypothesis. A model with (knowledge of) both GenSg and NomPl as predictors fits the responses significantly better than

models with only one of the two. Meanwhile, even in the {DatPl, GenSg, NomPl} presentation condition subjects fell far short of the near-100% accuracy that is in principle attainable given the situation in Table 1. Rather, responses even in this maximally informative condition were strongly biased toward the relative lexical frequencies of the four AccPl exponents—even though, as a result, lexemes often came to exhibit novel arrangements of exponents. This suggests that inference between paradigm cells is constrained by such frequencies, independent of the system of attested inflectional classes.

Proposal: Having rejected both hypotheses under scrutiny, we instead propose a Bayesian model of inflectional inference. This states that the (predicted) probability of a lexeme selecting a certain exponent in some paradigm cell depends not only on the probability of its known forms (its "bases" in other paradigm cells) given that exponent, but also on the prior probability of that exponent itself: $p(exponent|bases) \propto p(bases|exponent)p(exponent)$. The relative influence of the known bases versus the prior probabilities can be modulated by implementing regularization on $p(bases|exponent)$, as proposed, e.g., for maximum entropy models (Hayes and Wilson 2008). Our model accurately predicts the complex patterns observed in the experimental data while introducing no mechanisms not already motivated in the statistical learning or linguistics literature.

References

Ackerman, Farrell, James P. Blevins, Robert Malouf, and Juliette Blevins. 2009. Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In *Analogy in grammar: Form and acquisition*, ed. by James P. Blevins and Juliette Blevins, 54–82. Oxford: Oxford University Press.

Albright, Adam. 2002. The identification of bases in morphological paradigms. Doctoral Dissertation, University of California, Los Angeles.

Berko, Jean. 1958. The child's learning of English morphology. Doctoral Dissertation, Radcliffe College.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Rusudan Asatiani

An Optimal Cognitive "Decision": The Generative Model of Formal Representations of an Active~Passive Continuum in Georgian

Two different formal models traditionally called 'Active' and 'Passive' ones are clearly distinguished in Georgian (see: Shanidze 1973).

Active transitive constructions and their conversive passives, as a rule, are grammaticalized correspondingly by the active (1a) and passive (1b) formal models; e.g.:

(1) (a) *k'ac-**i*** *xat'-**av-s*** *surat-**s***
man-NOM paint-THM-PRS.ACT.S.3.SG picture-DAT
'The man paints the picture.'

(b) *surat-**i*** ***i**-xat'-**eb-a*** *k'ac-is* *mier*
picture-NOM PASS-paint-THM-PRS.PASS.S.3.SG man-GEN by
'The picture is painted by the man.'

But there are a lot of cases where simple semantic-functional interpretation of the models cannot be given inasmuch as the constructions pointed out as 'Active' or 'Passive' can actually represent a variety of verb semantics: non-conversive passives (as dynamic so static), active (intransitive/dynamic), reciprocals, potentials, deponents, and etc. (see: Asatiani 2013). Thus, the problem with the paradigms is that it is difficult to predict the meaning from the form and, to such an extent, traditional terms 'Active' and 'Passive' actually have a conventional character.

Attempts at new theoretical approaches have been undertaken to explain such cross-linguistically known phenomenon. M. Shibatani's (Shibatani 2006) interpretation seems to be more efficient from this point of view. He considers the active-passive opposition as a continuum, where polar dimensions fit in with the prototypical active and passive constructions, while non-polar, inter-medial cases share only some semantic-categorical features of the categories characteristic for the prototypical ones.

Languages turn to various strategies to represent such non-polar, medial cases; they either create new formal models, or come to an optimal decision and choose from the existing ones a model which is conventionally regarded as the most appropriate and proximate according to the certain semantic-categorical features.

During the formal representation of medial forms Georgian creates the new model showing the auxiliary conjugation for Static verbs, while for other medial forms it chooses either the active or the passive formal model.

The paper offers a generative strategy based on the decisive semantic features defining the choice: 'DINAMICITY', 'AORIST' and 'TELICITY'. The latter was used by Dee Ann Holisky (Holisky 1981) for medio-active verbs, although we suppose it is decisive for the whole process of formal representations of an active-passive continuum.

The character of 'VERB ARGUMENTS' is also important. The arguments are distinguished by the feature 'FREE WILL': an argument whose 'FW' (cmp.: Harris 1981) is included in the event is pointed out as +[FW] (resp. Ag) while an argument whose 'FW' is not included in the event is pointed out as −[FW] (resp. Ad). What about the argument which is semantically 'undergoer' and does not exist independently of the event, or not at all (Dowty

1991) (resp. Patient), the feature FW seems to be redundant for it; that is, it might be structurally qualified as an argument with a priory *zero*-FW. Thus, it is pointed out as Ø.

A choice is organized as an algorithm with the four stages of implicational rules and mirrors the hierarchically organised optimal generative process of linguistic structuring of an Active~Passive continuum in Georgian.

The certain combinations of rivalry features are formalized according to the following strategy: −[DYNAMIC] (resp. static) verbs follow the passive model without any restrictions (I stage), while for +[DYNAMIC] verbs it is necessary to check: 1. Do they equate with events including Ø-argument or do not? (II stage); 2. Do they represent events in Past (resp. +[AORIST]) or do not? (III stage) 3. Do they represent an action or event as being complete in some sense (resp. +[TELIC]) or do not (IV stage)?

The proposed model is the first attempt describing Georgian verb forms within the generative frame that has more explanatory validity. The whole process of formal models choices is governed by a general, optimal, cognitively defined "conventional linguistic decision": *the definite model representing some core semantics serves better to represent the certain marginal semantics as well.*

References

Asatiani, Rusudan. 2013. The information structure and typological peculiarities of the Georgian passive constructions. *Proceedings of 9th International Symposium on LLC,* 17-35. Berlin, Heidelberg: Springer-Verlag.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language 67.* 547-619.

Harris, Alice. 1981. *Georgian syntax.* Cambridge: Cambridge University Press.

Holisky, Dee Ann. 1981. *Aspect and Georgian medial verbs.* Delmar, N.Y.: Caravan Books

Shanidze, Akaki. 1973. *Fundamentals of the Georgian language grammar*. Tbilisi: TSU Press. (in Georgian).

Shibatani, Masayoshi. 2006. On the conceptual framework for voice phenomena. *Linguistics* 44-2. 217-269.

Matthew Baerman

Morphomic compositionality

As Stump (2001) points out, approaches to morphology can be either incremental (the meaning of a form is built up out of smaller units of meaning) or realizational (the meaning of a form is a property of the whole form). Since incremental approaches are more constrained in what they can model, much of the discussion within contemporary morphological theory concerns whether or not a given example can be treated in incremental terms, or whether it requires a realizational approach. We argue here that there are some examples best described by combining the two approaches, suggesting that the contrasting approaches are not mutually exclusive.

Our illustration involves multiple number marking on verbs. In some languages of North America verbs mark both subject number and event number using a largely undifferentiated set of plural formatives, drawn in whole or part from the same set used for marking nominal plurality. Thus verbs in the Yuman language Hualapai (Wataghomie et al. 1982) employ two main strategies for marking plurality: (i) suffix *-j*, and (ii) stem alternation (lengthening or initial reduplication). But exactly how these map onto the morphosyntactic paradigm varies across (apparently arbitrarily defined) classes of lexemes. Thus in Table 1, 'dig' lengthens the stem for all non-singular values, and adds *-j* for any kind of plurality, while 'tie s-thing large' uses stem lengthening to mark anything involving the plural, and then adds *-j* to both the short and long stems to yield forms with greater plurality for each stem alternant. In both types the morphological elements are the same and the compositionality is transparent, but the morphosyntactic mapping is different. Within those verbs whose plural stem is marked by reduplication, there is an additional distinction between the ones where the 2nd degree of plurality specifically denotes plurality of the event, and those for which it denotes plurality of the subject (Table 2). Again, this does not appear to be lexically specified and not semantically motivated.

'dig'		'tie s-thing large'	
content	*form*	*form*	*content*
SG SBJ, SG EVENT	hwal	gilgyo	SG SBJ, SG EVENT
	------	gilgyo-j	PAUCAL SBJ, SG EVENT
PAUCAL SBJ, SG EVENT	hwa:l	gilgyo:	SG SBJ, PL EVENT
PL SBJ and/or EVENT	hwa:l-j	gilgyo:-j	PL SBJ, PL EVENT

Table 1: Variant paradigmatic mappings in Hualapai (lengthening and/or *-j*)

'grind sthing crunchy in mouth'		'suck something hard in mouth	
content	*form*	*form*	*content*
SG SBJ, SG EVENT	juhu:d	jiboq	SG SBJ, SG EVENT
PAUCAL SBJ, SG EVENT	juhu:d-j	jiboq-j	PAUCAL SBJ, SG EVENT
SG SBJ, PL EVENT	jijuhu:d	jijboq	PL SBJ, SG EVENT
PL SBJ, PL EVENT	jijhuh:d-j	jijboq-j	PL SBJ, PL EVENT

Table 2: Variant paradigmatic mappings in Hualapai (reduplication and/or *-j*)

In Seri (a language isolate which along with the Yuman languages formed the core of the once-hypothesized Hokan family; Moser & Marlett 2010, Marlett 2015), number marking in both nouns and verbs involves dozens of distinct suffixes (alongside various types of stem

alternations). These display a clear cline of plurality with respect to each other, as in Table 3, which compares the two suffixes *-(t)ołka* and *-(t)ox*. Either suffix can be used for any of the three forms involving a plural value – this appears to be lexically specified – but crucially, when used together they line up so that *-(t)ołka* always realizes a value which is 'more plural' than *-(t)ox*. The suffixes thus encode degrees of plurality, but whether and how this is exploited by an individual paradigm is a matter of pure morphological stipulation.

→ *more plural*				
SG SBJ SG EVENT	SG SBJ PL EVENT	PL SBJ SG EVENT	PL SBJ PL EVENT	
-monxk	-mónax-ox	-mónx-ox	-mónx-ox	'go in zigzag'
-míːʔ	-míːʔ	-míːʔ-tox	-míːʔ-tołka	'be scarce'
-aksípχ-a	-aksípχ-ox	-aksípχ-ołka	-aksípχ-ołka	'glue'
-íχpχ-ax	-íχpχ-ołka	-íχipχ-ołka	-íχipχ-ołka	'be soft'

Table 3: Plurality cline in Seri: *-(t)ox* vs. *-(t)ołka*

In these examples the morphological paradigm is transparently composed through the accumulation of plural marking, but the precise linking to individual morphosyntactic values is lexically stipulated. This shows that incremental and realizational principles can operate in parallel, producing paradigms whose structure is at the same time functionally motivated and morphologically stipulated (i.e. morphomic).

References

Marlett, Stephen. 2015. *Cmiique Iitom: the Seri language*. Ms, University of North Dakota.

Moser, Mary & Stephen Marlett. 2010. *Comcaac quih yaza quih hant ihiip hac*. Mexico City and Sonora.

Stump, Gregory. 2001. *Inflectional morphology*. Cambridge.

Watahomigie, Lucille, Jorigine Bender & Akira Yamamoto. 1982. *Hualapai reference grammar*. Los Angeles.

Sonia Ben Hedia & Ingo Plag

Gemination and degemination in English affixation: Lexical strata, semantics, and phonetic evidence

In English, affixation may lead to the adjacency of two identical consonants across a morpheme boundary (e.g. *un#necessary*). The standard view of what happens in those cases is that with certain affixes the sequence of two identical consonants leads to a longer duration of the segment in question (gemination), while with other affixes the double consonant is of the same duration as a single consonant (degemination). A standard view in the literature is, for example, that the prefix *un-* geminates, whereas the prefix *in-* does not. Hence, the nasal in *unnatural* is predicted to be longer than the one in *uneven*, while the duration of the nasal is the same in *innumerous* and *inevitable* (cf. Cruttenden & Gimson 2014).

The idea of *in-* and *un-* displaying different behavior is in line with the theory of Lexical Phonology which assumes two lexical strata with two different phonological processes happening at each of these levels. While Level 1 affixes like *in-* integrate phonologically with their base and feature a weak morpheme boundary, Level 2 affixes like *un-* and *-ly* display less integration and a strong boundary (cf. Kiparsky 1982). Thus, *in-* is predicted to display degemination while *un-* and *-ly* are expected to geminate.

Until now, however, large-scale empirical evidence for this claim has been lacking. There is only one study (Oh & Redford 2012) that has empirically investigated gemination in English *in-* and *un-*prefixed words. They found that, while some *in-*prefixed words showed gemination, some did not. This clearly contradicts the assumption that *in-* always degeminates when a sequence of two identical consonants emerges.

The present paper investigates gemination and degemination with the three affixes *-ly*, *un-* and *in-* (the latter in the form of its allomorph *im-*) by analyzing English words taken from conversational speech (i.e. the Switchboard Corpus). Linear regression models were used to compare the duration of two identical cross-boundary consonants (e.g. *n#n* in *unnatural* or *l#l* in *really)* with the duration of a corresponding single consonant (e.g. *n#V* in *uneasy* or *C#l* in *probably)*. In our models, we controlled for the influence of pertinent phonetic aspects, such as word duration and the number of segments in the word, by including them as covariates in the models.

The results reveal that the Level 2 affixes *un-* and *–ly* differ in their gemination pattern. While *un-*prefixed words geminate, *ly-*suffixed words display degemination. This result clearly refutes claims according to which the two affixes should both geminate. The difference between *un-* and *-ly-*affixed words shows that the phonological behavior of complex words cannot solely be explained by assigning them to a particular stratum.

For the Level 1 prefix *in-* we found variation in gemination. In contrast to previous assumptions, our results reveal that not all *in-*prefixed words degeminate. We detected an effect of the number of nasals (m#m vs. m#C) in interaction with the semantics of the affix. For negative *in-*prefixed words, the nasal duration in words with two underlying nasals (e.g. *immobile*) was longer than in words with a single underlying nasal (e.g. *impossible*). For locative *im-* (e.g. *immigrate*), this effect could not be observed. Thus, negative *in-* geminates while locative *in-* degeminates. This result contradicts previous assumptions according to which negative *in-* should degeminate and thus also challenges the theory of Lexical Phonology, which predicts the degemination of *in-* as a Level 1 affix.

In sum, this paper calls simple categorical effects of lexical strata into question. Both, the difference in the gemination of Level 2 affixes (*un-* and *–ly*) and the gemination found with negative *in-* clearly refute previous assumptions made on the basis of such strata. It seems like the phonological/phonetic behavior of complex words cannot solely be explained by the existence of different lexical levels to which the different affixes belong.

References

Cruttenden, Alan & Alfred Charles Gimson. 2014. *Gimson's pronunciation of English*. 8th edn. London, New York: Routledge.

Godfrey, John J. & Edward Holliman. 1997. *Switchboard-1 Release 2*. Philadelphia, Pa: Linguistic Data Consortium.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In The linguistic society of Korea (ed.), *Linguistics in the morning calm. Selected papers from SICOL-1981 [Seoul International Conference on Linguistics]*, 3-91. Seoul, Korea: Hanshin Pub. Co.

Oh, Grace E.& Melissa A. Redford. 2012. The production and phonetic representation of fake geminates in English. *Journal of Phonetics* 40 (1). 82–91.

Neil Bermel, Luděk Knittl & Jean Russell

Near neighbours, analogy and competing forms in a nonce-word study of Czech morphology

Competing forms within a single morphological slot deviate from canonicity in Corbett's (2005: 33–34) sense of the word, and in general pose difficulties for models that seek to explain how speakers select a form for a particular slot. Additionally, for some lexemes these multiple forms are in competition. One such instance is the nominative plural of masculine animate nouns in Czech, where three desinential affixes occur with the stem: -i, -ové, or -é. In addition to phonological criteria, semantic criteria are also said to operate and some lexemes allow multiple forms.

Corpus data from the 100m-token SYN2005 subsection of the Czech National Corpus show that all three affixes are encountered frequently. The most common affix, -i, represents 61.7% of the tokens and is found with 56.8% of the types in the corpus. However, the two minority affixes have a strong position as well. The second most frequent, -é, appears with 28.3% of the tokens and 10.5% of the types, and the least-common, -ové, is found in 10.0% of tokens and 32.7% of types. These figures include numerous overlaps: many nouns are found with two or even all three affixes.

A classic dual-route model that posits a regular rule vs. exceptions will struggle to provide a realistic account of speaker choice here. We thus start from the presumption that a single-route model provides the best explanation: speakers store most forms, and rely on analogy (Skousen 1989), memory-based learning (Daelemans 2002), or another mechanism such as minimal generalization (Albright & Hayes 2003) to plug gaps. This assumption itself carries some baggage; analogical models, for example, will rely on near-neighbour data to produce these rare or unheard forms and inter-speaker variation may result from e.g. differences in vocabulary size (Dąbrowska 2008).

Our pilot study was designed to test whether near-neighbour data was more dependent on low-level schema like stem shape or higher-level ones like observed behaviour. Lacking a CELEX database or similar for Czech, we sought behaviour frequency data in the CNC. We identified three parameters: token frequency, type frequency, and competition (number of affixes represented in that neighbourhood), which yielded three clusters of environments (frequent/little variation; less frequent/higher variation; infrequent/higher variation). Four nonce words were created for each cluster and set in typical contexts; respondents saw the head form in full and were prompted to supply the plural. Finally, respondents read and answered comprehension questions on two texts to assess their reading skills as a proxy for language proficiency.

A repeated-measures ANOVA run on data from a pilot of 32 users in June 2015 showed that, as expected, the (low-level) stem ending explained most of the variance in the choice of affix ($F(1, 30) = 60.87$, $p < .001$). However, the (higher-level) frequency data represented a second significant, medium-sized factor ($F(2, 30) = 10.98$, $p < .001$). Semantic neighbourhoods did not play a significant role in choice ($p = .175$), suggesting that no recourse to additional information, as trialled in Keuleers et al. (2007) or Dodge & Lonsdale (2006), is necessary here. When added to the model, reading task success did not contribute to explaining the variance, suggesting that proficiency is not a factor here in speakers' choices.

References

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90. 119–161.

Corbett, Greville. 2005. The canonical approach in typology. In Frajzyngier, Zygmunt, Adam Hodges, David Rood (eds.), *Linguistic diversity and language theories*, 25–49. Amsterdam/Philadelphia: John Benjamins.

Czech National Corpus. 2000 - 2015. Prague: Czech National Corpus Institute.

Dąbrowska, Ewa. 2008. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language* 58. 931–951.

Daelemans, Walter. 2002. A comparison of analogical modeling of language to memory-based language processing. In Skousen, Royal, Deryle Lonsdale & Dilworth Parkinson (eds.), Analogical modeling: an exemplar-based approach to language, 157–179. Amsterdam & Philadelphia: John Benjamins.

Dodge, Inna Danielyan & Deryl Lonsdale. 2006. Modeling Russian Verbs of Motion: An Analogical Account. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 1239-1244. Mahwah, NJ: Erlbaum..

Keuleers, Emmanuel, Dominiek Sandra, Walter Daelemans, Steven Gillis, Gert Durieux, Evelyn Martens. 2007. Dutch plural inflection: The exception that proves the analogy. *Cognitive Psychology* 878. 283–318.

Skousen, Royal. 1989. *Analogical modeling of language.* Dordrecht: Kluwer.

Anja Binanzer & Klaus-Michael Köpcke

Competing factors for gender agreement: The Agreement Hierarchy from a learner's perspective

Due principally to the fact that German noun classification is based on multiple and thereby competing formal and semantic assignment rules (Köpcke & Zubin 2009), L2-acquisition presents learners with a complex challenge. The debate surrounding the order of acquisition of formal assignment rules – whether prior or subsequent to semantic rules (e.g. Wegener 2000) – has been informed in large part by results on L2-learners' use of determiners, whereas the use of other gender sensitive targets (attributive adjectives, relative, possessive and personal pronouns) has remained largely uninvestigated.

Therefore, we will treat German gender as an agreement category and will focus on the interaction between formal and semantic cues for the L2-development of the agreement system. Within the framework of usage-based and functionalist approaches (e.g. Ellis et al. 2015) we assume that form function mapping of different gender sensitive targets at the first acquisition stage is motivated semantically before the agreement system undergoes re-analysis as a grammatical system. In accordance with Corbett's (2006) Agreement Hierarchy we will show that this re-analysis begins with those targets which appear in the same syntactic domain as the nouns (determiners, attributive adjectives). In a subsequent step, grammatical agreement is marked on targets outside this syntactic domain.

To obtain evidence, we investigated 38 L2-learners (age 8-10, onset of German with about 3 years). The learners were divided into four groups according to their proficiency in German (low/high, tested by the C-test) and according to their L1 (Turkish/Russian). A monolingual control group of ten children was tested as well.

Over the course of six months, the children performed three written elicitation tasks. In each task five nouns were embedded in a short text. The children had to fill in a determiner in a blank preceding the nouns (cloze test). Furthermore, the children had to chose for three targets in different syntactic domains (attributive adjective, relative pronoun and personal pronoun) among three different gender marked forms (multiple choice), e.g.:

(1)	____	große$_F$	großes$_N$	großer$_M$	**MANN$_M$,**	der$_M$	die$_F$	das$_N$	nett	war, …	Es$_N$	Er$_M$	Sie$_F$	…
	DET	tall$_F$	tall$_N$	tall$_M$	man$_M$	who$_M$	who$_F$	who$_N$	nice	was	It$_N$	He$_M$	She$_F$	…

The 15 nouns (five for each gender) were chosen from the 130 most frequent nouns according to Pregel & Rickheit's (1987) productive lexicon of primary school children. Six of the nouns revealed the feature [+animate], nine of the nouns the feature [–animate]. For the two masculine and the two feminine animate nouns (*Mann, Vater – man, father; Frau, Mutter – woman, mother*), semantic and grammatical gender converged. For the two neutral animate nouns (*Baby, Kind – baby, child*) no semantic feature indicated their grammatical gender. For the inanimate nouns, only formal features (monosyllabic nouns; ending in open/closed *Schwa*-syllable) indicated their grammatical gender.

The results show that the development of the agreement system starts with a semantically based form function mapping. At the first acquisition stage, nearly complete agreement on all gender sensitive targets was observed for the nouns solely with converging semantic and grammatical gender, independently of their syntactic domain. At the same time, in the context of neutral animate and inanimate nouns, learners deviated from grammatically appropriate agreement patterns. At this point, they relied in their decisions systematically on the semantic

properties [±animate] of those nouns. Target agreement patterns based on their formal features were observed in a later stage of acquisition. Thereby, grammatical agreement was demonstrated to unfold sequentially on the different gender sensitive targets in the order predicted by Corbett's Agreement Hierarchy.

Furthermore, between the four groups correlations concerning the onset of marking grammatical gender were found: In general Russian learners performed better than Turkish learners. This observation supports the assumption that learners familiar with gender from their L1 transfer this knowledge to their L2-acquisition.

References

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Ellis, Nick C., Matthew B. O'Donnell & Ute Römer. 2015. Usage-based language learning. In MacWhinney, Brian & William O'Grady (eds.), *The Handbook of Language Emergence,* 163-180. New York: Wiley.

Köpcke, Klaus-Michael & David A. Zubin. 2009. Genus. In: Hentschel, Elke & Petra M. Vogel (eds.), *Deutsche Morphologie,* 132-154. Berlin: De Gruyter.

Pregel, Dietrich & Gert Rickheit. 1987. *Der Wortschatz im Grundschulalter. Häufigkeitswörterbuch zum verbalen, substantivischen und adjektivischen Wortgebrauch*. Hildesheim: Olms.

Wegener, Heide. 2000. German gender in children's second language acquisition. In Unterbeck, Barbara, Matti Rissanen, Terttu Nevalainen & Mirja Saari (eds.), *Gender in Grammar and Cognition,* part 1, 511-544. Berlin: De Gruyter Mouton.

Tomislava Bosnjak Botica & Gordana Hrzica

Competition and suppletion: Evidence from the Croatian verbal system

Contrary to the postulates of the blocking effect (Aronoff, 1976) according to which morphological doublets exclude each other, parallel forms with the same meaning are evidenced in languages. According to models of morphological competition (e.g. the Constant Rate Hypothesis - Kroch 1989), it is inevitable for such doublets to resolve resulting in the decline of one of the forms. Theoretical models that explain such competition revolve around two approaches. One is polarized between regular (default) and irregular paradigms. Irregular paradigms show evidence of overgeneralisation and paradigm change when frequency is taken into account (e.g. Pinker 1984). The other approach introduces language typology as a relevant factor, stating that morphological change is different in morphologically diverse languages, but always governed by similar principles (such as morphological productivity (Dressler 2005) or transparency (Slobin 1985)).

Croatian verbal morphology is typically described using verb class distinctions. The number of classes differs among approaches, but the basic criterion for the class division is the presence or absence and the type of suppletion in verb stems. Weak suppletion has often been analysed as a thematic vowel following the stem in some inflected forms, but the approach employed here will treat the thematic vowel (and more elaborate changes in verb forms) as a part of a stem. Verbs in one of the verb classes show no stem suppletion (eg. inf. hoda-ti 'to walk', pres.1.sg. hoda-m, pres.3.pl. hoda-ju), verbs in two classes have two stems (eg. inf. misli-ti 'to think', pres.1.sg. misli-m, pres.3.pl. misl-e) while verbs in other classes have three different stems (eg. inf. pisa-ti 'to write', pres.1.sg. piše-m, pres.3.pl. piš-u). Additionally, factors such as frequency, predictability and transparency of a verb class have been used to further explain the verb class system. The goal of this research was to describe verbs with dual-class membership (eg. inf. šetati 'to stroll', pres. 1. sg. šeta-m/šeće-m) and to determine whether the competition in such forms can be explained by their tending to conform to one default class or by other factors, namely, transparency.

Fifty dual-class verbs were found in two types of language sources: dictionaries of Croatian language and the hrWaC corpus (1,9 billion tokens, written language – Ljubešić and Klubička 2014). To obtain information on current usage of dual-class verbs, two forms for every dual-class verb were checked in hrWaC (1^{st} and 3^{rd} person singular or 1^{st} and 3^{rd} person plural).

The majority of the verbs were dual members of the one-stem class and of one of the three-stem classes. They showed a small preference towards the usage in the one-stem class (59%). Verbs with one-stem preference showed a high percentage of use in their preferred class (median 95%) compared to verbs with three-stem preference that showed a somewhat lower percentage of use in their preferred class (median 85%). This singles out the one-stem class as a good candidate for the default class. In such a case, according to dual-mechanism models, a frequency effect in class preference should be observed, i.e. more frequent verbs should be more inclined to three-stem usage while less frequent verbs should show a tendency towards one-stem (potential default) class usage. However, regression analysis showed that frequency could not significantly predict class-preference ($F=(1, 21)=.023$, $p=.880$). On the other hand, when transparency in the transforming verb stem was taken into account, it was a significant predicting factor for the one-stem class preference ($F=(1, 30)=11.399$, $p=.002$, $R^2=.275$). Dressler (1999) states that in strongly inflectional languages there is often no default or only a weak default among competing morphological patterns. An approach which takes

morphological transparency into consideration would be more comprehensive in describing the outcomes of morphological competition in dual-class verbs in Croatian than one focused on the default class.

References:

Aronoff, M. (1976). Word Formation in Generative Grammar. Cambridge: MIT Press.

Dressler, W. U. (2005). Morphological Typology and First Language Acquisition: Some Mutual Challenges. In: G. Booij et al (eds.), Morphology and Linguistic Typology (MMM4). University of Bologna.

Dressler, W. U. (1999), Why collapse morphological concepts?, Behavioural and Brain Sciences, 22, 1021.

Kroch, A. (1989). Reflexes of Grammar in Patterns of Language Change. Language Variation and Change, 1:199–244.

Ljubešić, N. and Klubička, F: {bs,hr,sr}WaC — Web corpora of Bosnian, Croatian and Serbian**.** Proceedings of the 9th Web as Corpus Workshop (WaC-9).

Pinker, S. (1984). Language Learnability and Language Development, Cambridge: HUP.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In: D. I. Slobin (ed.). The Crosslinguistic Study of Language Acquisition, Vol. 2. Theoretical Issues. Hillsdale, NJ: LEA, 1157–1256.

Gilles Boyé & Gauvain Schalchli

Competitive Predictions: The case of the past-participle converts in French

1 Data

Plénat (1987) proposes an analysis of French past participles based on a single stem and abstract derivations in the SPE framework. In the course of the paper, he remarks on deverbal nouns made by converting the feminine form of the past participles regardless of their conjugation class:

base verb	past participle	deverbal noun
ARRIVER ('to arrive')	*arrivée*	ARRIVÉE ('arrival')
PERCER ('to pierce')	*percée*	PERCÉE ('breakthrough')
ÉCLAIRCIR ('to clear')	*éclaircie*	ÉCLAIRCIE ('clearing')
SAISIR ('to grab')	*saisie*	SAISIE ('grabbing')
METTRE ('to set')	*mise*	MISE ('setting')
PRENDRE ('to take')	*prise*	PRISE ('taking')
TENIR ('to hold')	*tenue*	TENUE ('holding')
BATTRE ('to beat')	*battue*	BATTUE ('hunting beat')
CRAINDRE ('to fear')	*crainte*	CRAINTE ('fear')
FEINDRE ('to feign')	*feinte*	FEINTE ('feint')

This conversion has been studied in competition with other constructions (see for example Villoing *et al.*, 2010; Tribout and Villoing, 2014) but Plénat also notes an internal competition. Some verbs from the third conjugation lack a past participle convert but display another deverbal noun in *-te* in a kind of complementary distribution:

base verb	past participle	deverbal noun
VENDRE ('to sell')	*vendue*	VENTE ('sale')
PONDRE ('to lay eggs')	*pondue*	PONTE ('egg-laying')
SUIVRE ('to follow')	*suivie*	SUITE ('sequel')
DEVOIR ('to owe')	*dûe*	DETTE ('debt')
DESSERVIR ('to service')	*desservie*	DESSERTE ('service')
RECEVOIR ('to receive')	*reçue*	RECETTE ('proceeds')

2 Analysis

Our study is based on the observation of direct analogies between forms in morphological paradigms and aims to answer the Paradigm Cell Filling Problem (Ackerman *et al.*, 2009) in the context of this derivation by predicting relations between forms in the verb paradigm and the resulting nominalization.
Our analysis works in three steps. First, we extract analogies and distribution classes from our training lexicon then we populate paradigms by organizing the analogy competition and finally we harvest the complete paradigms including the nominalizations.

2.1 Extracting Analogy Classes

Building on the Minimal Generalization Learner (Albright, 2002; Albright and Hayes, 2003) and proposals by (Bonami and Boyé, 2014; Bonami and Luís, 2014), we extract minimally general analogies between every pair of forms in our training set. These analogies have partially overlapping scopes and we further organize them by making classes of analogies in competition and calculating their distribution.

2.2 Populating Cells

Populating cells for each lexical entry is based on a two-round mechanism. In the first round, we use the analogy classes and their distribution (red arrows in Fig. 1) to populate all cells with all possible candidates starting from the sparse lexical information (red dots). In the second round, all cells (orange dots) have been filled and we use this information (green dots) to populate the cells again, using all the knowledge about analogies between all cells (orange arrows).

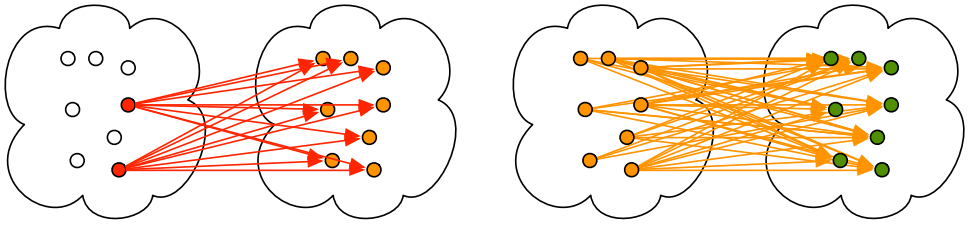


Figure 1: The two-round election

2.3 Harvesting Paradigms

From this anarchic configuration, we extract complete paradigms, the inflected forms and the nominalization, using the following morphological hypothesis:

(1) paradigms are structures were **all** members are analogically related

A candidate paradigm is a set of form candidates where all the forms in all the cells have mutual analogies. In Fig. 2, among the forms in the various cells (green dots), the blue ones form a clique (completely interconnected network in blue) where all forms are connected to all the other forms.

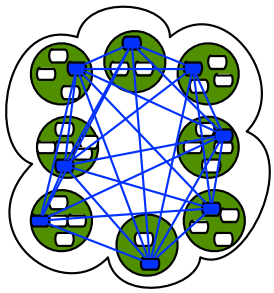


Figure 2: Extracting a candidate paradigm

3 Conclusion

With this competitive prediction method, we start from the partial inflectional paradigms of the French verbs in our lexicon and generate complete paradigms (inflected forms and nominalization) capturing the complementary distribution in the deverbal nouns (past participle converts vs *-te* deverbal nouns) replicating the result of the abstract phonological model of Plénat (1987) with only surface knowledge of a training set and an organized competition of analogies. The analogies used are completely local to form pairs and inflectional classes are not encoded in the analysis but rather emerge from the data allowing in turn the discrimination among the deverbal nouns allomorphs.

References

Ackerman, F., Blevins, J. P., and Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins and J. Blevins, editors, *Analogy in grammar: Form and acquisition*, pages 54–82. Oxford Scholarship Online.

Albright, A. (2002). *The identification of bases in morphological paradigms*. Ph.D. thesis, UCLA.

Albright, A. and Hayes, B. (2003). Rules vs. analogy in english past tenses: A computational/experimental study. *Cognition*, **90**, 119–161.

Bonami, O. and Boyé, G. (2014). De formes en thèmes. *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux*, pages 17–45.

Bonami, O. and Luís, A. R. (2014). Sur la morphologie implicative dans la conjugaison du portugais : une étude quantitative. *Mémoires de la Société de Linguistique de Paris*, **22**, 111–151.

Chomsky, N. and Halle, M. (1968). *The Sound Patterns of English*. M.I.T. Press, Cambridge.

Plénat, M. (1987). Morphologie du passé simple et du participe passé des verbes de l' ≪ autre ≫ conjugaison. *ITL Review of Applied Linguistics*, **77-78**, 93–150.

Tribout, D. and Villoing, F. (2014). La composition vn et la conversion v>n en français: un nouveau cas de concurrence morphologique ? In F. Villoing, S. David, and S. Leroy, editors, *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux*. Presses Universitaires de Paris-Ouest.

Villoing, F., Ferret, K., and Soare, E. (2010). Les noms d'événement en –age et en –ée : une différenciation fondée sur l'aspect grammatical. *Actes du 2ème Congrès Mondial de Linguistique Française*, pages 945–968.

Gjorgje Bozhoviq

Split plurality, historicity and the *longue durée*

Since Ritter (1991), it has been argued for an additional functional projection between DP and NP that accommodates number inflection (NumP) and is generally assumed to be the default syntactic location of number features within the DP. On the other hand, more recent research on idiosyncratic plurality such as irregular and lexical plurals argues for a possibility that number features may be found on the categorial head *n*, closer to the noun (Kramer 2012; Acquaviva 2008; among others). Kramer (*ibid.*) argues that number is split between Num for regular plurals and the nominalizing head *n* for idiosyncratic plurality.

Following Kramer (*op. cit.*), we present a split analysis of plurality in Albanian, which possesses a highly elaborated plural morphology in nouns, ranging from number syncretism (1), to different allomorphy patterns in plural stems (2), to an extensive inventory of various affixal plural markers (3).

(1) a. *mësues(e)* 'teacher' → pl. *mësues(e)*
b. *kohë* 'time' → pl. *kohë*

(2) a. *mik* 'male friend' → pl. *miq*
b. *dash* 'ram' → pl. *desh*
c. *ishull* 'island' → pl. *ishuj*

(3) a. *fshat* 'village' → pl. *fshatra*
b. *prind* 'parent' → pl. *prindër*
c. *njeri* 'man' → pl. *njerëz*
d. *hero* '*id.*' → pl. *heronj*
e. *shok* 'male friend, comrade' → pl. *shokë*
f. *fushë* 'field' → pl. *fusha*
g. *vend* 'place, town, land' → pl. *vende*

Moreover, there are instances of doubled plural marking in some nouns (4), and nouns with more than one plural form are also not rare (5).

(4) a. *vëlla* 'brother' → pl. *vëlle-z-ër* (stem alternation with two plural suffixes)
b. *yll* 'star' → pl. *yj-e* (stem alternation and a plural suffix)

(5) a. *flamur* 'flag' → pl. *flamuj* / *flamurë* / *flamura*
b. *yll* 'star' → pl. *yj* / *yje*
c. *ujk* 'wolf' → pl. *ujq* / *ujqër*
d. *kohë* 'time' → pl. *kohë* / *kohëra*

We argue that the split plurality in Albanian has resulted from an ongoing competition between various plural markers, that was triggered by the phonological history and reanalysis of historical plurals in the language (Demiraj 1985). We argue that the older plural markers, surviving synchronically as stem allomorphy in Albanian, have been pushed down from Num to *n* by reanalysis, allowing for the novel affixal plural markers to redundantly reappear and compete among themselves for the realization of the Num head of NumP; as the singular and plural paradigms leveled and number syncretism occurred, in line of the principle of contrast (Carstairs-McCarthy 1994). In that sense, splitting the number features between Num and *n* has worked as a competition resolution strategy in Albanian.

From a more general point of view, we argue that the data from Albanian conforms to Carstairs-McCarthy's (2010) observation that "morphology exists because morphophonology exists", i.e. "phonology gives rise to morphophonology which gives rise to morphology". In other words, a sharp (morpho)phonological split between the historical and emerging plurals in Albanian has led to a state of equilibrium, in the sense of Aronoff (*online*), "in which no competitor has a clear advantage and such states may persist for long periods".[1] This state of equilibrium is, however, not static itself, but constantly under pressure to resolve, as the data from Albanian shows under thorough examination.

References

Acquaviva, P. (2008). *Lexical Plurals*. Oxford: Oxford University Press.

Aronoff, M. (online). Competition and the Lexicon. To appear in *Proceedings of the Annual Meeting of La Società di Linguistica Italiana 2013*. Available at https://linguistics.stonybrook.edu/fa-culty/mark.aronoff as of June 2015.

[1] Aronoff (*ibid.*) makes a note of a similar state of equilibrium in Maltese plural morphology.

Carstairs-McCarthy, A. (1994). Inflection classes, gender, and the principle of contrast. *Language* 70(4): pp. 737–788.

Carstairs-McCarthy, A. (2010). *The Evolution of Morphology*. Oxford: Oxford University Press.

Demiraj, Sh. (1985). Singularizimi i shumësit të emrave në gjuhën shqipe (vështrim historik). La singularisation du pluriel des substantifs en albanais (étude historique) / L'évolution vers une neutralisation de l'opposition de nombre des substantifs albanais. *Studime Filologjike* 39(4): pp. 13–22.

Kramer, R. (2012). A split analysis of plurality: Evidence from Amharic. In: N. Arnett & R. Bennett (eds.), *The Proceedings of WCCFL 30*. Somerville, MA: Cascadilla Press, pp. 226–236.

Ritter, E. (1991). Two functional categories in noun phrases: Evidence from Modern Hebrew. In: S. D. Rothstein (ed.), *Perspectives on Phrase Structure: Heads and Licensing*. San Diego: Academic Press, pp. 37–62.

Marco Bril & Martine Coene

Written inflections in L2 production in advanced learners of French

Inflectional morphology is very hard to acquire in L2, even for learners who have reached an advanced level of L2 (e.g. Bartning 2000). From a psycholinguistic approach, inflection errors made by L2 learners, can be related to locality: agreement processes which are more local in the sentence (e.g. between N and Adj) are mastered earlier than those that are less local (e.g. between N and participle) (*Processability Theory,* Pienemann 1989).

Here, we focus on data from advanced Dutch L1 – French L2 learners to test this locality effect in adjectival inflection. In French, inflection is present both in adjectives and participles; these exhibit agreement with the noun in number and gender. In contrast, inflection in Dutch is only present in adjectives in [+ neuter, + singular, + definite, + attributive] context. In all other contexts no inflection shows up. For Dutch L1 – French L2 learners this typological difference can affect the complexity of the acquisition of French adjectival inflection. If agreement with adjectives is indeed the result of a more local syntactic relation than with participles, this leads to the hypothesis that the first is fully mastered at an advanced level, while the latter is not. We further focused on past participle constructions comparing agreement in dislocation and relative clauses. Here also, differences in locality are expected to lead to different error rates, as agreement in dislocation constructions is known to be more locally than in relative clauses.

In addition to locality, the overt/covert nature of phonological expression of the inflection morpheme has also been related to the error rate of written inflection. Namely, advanced L2 learners seem to use phonological cues in processing inflectional morphology (e.g. Carrasco-Ortiz & Frenck-Mestre 2014). For L1 learners it has been shown that written inflection errors in French are related to the homophonic nature of the inflection morpheme (Largy & Fayol 2001). By testing the effect of locality on the one hand and of 'audibility' of the inflection morpheme on the other hand, this study aims to provide new insights with respect to the role of potential syntactic and phonological features on the acquisition of written inflection in advanced L2 learners.

The participants in this study consisted of advanced Dutch L1-French L2 learners who were recruited at a secondary school in The Netherlands. All participants studied French for 6 years at the highest level of Dutch education. The test consisted of a fill-in-the-gap elicitation task covering three conditions: agreement in two lexical categories (adjective (1a) vs. participle (1b)), in utterances with different syntactic complexity (dislocation (2a) vs. relative clause (2b)), and exhibiting differences in 'audibility' (inflection morpheme phonologically expressed (3a) or silent (3b)). Participant's scores were computed as correctness scores per condition.

(1a) La seul_____ période où on est partis, était l'été
(1b) Une page que nous avons lu_____, concerne l'anglais
(2a) Je l'ai arrosé_____, la plante
(2b) La clé que j'avais perdu_____, était celle de ma voiture
(3a) Une remarque que nous avons fait_____ à ce propos, n'a pas résolu le problème
(3b) La boutique qu'on a trouvé_____ au bout de la rue, vend des vêtements de marque

The results show that more inflection errors occur with participles than with adjectives ($t(25) = 9{,}20$; $p = .000$). Also, participles in relative constructions show more inflection errors than in dislocation constructions ($t(25) = 4{,}45$; $p = .000$). However, no significant effect was found for the 'audibility' condition, i.e. no more correct responses were found for inflection

morphemes that are phonologically expressed, compared to their silent counterparts ($t(25) = .166$; $p = .87$). Based on these findings, we take the effects of locality to be in line with Processability Theory. The fact that no significant effect is found for the 'audibility' condition, is in contrast to what has been found for the phonological influence in processing L2 inflectional morphology. Namely, the phonological effect which occurs in processing French inflection, does not in written production of inflection morphemes in L2 French.

References

Bartning, Inge. 2000. Gender Agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica*. 225-237.

Carrasco-Ortiz, Haydee. & Cheryl Frenck-Mestre. 2014. Phonological and orthographic cues enhance the processing of inflectional morphology. ERP evidence from L1 and L2 French. *Frontiers in psychology* (5). 888.

Largy, Pierre. & Michel Fayol. 2001. Oral cues improve subject-verb agreement in written French. *International Journal of Psychology* 36(2). 121-131.

Pienemann, Manfred. 1989. Is language teachable? Psycholinguistic experiments & hypotheses. *Applied Linguistics* 10(1). 52-97.

Shoju Chiba

The lexical realization patterns and their relationship to morphological productivity: The Finnish symbiotic derivations

How multiple near-synonymous derivations co-exist (and why so) or, more broadly, how the distributions of different derivational types overlap each other has received little attention. Pala & Hlaváčková (2007), for instance, attempt to enrich the Czech WordNet with derivational relations (subnets or derivational nests), but a closer examination of the clusters within different derivational relations in their database is still required. The notion of distributional semantics (Harris 1970) has recently adopted new computational methodologies and extended the compositional perspective of semantic distributionality to the analysis of derivational words (Lazaridou et al. 2013; Marelli & Baroni 2015).

Finnish, a Uralic language not related to the Indo-European family, has a rich inventory of morphological affixes (Karlsson 1983; Koivisto 2013). Its morphology includes many near-synonymous derivational affixes as well. Focusing on the verb-derived derivations of Finnish, this study explicates the lexical realization patterns of derived words – how the different derivational counterparts are lexically distributed – and argues that their lexical overlaps are interestingly interrelated to the morphological productivity of the derivational types attested in corpus data.

This paper first illustrates the derivational overlaps using dictionary entries (*Nykysuomen sanalista*). The data was analyzed morphologically and the bases and affixes of the derived words were listed in a database. Next, using a large-scale corpus of modern written Finnish (*The Finnish Text Collection*, or *FTC*, which includes 180 million words), I demonstrate how the co-existence patterns of certain derivational counterparts (doublets, triplets, etc.; words more or less semantically synonymous) are observed through the corpus data. The lexical-derivational database was built on the basis of the inventory of the derivational types compiled above, and enhanced and augmented with the *FTC*. The distributions of the different derivational affixes are then analyzed quantitatively and their morphological productivity is measured.

The analysis reveals that there exist interesting lexical-derivational relationships among derivational affixes: Firstly, we can observe a kind of "symbiosis" among certain derivation types (e.g., between the nominalization affix pair *-nta*/*-ntä* and *-ma*/*-mä*), which arguably contributes to reinforce the productivity of the mutually symbiotic derivation types. Not every type of symbiotic derivation shows a high productivity index (*P* in Baayen 2001). Instead, some demonstrate similar (but moderate) productivity on the scale, which is observable when we count the growth rate of the vocabulary with a bootstrapping method. In contrast, there are other types of seemingly symbiotic derivations where one of the derivation patterns shows a higher productivity index, which possibly facilitates the popularity of the other symbiotic affixes. Both types of lexical-derivational overlaps are persistently found throughout even low frequency lexemes in a corpus.

The question of derivational rivalries has been approached either from the process-based view of word formation, i.e. how a rule is applied or hampered in what contexts/conditions (Aronoff 1976; Marle 1985; Rainer 1988), or on the descriptive ground of word-based morphology (e.g., Haspelmath 2002: 50-51, 173-174). Neither approach is sufficient, however, to account for the distributions of the variety of near-synonymous derivations, as posited by Carstairs-McCarthy

(1992: 151-152). This study shows there may exist a kind of paradigmatic-distributional relationship between certain derivational types, which facilitates their productivity.

The findings of this study extend to even more "subtle," perhaps at first glance unrelated or trivial, relationships, where we hardly perceive any direct semantic, formal, or grammatical connections between the observed symbiotically distributed derivation types (e.g., nominalization *-nti* and reflexive intransitive *-u-*/*-y-*). The question of how they become such "sociable" derivation types remains to be answered, however.

References

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

Baayen, Herald. 2001. *Word frequency distributions*. Dordrecht: Kluwer.

Carstairs-McCarthy, Andrew. 1992. *Current morphology*. London: Routledge.

Harris, Zellig. 1954. Distributional structure. *Word*. 10(23). 146-162.

Haspelmath, Martin. 2002 *Understanding morphology*. London: Arnold.

Karlsson, Fred. 1983. *Suomen kielen äänne- ja muotorakenne*. Porvoo: Werner Söderström.

Koivisto, Vesa. 2013 *Suomen sanojen rakenne*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Lazaridou, Angeliki, Marco Marelli, Roberto Zamparelli & Marco Baroni. 2013. Compositional-ly derived representations of morphologically complex words in distributional semantics. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Proceedings of ACL), 1517-1526.

Marelli, Marco & Baroni, Marco. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122(3). 485-515.

Marle, Jaap van. 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.

Pala, Karel & Dana Hlaváčková. 2007. Derivational relations in Czech WordNet. *Balto-Slavonic Natural Language Processing 2007* (BSNLP), 75-81.

Rainer, Franz. 1988. Towards a theory of blocking: Italian and German quality nouns. In Booij, Geert & Jaap van Marle (eds.), *Yearbook of Morphology 1988*, 155-185. Dordrecht: Foris.

Greville G. Corbett & Sebastian Fedden

One system or two? A Canonical Typology approach

Depending on the research tradition, there are types of data which are handled as representing a single (complex) system or elsewhere as two related systems. The particular analysis is often assumed rather than argued for. We need a more principled approach here, since we are aiming at a general typology of nominal classification (including gender and classifiers), and a prerequisite for this is a means of determining the number of systems involved in a given language.

Let us take an example which is relatively straightforward. We analyse Italian as having gender and number, despite the fact that these features are realized cumulatively. This can be seen in the following adjective paradigm.

nuovo 'new'	SINGULAR	PLURAL
MASCULINE	nuov-o	nuov-i
FEMININE	nuov-a	nuov-e

Why does it seem evident that our analysis should indeed have two features, each with two values, rather than a single feature with four values?

First, gender and number in Italian are separate systems because they are orthogonal to each other. Given a noun which takes masculine agreement we cannot predict its number value. The same holds for a noun that takes feminine agreement. And equally, given a number value we cannot predict gender.

The **second** argument concerns form. While our example shows cumulation of number and gender, there are number forms (of synthetic verbs) which show number without gender, as in example (1).

(1) Lui/lei parl-a oggi
3SG.M.SBJ/3SG.F.SBJ speak-3SG.SBJ today
'He/she speaks today.'

These observations suggest two parameters for our typology: (i) the degree to which the semantics of the two candidate systems are orthogonal to each other, and (ii) the degree to which their means of realization are distinct. We adopt a canonical approach (Corbett 2013, Corbett & Fedden 2017[2015]). In canonical terms, the higher the degree of orthogonality of the semantics and the more distinct the formal realization, the more clearly different the two systems are. We give a typology of nine possibilities, from a canonical single system to a situation with canonically two systems, specifying degrees of semantic and formal overlap in between. We then apply this typology to a sample of languages (from a larger project set). Each of these is traditionally analysed as having a gender system and at least one system of classifiers.

A key language in our typology is the Papuan language Mian, which is analyzed as having four genders as well as six classifiers that appear as prefixes on a subset of verbs (Fedden 2011). The forms which are used in gender agreement and in the verbal classifiers are completely different. On the other hand, and more importantly, there is considerable overlap in the semantics. In many cases, if we know the value of a noun

in one system, we can predict the value it has in the other. For example, from the perspective of gender, all masculine nouns take the M-classifier. From the perspective of the classifiers, all nouns which take the long, covering and bundle classifier have neuter 1 gender. The degree of orthogonality of the two candidate systems in Mian is actually rather low. If we multiply four genders with six classifiers we get 24 theoretical possibilities of which only nine are attested. The cells filled with examples in the system matrix (Table 1) are the attested combinations.

Table 1. Mian gender and classifiers: orthogonality

	MASCULINE	FEMININE	NEUTER 1	NEUTER 2
M-classifier	man, boy, boar	---	sleeping bag, plate, mosquito net	---
F-classifier	---	woman, girl, sow	---	house, steel axe, money (kina note)
Long	---	---	tobacco, eating implement, bush knife	---
Bundle	---	---	string bag (large), tobacco pouch	---
Covering	---	---	blanket, band aid	---
Residue	---	tortoise	cassowary egg, plane, hat	---

We conclude that our method provides the means to be clear and explicit about analyses into one or more than one system. It provides the means for analysing traditional gender systems and traditional classifiers according to a common standard. This is the necessary basis for a comprehensive typology of nominal classification, and in particular of languages which have been argued to have both systems concurrently.

References

Corbett, Greville G. 2013. Canonical morphosyntactic features. In Dunstan Brown, Marina Chumakina & Greville G. Corbett (eds.), *Canonical morphology and syntax*, 48–65. Oxford: Oxford University Press.

Corbett, Greville G. & Sebastian Fedden. 2017 (2015 online). Canonical gender. *Journal of Linguistics*, 1-37. [doi:10.1017/S0022226715000195]

Fedden, Sebastian. 2011. *A grammar of Mian*. Berlin: De Gruyter Mouton.

Berthold Crysmann

Positional competition in Chichêwa: A unified approach to templatic and scopal order

This paper connects to an ongoing discussion as to the impact of semantic factors on morph order: while, Rice (2011) or Aronoff & Xu (2010) claim that scope facts are not restricted to derivational morphology, Stump (2001) or Spencer (2003) argue that in inflectional morphology, order is largely conventionalised. Recently, Crysmann & Bonami (2015) suggested that most cases of purported scope effects in inflection are either not semantic in nature, or do not show alternation in surface patterns, making it impossible to encode semantic differences by means of order.

Templatic and scopal order in Chichêwa (Hyman, 2003) constitutes one of the rare cases where difference in order may truly encode differences in interpretation: markers of grammatical function change (Causative-Applicative-Reciprocal-Passive) may either occur in a templatic order, or, in circumscribed cases, in an anti-templatic order, marking scope. An important finding of Hyman's is that templatic order encodes a superset of the scopal readings available for anti-templatic order. As shown in Table 1, order conforming to the Bantu CARP template is neutral with respect to relative scope: Thus, the sequence *its-il* may encode both the applicative of a causative, or the causative of an applicative. The same observation holds for template-conforming AR and CR orders (cf. Hyman, 2003).

Anti-templatic order may arise as the result of two different deviations: displacement of the C-marker to the right, or by means of doubling R to the left of A. As for interpretation, all anti-templatic orders establish narrow scope of the reciprocal, either with respect to causative or the applicative.

Surface	Pattern	Scope	CARP
its-il	CA	{ }	yes
its-an	CR	{ }	yes
an-its	RC	$\{\rho < \kappa\}$	no
its-il-an	CAR	{ }	yes
il-an-its	ARC	$\{\rho < \kappa\}$	no
an-il-an	RAR	$\{\rho < \alpha\}$	no
its-an-il-an	CRAR	$\{\rho < \alpha\}$	no
its-an-il-an	RCAR	$\{\rho < \alpha, \rho < \kappa\}$	no
its-an-il-an	RARC	$\{\rho < \alpha, \rho < \kappa\}$	no

Table 1: Possible orderings and scopings

Hyman (2003) proposes an OT-analysis involving different rankings of scope constraints with respect to the template constraint. Moreover, to capture doubling of RAR, he invokes conjunction of the relevant constraints. In all those cases that feature overabundance, he postulates an indeterminate ranking of templatic and scopal constraints. While, at first sight, capable of capturing the facts, there are conceptual problems with his analysis: first, low-ranked mirror constraints never have any impact on the observable data. Second, Hyman's constraint ranking for doubling violates Pāṇini's Principle, which is actually embraced by OT (Prince & Smolensky, 1993). Abandoning this principle would turn OT into one of the very few theories that reject the Elsewhere Condition in favour of purely stipulated constraint rankings. Assuming Pāṇini, however, Hyman's theory is unstatable in OT. This confirms again that the ranking architecture of OT does not lend itself naturally to an account of overabundance.

Instead, I shall build on the neo-templatic framework laid out in Crysmann & Bonami (2015), which captures variation in order by positional competition between partially specified inviolable order constraints, and show that the observable patterns are accounted for uniformly with an enriched positional template.

Since all deviations from templatic order involve the reciprocal, the core idea behind the formal analysis is to provide additional options for marking narrow scope of ρ: either by positional underspecification of C, or by a morphotactic construction licensing RAR. To capture the associated scope facts, I follow

Spencer (2003) and incorporate outscoping relations into the morphosyntactic property set realisation rules operate on.

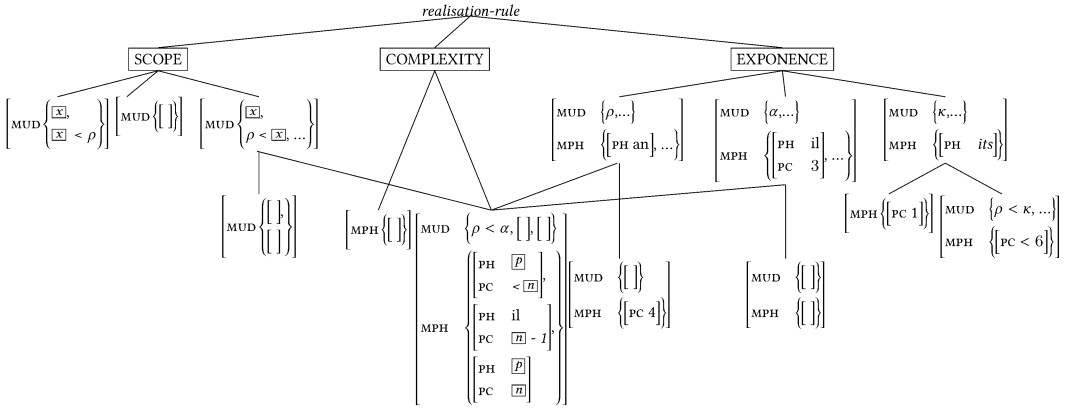


Figure 1: Formalisation

The analysis proposed puts positional competition centre stage, tying placement possibilities directly to interpretation. Thus, my treatment of Chichêwa resembles both positional disambiguation in Swahili and Mari partial ordering (Crysmann & Bonami, 2015), accounting for overabundance by having positionally underspecified exponents occupy the slots left by the more restricted ones.

References

Aronoff, Mark & Zheng Xu. 2010. A realization optimality-theoretic approach to affix order. *Morphology* 20. 388–411.

Crysmann, Berthold & Olivier Bonami. 2015. Variable morphotactics in information-based morphology. *Journal of Linguistics* FirstView. 1–64. doi:10.1017/S0022226715000018.

Hyman, Larry. 2003. Suffix ordering in bantu: A morphocentric approach. *Yearbook of Morphology 2002* 245–281.

Prince, Alan & Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Tech. rep. Rutgers Piscataway.

Rice, Keren. 2011. Principles of affix ordering: An overview. *Word Structure* 4. 169–200.

Spencer, Andrew. 2003. Putting some order into morphology: reflections on Rice (2000) and Stump (2001). *Journal of Linguistics* 39. 621–646.

Stump, Gregory. 2001. *Inflectional morphology*. Cambridge: CUP.

Alexis Dimitriadis, Natalie Boll-Avetisyan & Tom Fritzsche

Learnability of fusional and agglutinative morphology: Insights from artificial language learning

The opposition of fusional and agglutinative morphological systems is one of the oldest concepts in language typology and comparative linguistics, dating back to Schlegel and Humboldt in the 1800s. The trade-off between the two morphological systems is often talked about in terms of learnability: it should be easier to acquire transparent agglutinative morphology than opaque fusional morphology (e.g., Pinker 1996; Zuidema 2003; Neeleman & Szendrői 2005; Fasanella 2014). However, these works do not support their claims with evidence from human learning behavior.

In the present study, we directly compare the learnability of fusional and agglutinative systems in an artificial language learning task (Reber 1967). In addition, we assess the influence of the native morphological system of the learners. Two predictions will be evaluated: (1) learning of an agglutinative compared to a fusional system proceeds quicker and yields higher accuracy rates and (2) the native morphological system affects learning rate and outcome.

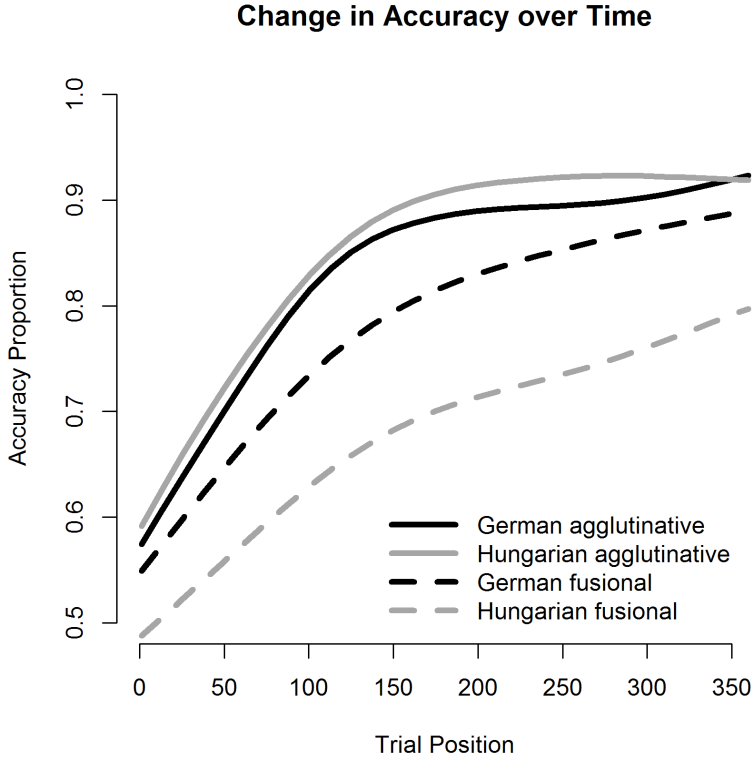
20 speakers of German and 20 of Hungarian (typical fusional and agglutinative languages) participated in the study. Each participant was randomly assigned to either the agglutinative or the fusional condition. The artificial languages were made up of morphologically complex nonwords, consisting of stems marked with prefixes that indicate number (singular, dual, plural) and gender (specifically, semantically-based noun classes: objects, foods, and animals; cf. Corbett 1991 for their characterization as “genders”). This grammatical system is typologically attested, being loosely patterned after that of Swahili and related Bantu languages. Stems are bisyllabic (e.g., /ki:tɛp/, /pʊro:f/) and prefixes are monosyllabic (e.g., /vu:/, /zy:/). In the fusional condition, stems have a single prefix that indicates a specific number-gender combination; while in the agglutinative condition there are two prefixes, with the exponent of number preceding gender.

a.	bo:–ki:tɛp sg+class3–cookie ‘one cookie’	(fusional)	b.	vu:–zy:–ki:tɛp sg–class3–cookie ‘one cookie’	(agglutinative)

In the fusional condition there are nine different morphemes, each representing one number-gender combination of the 3x3 paradigm. In the agglutinative condition, there are six morphemes (3+3), each representing one value for either number or gender yielding nine different prefix-prefix combinations (i.e., there is no syncretism in these artificial languages.)

In contrast to classic artificial language learning paradigms with separate training and test phases, we employ a novel paradigm of concurrent training and testing. Participants are exposed to 360 trials during the experiment. In each trial, an inflected word is presented auditorily along with two images on the screen: the target and the distractor, which differs from the target in number, gender, or both. Participants must identify the image matching the stimulus word, by pressing one of two buttons. They receive visual feedback after each response.

By design, it is difficult or impossible to learn the stems. Participants need to acquire the morphology in order to improve their performance during the experiment. Any opportunity for lexical learning is in any event identical across conditions.



Data were analyzed using Generalized Additive Mixed Models (van Rij et al. 2015), which model non-linear effects over time by simultaneously accounting for random effects. The figure shows the modelled learning curves for each of the four participant groups. It can be seen that learning takes place in all groups: Accuracy increases over time and settles at around 80 to 90 percent. Learning rate and ultimate attainment are highest for the agglutinative system, with no difference between German and Hungarian participants ($z=0.94$, $p=.35$, n.s.). Within each language group, accuracy is lower for fusional than for agglutinative morphology. This difference is significant from trial 16 until the end in Hungarian speakers ($z>1.96$, $p<.05$) and from trial 99 to 215 in German speakers ($z=2.04$, $p<.05$).

While the agglutinative system was learned equally well by both groups of speakers, German speakers scored higher than Hungarians in learning the fusional system (from trial 77 onwards; $z>1.96$, $p<.05$). We interpret this as indicating that acquisition of the fusional system, being harder, benefits more from native-language experience.

References

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

Fasanella, Adriana. 2014. *On How Learning Mechanisms Shape Natural Languages*. Ph.D. Thesis, Universitat Autònoma de Barcelona.

Neeleman, Ad & Kriszta Szendrői. 2005. Pro-Drop and Pronouns. *Proceedings of WCCFL 24.*

Pinker, Steven. 1996. *Language Learnability and Language Development*. 2nd edn. Harvard University Press.

Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6(6). 855–863.

Rij, Jacolien van, Martijn Wieling, R. Harald Baayen & Hedderik van Rijn. 2015. *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. R package v.1.0.5.

Zuidema, Willem. 2003. Modeling language acquisition, change and variation. In Simon Kirby (ed.), *Language Evolution and Computation: Proceedings of the Workshop/Course at ESSLLI*, Vienna 2003.

Wolfgang U. Dressler, Lavinia Merlini Barbaresi, Sonja Schwaiger, Jutta Ransmayr & Katharina Korecky-Kröll

Rivalry and synergy among Italian and German diminutives

Different productive diminutive suffixations of the same language (or language stage in acquisition) compete with each other for the same input, thus they might mutually restrict their productivity. However, if one diminutive suffix is attached to a new word, this new diminutive formation opens up the potentiality of adding a rivalling suffix, thus establishing synergy between these suffixations. Therefore the competition of productive rules does not curb (in Bauer's 2001 terms) the availability of variants but their profitability in terms of type and especially token frequency.

Our approach is based on our two-level model of morphological productivity (Dressler et al. 2014) and on our ten-step scale of morphosemantic transparency/opacity of derivational morphology (Ransmayr et al. 2015), which both allow a new look on the conditions of rivalry and synergy, on synonymy vs. semantic differentiation of variants and on characterising actual vs. potential vs. impossible words and the role of lexical series.

We oppose morphosemantic and morphopragmatic meaning of evaluatives according to the model of Dressler & Merlini Barbaresi (1994, 2001**,** cf. Merlini Barbaresi 2015, Dressler & Korecky-Kröll 2015), unlike Potts's (2007) or Fortin's (2011).

Productive **Italian** diminutive formation (especially suffixes *–ino, -etto, -ello, -uccio,* see Merlini Barbaresi 2004) show both rivalry (thus mutual exclusions, semantic or pragmatic specialisation, terminological selection) and synergy, as apparent in the diminutivisation of recent loan words, which are usually first diminutivised with *–ino,* which then opens the base up for further diminutive formation, e.g. *vipp-ino/-etto/-ello/-uccio/-ar-ello, hacker-ino/-etto/-ello/-uccio*.

For **Austrian German** the most productive German suffixation in *-chen* and the most productive Austrian suffixation in *-erl* will be compared with corpus-linguistic methods. Synonymous diminutive variants are always possible in case of transparency degrees 1-2: 2 (usually restrictd to women and children): *Häub-chen = Haub-erl* 'bonnet-DIM', *Jäck-chen = Jack-erl* 'jacket-DIM', much less so for more opaque degrees.

Variants of diminutivised neologisms are much rarer in German than in Italian, e.g. *Projekt-chen = Projekt-erl, Computer-chen* and potential *Computer-l*, which appears to represent a very different degree of profitability rather than of productivity.

For **first language acquisition of German diminutives** (cf. Dressler & Korecky-Kröll 2015, Korecky-Kröll & Dressler 2007) diminutive competition and synergy (e.g. Bauch-i 'tummy' = Bauch-erl) is reanalysed in various phases of German first language acquisition, adding the new material from our INPUT project.

These results can be compared with published Italian acquisition facts (Noccetti et al. 2007): The most productive Italian suffixation in *-ino* emerges first. This does not impede the emergence of productive suffixation in *-etto* and *-ello* soon afterwards, although *-ino* maintains a default status at least up to the age of 3;0 and remains the most profitable in adult diminutive variation. Only at the beginning of diminutive acquisition there is no competition between different diminutives of the same basis. Then *libr-ino* 'book-DIM' occurs alongside *libr-etto*, *can-ino* 'dog-DIM' alongside normal irregular *cagn-ol-ino.* Such early diminutive doublets clearly contradict Clark's (1993) principle of contrast.

The comparison of Italian (with higher productivity and profitability of diminutive formation) and Austrian German (with its socio-regional variation between the two diminutive suffixes studied) allows new, corpus-linguistically based insights into issues of lexical blocking (Rainer 2012, 2014, Bauer et al. 2013: 568ff) because each of the two characteristics is liable to increase the number of diminutive doublets, at least in the

allowance of potential forms. Which one among potential words becomes an actual word is then largely a matter of the pragmatics of usage. The importance of pragmatics for diminutives may also explain why in the area of rule competition diminutives do not seem to have an intermediary position between inflection and prototypical derivational morphology (cf. Dressler et al. 2014).

Diminutives appear to allow many more variants and be much less subject to lexical blocking than other patterns of derivational or inflectional morphology. This can be explained by their basic and frequently preponderant pragmatic meaning (Dressler & Merlini Barbaresi 1994, 2001, Merlini Barbaresi 2015). Thus, words with pragmatic meaning are not excluded in rule competition by lexical blocking, which regards lexical reference. Since Italian has many more actual and potential diminutives than German and uses diminutives much more for pragmatic effects, this higher impact of morphopragmatics explains the otherwise surprising greater suffix rivalry of Italian. In this light also Malicka-Kleparska's (1985) Polish findings on unconditioned variation between Polish diminutive doublets should be reanalysed.

Bernard Fradin

Competition in derivation: What can we learn from doublets?

1. We are entitled to speak of competition in inflection when the same set of morpho-syntactic features can be correlated to more than one exponent e.g. chm *pört-na-lan* vs. *pört-lan-na* 'house {to[DAT], my}' (Crysmann & Bonami 2015) or equivalently when the cell of an inflectional paradigm is occupied by several exponents e.g. spa *pens-ases* vs. *pens-aras* think-SUBJV.PST.2SG. In derivation, a competition may arise when two lexemes (i) morphologically correlated to the same lexical base present distinct exponents (ii) while expressing exactly the same semantic content e.g. fra *camionn-ier* vs. *camionn-eur* 'truck-driver'. Competing forms give rise to situations that range from blocking (one form is dispreferred e.g. *camionnier*) to free variation (the past subjunctive in Spanish).

Doublets are conceived of as derived expressions which strictly satisfy properties (i) and (ii). More concretely, their bases are the same to the extent that (a) they have the same meaning and (b) they occur in the same construction or have the same distribution. In a parallel way, their derivational exponent has "the same semantic content" only if (c) the overall interpretation of the derived lexemes entails the same conclusions in similar contexts. These properties allow us to sort out the various cases of morphological competition. The talk addresses two issues: can we predict when doublets will appear? When they will, can we predict their meaning? The data are based on French nominalizations (NZNs) in *–age* / *–ment* collected from frWaC, TLFi and the Web.

2. (1) When conditions (a) and (b) are not fulfilled no doublets and no competition occur e.g. fra *émousser*[1] 'X remove moss from Y' and *émousser*[2] 'X make Y blunt'. These Vs are the same morphological V since they share the same inflectional paradigm but constitute distinct verbal lexemes, and their respective NZNs *émoussage* and *émoussement* have nothing in common. (2) All conditions but (b) are fulfilled e.g. *enterrer* 'X put Y under ground' i.e. 'to bury', which is correlated to *enterrement* and *enterrage*. These NZNs denote the same event, but the former is used for animates and the latter for plants. They do not constitute doublets and do not compete. (3) All conditions are fulfilled. In the simplest case, two NZNs with exactly the same meaning and distribution are correlated to the same V: we have true doublets occurring in free variation e.g. *pavage* / *pavement* 1) 'action of paving', 2) 'paved area'. But quite often the base V heads several constructions and the NZNs functioning as doublets are correlated to one of them only e.g. fra *recollage* / *recollement*[1] 'action of gluing Y again' facing *recollement*[2] MED 'action of cicatrizing'. (4) All conditions are fulfilled but one of the forms is dispreferred for prosodic reasons e.g. *??enneigeage* vs. *enneigement* 'quantity of fallen snow'.

3. Hypothesis (H1): the more the situation described by a base verb involves parameters, the reflect of which are constructions, the more doublets are likely to emerge, because complex constructions offer more room to express specific properties than those describing simple events. For instance, *encaver* 'to cellar' has one construction only 'X put Y[vine] into cellar' and denotes a very specific event whose participants and circumstances are fully determined. H1 rightly predicts that its NZNs are true doublets. On the contrary fra *perler* has several constructions: *perler*[1] 'X sew beads on Y'; *perler*[2] 'X[liquid] form beads (on Y)'. H1 predicts that the only true doublets will be NZNs correlated to constructions that are identical. This happens with *perlage*[1] / *perlement*[1] 'action of sewing beads on Y' vs. *perlement*[2] 'action of forming beads (of liquid)'.

4. Although many doublets are strictly equivalent e.g. fra *triplage* / *triplement* ‘action of tripling’, many others are not and the distribution of each exponent is not completely random. For example, when the situation denoted by the bse-V involves an agent or an identified external cause, *-age* is preferred; when no agent or cause is involved *-ment* is preferred e.g. fra *le roussissage des étoffes* ‘the dying red of fabrics’ vs. *le roussissement des feuilles* ‘the turning brown of leaves’. Insofar as constructions which are different describe situations which are different, the corresponding NZNs tend to have distinct meanings. I hypothetize (H2) that these NZNs form series that are organized in function of constraints bearing on their arguments and of the verbal constructions they are correlated with. The more frequent the correlation, the more entrenched the NZN and its meaning type will be (Fradin 2014).

Isidora Gatarić & Dušica Filipović-Đurđević

Processing effects of semantic ambiguity of Serbian derivational suffixes

In previous studies suffix ambiguity was operationalized as a phenomenon in which a given suffix (e.g. suffix *-er*) has two roles in one language – inflectional (syntactic; e.g. *higher*) and derivational (semantic; e.g. *worker*). Authors of these studies have shown that nouns with ambiguous suffixes are processed slower than those with an unambiguous suffix. This effect has been observed in Finish, Dutch and English (Bertram et al., 1999; 2000). However, in addition to being ambiguous by simultaneously having inflectional and deriational roles, a suffix can have multiple inflectional roles, as well as multiple derivational roles. The phenomenon of inflectional ambiguity of a suffix (syncretism) has been investigated as a separate linguistic phenomenon (e.g. Blevins, 1995). In theoretical study conducted in English, Plag (1998) proposed the concept of derivational suffix polysemy, but did not equate this phenomenon with suffixal ambiguity. However, to our best knowledge, the phenomenon of multiple derivational suffix roles (semantic ambiguity of derivational suffix) has not been investigated empirically.

The aim of this research was twofold: to operationalize semantic ambiguity of derivational suffixes, and to investigate cognitive processing of suffixal semantic ambiguity. We defined suffixal semantic ambiguity as a feature of suffix to have more than one meaning, regardless of whether they are related or not (e.g. suffix *-ica* can mark gender, diminutive and human features). The list of suffix meanings was created on the basis of linguistic literature on the word formation process in Serbian (Klajn, 2003; 2005). Pairs of two derived nouns (e.g. *anđelak-anđelče*) were presented to participants (N=46) in visal lexical decision task. Pairs were created of the same stem, but different suffixes: one noun in pair had ambiguous suffix (N=44), and the other one had unambiguous suffix (N=44). Nouns were selected from Frequency Dictionary of Contemporary Serbian Language (Kostić, 1999). Pairs of nouns were formed in order to control for the effects that could originate from various characteristics of a stem (e.g. morphological family size). Nouns were split into two experimental lists, and the Latin square design was used to ensure that each participant would see each stem and each suffix without repeating the stimuli. After controlling for the random effects of stimuli and participants, as well as for the fixed effects of several covariates (trial order, word length in letters/phonemes, lemma frequency, and suffix frequency) Linear Mixed-Effects Regression revealed significant effect of suffix ambiguity. Words containing semantically ambiguous suffixes were processed faster than those with unambiguous ones. The same analysis revealed significant effect of suffix frequency, but it's direction was dependent upon suffix ambiguity. Facilitatory effect of suffix frequency was observed for unambiguous suffixes, whereas inhibitory effect was observed for ambiguous suffixes. However, having in mind that this type of finding is reported for the first time, our next goal would be to evaluate this effect by presenting new set of stimuli, and to try to simulate the observed effects by using Naive Discriminative Learning procedure (Baayen et al., 2011).

References

Baayen, R. Harald, Petar Milin, Dušica Filipović-Đurđević, Peter Hendrix & Marco Marelli. 2011. An Amorphous Model for Morphological Processing in Visual Comprehension based on Naive Discriminative Learning. *Psychological Review* 118. 438-481.

Bertram, Raymond, Matti Laine & Katja Karvinen. 1999. The Intreplay of Word Formation Type, Affixal Homonymy, and Productivity in Lexical Processing: Evidence from Morphologically Rich Language. *Journal of Psycholinguistic Research* 28. 213-226.

Bertram, Raymond, Matti Laine, Harald R. Baayen, Robert Schreuder & Jukka Hyönä. 2000.

Affixal homonymy triggers full-form storage, even with inflected words, even in a morphologically rich language. *Cognition* 74. 13-25.
Blevins, James. 1995. Syncretism and Paradigmatic Opposition. *Linguistics and Philosophy* 18(2). 113-152.
Kostić, Đorđe (ed.). 1999. *Frekvencijski rečnik savremenog srpskog jezika*. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora i Laboratorija za eksperimentalnu psihologiju Filozofskog fakulteta u Beogradu.
Klajn, Ivan (ed.). 2003. *Tvorba reči u savremenom srpskom jeziku - sufiksacija i konverzija*. Beograd: Institut za srpski jezik SANU.
Klajn, Ivan (ed.). 2005. *Gramatika srpskog jezika*. Beograd: Zavod za izdavanje udžbenika.
Plag, Ingo. 1998. The polysemy of -ize derivatives: On the role of semantics in word formation. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1997*, 219-242. Dordrecht: Kluwer Academic Publishers.

Stefan Hartmann

Native and non-native alternatives:
Competition and reorganization in the German word-formation system

In German nominalization, the loan suffix *-ation* is quite frequent (cf. e.g. Fleischer & Barz 2012). In many cases, derivatives in *-ation* compete with formations in *-ierung*, e.g. *Isolierung* vs. *Isolation* 'isolation', *Realisierung* vs. *Realisation* 'realization'. In present-day German, there seems to be a division of labor between both suffixes: While combining the non-native stem ending in *-ier-* with the native suffix *-ung* tends to evoke a processual reading, its non-native competitor tends to yield resultative readings (cf. Duden 2011). For instance, *Isolierung*, if not used in a lexicalized reading ('insulation'), describes the process of isolating something (or someone). *Isolation*, by contrast, describes the state of being isolated.

But does the idea of this "division of labor" describe the current situation accurately, and if so, how did the present-day situation come about? The present paper addresses these questions with a quantitative corpus analysis drawing on multiple sources: First, the Early New High German period is investigated with the help of the Mainz Early New High German Corpus, a 400,000-word corpus covering 80 texts, balanced for five language areas and eight 30-year periods from the mid-sixteenth to the early 18th century. In this period, a wave of borrowings in *-ieren* entered the German language due to French influence (cf. Schmidt 2007), along with their French nominalized counterparts in *-ation*. At the same time, the non-native verbs began to combine with the highly productive native suffix *-ung*, arguably yielding *-ierung* as a suffix in its own right via reanalysis (cf. Haspelmath 2004).

Second, the GerManC corpus (Durrell et al. 2007), a 600,000-word corpus covering the time from 1650 to 1800 and balanced for five language areas, seven text types, and three 50-year periods, is used to investigate the early stages of the New High German period. Third, the developments until the early 20th century leading up to the present-day situation are analyzed drawing on data from the German Text Archive (Deutsches Textarchiv, DTA), a 100-million word corpus covering texts from the mid-seventeenth to the early 20th century.

The attestations of *-ierung* and *-ation* (as well as its variants, e.g. *-tion* in *Konsumption* 'consumption') in all three corpora are annotated for the aspectual construal evoked by the respective formations in context. Factors like text type and register are taken into account as well. In addition, data from synchronic web corpora are used to describe the present-day situation in more detail.

I will argue that the diachronic development of fully vs. partly non-native nominalizations in German provides a prime example of the reorganization of a word-formation system due to competition between functionally equivalent patterns. Drawing on theoretical notions and principles of Construction Morphology (e.g. Booij 2010), I will also explain why, from a constructionist perspective, non-native formations, which are often borrowed as a whole, can be seen as word-formation products, and I will argue that native and non-native word-formation in German do not constitute two different "systems" (*pace* Bergmann 1998), although the distinction between native and non-native word-formation patterns in German is still appropriate (cf. Kempf 2010). In the case of nominalizations in *-ierung* and *-ation*, both patterns form part of a complex constructional network that is diachronically reorganized to strike a balance between redundancy (or "degeneracy", cf. Van de Velde 2014) on the one hand and functional differentiation on the other.

References

Bergmann, Rolf. 1998. Autonomie und Isonomie der beiden Wortbildungssysteme im Deutschen. *Sprachwissenschaft* 23. 167–183.

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

Duden. 2011. Richtiges und gutes Deutsch: Das Wörterbuch der sprachlichen Zweifelsfälle. 7th ed. Mannheim: Bibliographisches Institut.

Durrell, Martin, Astrid Ensslin & Paul Bennett. 2007. The GerManC Project. *Sprache und Datenverarbeitung* 31. 71–80.

Fleischer, Wolfgang & Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache*, 4th edn. Berlin, New York: De Gruyter.

Haspelmath, Martin. 1995. The growth of affixes in morphological reanalysis. In: Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1994*. Dordrecht: Kluwer, 1-29.

Kempf, Luise. 2010. Warum die Unterscheidung fremd/nativ in der deutschen Wortbildung nicht obsolet ist. In Carmen Scherer & Anke Holler (eds.), *Strategien der Integration nicht-nativer Elemente und Strukturen*, 123–142. Berlin, New York: De Gruyter.

Schmidt, Wilhelm. 2007. *Geschichte der deutschen Sprache*: *Ein Lehrbuch für das germanistische Studium*, 10th edn. Stuttgart: Hirzel.

Van de Velde, Freek. 2014. Degeneracy: The Maintenance of Constructional Networks. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the Scope of Construction Grammar*, 141–180. Berlin, New York: De Gruyter.

Anja Hasse

Overabundance or conditioned allomorphy in Zurich German? A closer look

Zurich German exhibits competing dative forms of the indefinite article: feminine *ere*/*enere* and masculine/neuter *eme*/*emene*. It has been argued that the distribution of these forms is syntactically conditioned with the short forms *ere* and *eme* mostly used after prepositions (cf. Schobinger 2008, 30). This makes it a case of conditioned allomorphy. The data of the Swiss German Dialect Syntax Atlas (cf. Bucheli & Glaser 2002), however, reveals that this description is not appropriate. In the Zurich area, a lot of variation can be observed, not only between survey points but also at single survey points (cf. Hasse 2012). In order to collect the article forms, informants of the Swiss German Dialect Syntax Atlas were asked to translate four phrases from Standard German into their dialect (two with and two without a preposition) resulting in a data set of feminine and masculine articles each post-prepositional and in an NP. Some of the informants in the area of Zurich used the short forms *ere* and *eme* after the preposition but some used the long forms *enere* and *emene* instead, cf. (1) and (3). Likewise, some used short forms while others used long forms in the NP, cf. (2) and (4).

(1)	*mit*	***eme/emene***	*Maa*				‘with a man’
	with	**a-DAT.MASC**	man				
(2)	*Ich*	*ha=s*	***eme/emene***	*Maa*	*gää*.		‘I gave it to a man.’
	I	have-1.SG.PRS=it	**a-DAT.MASC**	man	give-PAST.PART.		
(3)	*mit*	***ere/enere***	*Frau*				‘with a woman’
	with	**a-DAT.FEM**	woman				
(4)	*Ich*	*ha=s*	***ere/enere***	*Frau*	*gää.*		‘I gave it to a woman.’
	I	have-1.SG.PRS=it	**a-DAT.FEM**	woman	give-PAST.PART.		

Thus, at the level of the dialect as a linguistic system, these varying article forms indicate Overabundance, where two forms are interchangeable within the same context (cf. Thornton 2011 *inter alia*). However, the question is if Overabundance also can be found at the level of individual speakers. We will therefore zoom in at individual survey points. Do all the informants who use short forms in a post-prepositional context use only long forms in the NP or the other way around? In this case, the single speakers of this dialect would show syntactically conditioned variation, although the pattern of which form to use would be variable between the speakers. According to that, this would rather be an instance of interpersonally distributed conditioned allomorphy. Alternatively, it might be that only some of these speakers have syntactically distributed forms while the others have overabundant article forms. This raises the question what makes speakers more prone to use overabundant forms. If we want to access the level of the inflectional system of speakers, additional data is needed: Only in-depth studies of the language use of single speakers can reveal if and to what extent Overabundance occurs. Furthermore, comparing different speakers can shed some light on how widespread Overabundance is within a speech community.

In my talk, I will present first results of such a study. I will discuss the morphological competition of the Zurich German article forms while comparing the inter- and intrapersonal variation and distinguishing between Overabundance and conditioned allomorphy. In doing so, I will present different data sets and methods to achieve a more fine-grained picture of Overabundance. Finally, I will argue that there is an additional parameter of determining Canonical Overabundance as defined by Thornton (2011). Overabundance is a phenomenon in the inflectional paradigm of a single speaker. Canonical overabundance is found with all speakers of a variety. The less speakers show it, the less canonical it is. The endpoint of this scale is Overabundance as a pure idiosyncratic feature.

References

Bucheli, Claudia & Elvira Glaser. 2002. The Syntactic Atlas of Swiss German Dialects: empirical and methodological problems. In Sjef Barbiers (ed.), *Electronic book on Syntactic Microvariation* http://www.meertens.knaw.nl/projecten/sand/synmic/. (27 June, 2015.)

Hasse, Anja. 2012. *Mit enem Mah oder mit anamu Ma. Variation des schweizerdeutschen Indefinitartikels im Dativ*. University of Zurich: Unpublished master thesis.

Schobinger, Viktor. 2008. *Züritüütsch. Säit me so oder andersch?* 5th ext. edn. Zurich: Schobinger-Verlag.

Thornton, Anna M. 2011. Overabundance (Multiple Forms Realizing the Same Cell): A Non-canonical Phenomenon in Italian Verb Morphology. In Martin Maiden et. al. (eds.), *Morphological Autonomy. Perspectives from Romance Inflectional Morphology*, 358-381. New York: Oxford University Press.

Nabil Hathout & Fiammetta Namer

A multi-level paradigm-based model of competition in word formation

In Word Formation (WF), competition occurs when there are "more than one form for the same morphosemantic complex so that rules may have multiple outputs" (Bauer 2014; cf. also Aronoff to appear). In French, for instance, prefixation in *anti-* applied on the base-noun *cancer* produces either *anticancer*$_{\text{ADJ}}$ or *anticancéreux*$_{\text{ADJ}}$ 'anticancer'. We present a WF model able to explain how several rival forms may emerge, with examples primarily taken from French.

Large corpora provide evidence that competition in WF often results in more than one winner: see Aronoff & Lindsay 2013 for cases of English suffix rivalry discovered via automatic Web search, such as *-ic* and *-ical* (*e.g.* both *historic* and *historical* derive from *history,* with the same meaning). So, competition aims to rank rival forms according to lexical and contextual criteria (Baayen & Renouf 1996; Plag 2006).

Traditional WF models are not well-equipped to predict and prioritize competing forms, and they do not explain how competition actually works. Derivation rules are binary ('all-or-nothing') devices, completely independent one from the others. They are designed to derive a unique derivative from a single base. In the absence of discriminant constraints, all possible rules apply uniformly so that the rival output forms are not ranked.

Yet, competition in WF raises many questions: what does explain the emergence of rival synonymous forms derived by the same rule from the same base? Why does the number of competing forms differ from one base-word to another, when the same rule applies? Is ranking of rival forms predictable? If it is, what mechanisms do account for differences in arbitration when selecting the preferred competitor?

Our answers will be illustrated by several examples, including the case of competing adjectives prefixed with *anti-:*

base	**meaning**	**realizations (Google frequency, june 26, 2015)**
cancer	against cancer	*anticancéreux* (400000), *anticancer* (260000), *anticancérigène* (8000)
fièvre 'fever'	against fever	*antipyrétique* (120000), *antifièvre* (10000), *antifébrile* (4000), *antifiévreux* (300), *antifébrigène* (1)
parlement 'parliament'	against parliament	*antiparlementaire* (10000), *antiparlement* (10000)
puce 'flea'	against fleas	*antipuce* (400000)

The analysis of this type of competition requires a WF decompartmentalization we achieve by adopting a paradigm-based approach (Štekauer 2014) characterized by two fundamental aspects.

The first is the extension to the semantic level of Bochner's (1993) paradigmatic approach of derivation and its cumulative patterns we call modules (see also Strnadova 2015). This extension gives our model new degrees of freedom, which allow to generate all the candidates.

The second aspect is the possibility of ordering candidates according to criteria combining a range of word properties and linguistic environments. Our model is set within a framework inspired by Optimality Theory (Ackema and Neeleman 2005), so that the decision making system, no longer binary, uses gradable constraints and operates within a paradigm-based framework. The ranking of

candidates results from a competition which involves both the individual properties of words, and the modules they belong to. For example, *anticancer* and *antifièvre* are optimal according to semantic transparency since they exhibit an exact match between the formal module {X, anti-X} and the semantic one {@, against(@)}. On the other hand, *anticancéreux* and *antipyrétique* are better candidates from the perspective of their integration in the category of adjectives, as *-eux* and *-ique* endings are productive adjectival suffixes. The importance speakers give to each constraint is reflected in the differences of the competitors' ranking from one base to another, as will be shown in the talk, where we will describe the paradigmatic structure of the model, develop the list of the constraints at work and the way they interact.

References

Ackema, Peter & Neeleman, Ad. 2005. Word-Formation in Optimality Theory. *Handbook of Word-Formation*. Springer. 285-314.

Aronoff, Mark. to appear, Competition and the Lexicon. In *Proceedings of the Annual Meeting of La Società di Linguistica Italiana 2013*.

Baayen, R. Harald & Renouf, Antoinette. 1996. Chronicling the Times: Productive Lexical Innovations in an English Newspaper. *Language* 72. 69-96.

Bauer, Laurie. 2014. Grammaticality, Acceptability, Possible Words and Large Corpora. *Morphology* 24. 83-103.

Bochner, Harry. 1993. *Simplicity in Generative Morphology*. Gruyter.

Lindsay, Mark & Aronoff, Mark. 2013. Natural Selection in Self-organizing Morphological Systems". *7th Décembrettes*, Lincom Europa. 133-153.

Plag, Ingo. 2006. Chapter 23. Productivity. *Handbook of English Linguistics*, Blackwell. 537-556.

Strnadová, Jana. 2015. Multiple Derivation in French Denominal Adjectives. *Carnets de Grammaire*. 22. 327-346.

Štekauer, Pavol. 2014. Derivational Paradigms. *The Oxford Handbook of Derivational Morphology*. In Lieber, Rochelle & Štekauer, Pavol (eds), Oxford University Press. 354-369.

Isa Hendrikx, Kristel Van Goethem & Fanny Meunier

Modeling "constructional transfer": A corpust based study of competing morphological and syntactic intensifiers in native and learner Dutch

In this contribution we will present a corpus-based comparison of the use of intensifying constructions in (written) native Dutch (*Corpus Hedendaags Nederlands*), Dutch by French-speaking learners (*Leerdercorpus Nederlands*) and native French (*Frantext*). The central focus will be on the competition between morphological and syntactic means to intensify adjectives. The analysis will be approached from the constructional perspective on language acquisition and multilingualism (cf. Tomasello 2003; Goldberg 2010; Höder 2012, 2014).

From such a usage-based point of view, second language acquisition is presumed to be more complex than L1 acquisition because of the competition between the specific constructions of the foreign language with the L1 constructions (Ellis & Cadierno 2009). In our study, we will examine the acquisition of Dutch intensifying constructions by French-speaking learners and will therefore compare the means to express intensity in both languages. In Dutch, adjectives can be intensified through adverbial modification [[ADV] [ADJ]]$_{AP}$ (e.g. *heel goed* 'very good'), prefixes [Pref [ADJ]]$_{ADJ}$ (e.g. *hypergevoelig* 'hypersensitive') and within adjectival 'elative' compounds [<N> [ADJ]]$_{ADJ}$ (e.g. *ijskoud* 'ice-cold') (Hoeksema 2012), the latter type being quite productive in Dutch. In French, intensification is more typically expressed through syntactic constructions, for example by adverbs (such as *très, tout* and adverbs in *-ment*: *absolument, totalement*) (Riegel e.a.1994: 620), or through reduplication [[ADJ] [ADJ]]$_{AP}$ (e.g. *rouge rouge* 'completely red') (Riegel e.a.1994: 622). Prefixation also occurs (e.g. *hypersympa* 'extremely nice'), but adjectival compounding is extremely uncommon as means of intensification, one exception being *ivre-mort* 'dead drunk'. In consequence, it can be assumed that French-speaking learners of Dutch will underuse the specific Germanic intensifying constructions (adjectival compounds) and, inversely, overuse syntactic intensifiers such as adverbs by what we call "constructional transfer".

In order to test this hypothesis, we have collected 21.200 adjectives from the *Corpus Hedendaags Nederlands* of which 308 are accompanied by an intensifier. In addition we have analyzed 498 intensified adjectives (out of a total of 8.454 adjectives) from the *Leerdercorpus Nederlands* written by French-speaking students in sixth grade of secondary school (Degand & Perrez 2014). These corpus data have been compared with a similar sample of 353 intensified adjectives (out of a total of 9.009 adjectives), drawn from recent essays in the *Frantext* corpus.

The findings of our data analysis confirm our hypothesis to a great extent. Learners clearly overuse adverbial intensification (96,6%) in comparison to native Dutch speakers (74,4%), probably influenced by their native language (similar proportion of intensifying adverbs found in *Frantext*). Furthermore, learners significantly underuse intensifying compounds in comparison to native speakers of Dutch (learners 1,8% vs Dutch natives 13,6%, while no attestations in the native French corpus were found), but also prefixes (learners 1,8% - again in line with the results for French natives (1,1%) vs Dutch natives 8,7%). In addition, the native Dutch corpus contains specific intensifying constructions that are not used by the learners, such as [*as* ADJ *as* X] (1,6%), although they do occur – marginally – in French (0,8%).

These results might be explained as transfer effects from French, as there appears to exist a cline between native Dutch, learner Dutch and native French. Besides the comparison of the construction types, we will also take a closer look at the extent of lexical variation and

productivity of the intensifiers in the different corpora (type/token ratio, hapax legomena), compare the most frequent intensifiers and intensifying collocations, and investigate the relation between intensifier type and semantic type of the adjective (scalar, limit, extreme) (cf. Paradis 1997, 2001).

Finally, the different corpus results will be confronted with Höder's diasystemic approach to multilingualism (Höder 2012, 2014). Concretely this theory would imply that a bilingual person has a unified semantic-cognitive representation of intensification to his disposition with a variety of formal means (morphological and syntactic constructions) to be specified for the different languages.

References

Corpus Hedendaags Nederlands. 2013. Instituut voor Nederlandse Lexicologie (ILN) 800.000 texts (1814-2013). Available online at Common Language Resources and Technology Infrastructure (CLARIN). https://portal.clarin.inl.nl/search/page/search

Degand, Liesbeth & Julien Perrez. 2004. Causale connectieven in het leerdercorpus Nederlands. *Tijdschrift n/f 4*. 115-128.

Ellis, Nick. & Teresa Cadierno. 2009. Constructing a Second Language. *Introduction to the special section. Annual Review of Cognitive Linguistics* 7, 111-139.

FRANTEXT, ATILF. 2014. CNRS & Université de Lorraine. 271.599.218 words. http://www.frantext.fr.

Goldberg, Adele. 2010 [2006]. *Constructions at Work. The nature of generalization in language.* Oxford: Oxford University Press.

Hoeksema, Jack. 2012. Elative compounds in Dutch: Properties and developments. In Oebel (ed.) 2012, 97-142.

Höder, Steffen. 2012. Multilingual constructions: A diasystematic approach to common structures. In *Multilingual Individuals and Multilingual Societies* [Hamburg studies on multilingualism 13], K. Braunmüller & C. Gabriel (ed). 241-257. Amsterdam: John Benjamins.

Höder, Steffen. 2014. Constructing diasystems: grammatical organisation in bilingual groups. In *The Sociolinguistics of Grammar* [Studies in language companion series 154], T. Afarli & B. Maehlum (ed). 137-152. Amsterdam: John Benjamins.

Paradis, Carita. 1997. *Degree modifiers of adjectives in spoken British English.* Lund: Lund University Press.

Paradis, Carita. 2001. Adjectives and boundedness. In *Cognitive Linguistics* 12-1. 47-65.

Quirk, Randolph, Sidney Greenbaum, Geoffry Leech, Jan Svartvik. 1997. *A comprehensive grammar of the English Language*. London: Longman.

Riegel, Martin, Jean-Christophe Pellat, René Rioul. 1994. *Grammaire méthodique du français*. Presses Universitaires de France: Paris.

Tomasello, Michael. 2003. *Constructing a language.* Boston: Harvard University Press.

Fabiola Henri, Gregory Stump & Delphine Tribout

Conversion relations and the morphological complexity of Mauritian, a French-based creole

Creolistic research persistently asserts the simplicity of creoles, citing as evidence the claimed simplicity of creole morphology (Vinson 1989, McWhorter 2001, Good 2012, Bakker 2014). Yet, creoles not only exhibit morphology, but evince a surprising degree of morphological complexity. Drawing on the unexpected evidence of conversion relations, we provide new support for this conclusion.

According to Manova & Dressler (2005), conversion relations among nouns and verbs may exhibit whole-word identity (Eng. $walk_V$ / $walk_N$) or stem identity (Fr. $danser_V$ / $danse_N$). One might imagine three degrees of conversion complexity:

	Complexity	Where L is a lexeme belonging to category A: A→B conversion
(i)	least	is invariably based on the same word/stem form of L
(ii)	↕	varies predictably with respect to the form of L on which it is based
(iii)	most	varies unpredictably with respect to the form of L on which it is based

French shows conversions of type (iii): V→N conversions may be based on three distinct verb stems, whose choice is determined neither by phonological nor grammatical nor semantic criteria (Tribout 2012). If unpredictability is a measure of morphological complexity (Ackerman & Malouf 2013, Stump & Finkel 2013) then the unpredictability of French conversions contributes to the morphological complexity of French.

We show that in Mauritian, conversions are as complex as in French: V→N conversions may involve different verb forms and the choice of verb form is unpredictable. Mauritian verbs have two forms —short and long (see Table 1)— whose distribution is morphomically rather than morphosyntactically conditioned (Henri 2010):

Verb	Short Form (SF)	Long Form (LF)
DANSE 'to dance'	*dans*	*danse*
PINTIRE 'to paint'	*pintir*	*pintire*
MARINE 'to marinate'	*marinn*	*marine*
BWAR 'to drink'	*bwar*	*bwar*

Table 1. Verb form alternations in Mauritian

V→N conversion in Mauritian may select either a verb's SF or its LF, as in Table 2. Overall, converted nouns have the same kinds of meanings whether they arise from a verb's LF or its SF; thus, form selection is not, in itself, predictable. Moreover, converted nouns are not simply inherited from the lexifier language: because nouns like $danse_N$ or $sante_N$ do not come from French, conversion is clearly a productive process in Mauritian. As they are based on verb-form selections that are both variable and unpredictable, V→N conversions in Mauritian are comparable in complexity to those of French.

	Verb		→ Noun	
	LF	SF		
'to dance'	*danse*	*dans*	*danse* *(la)dans*	'dancing; ball' 'dance'
'to sing'	*sante*	*sant*	*sante*	'song'
'to paint'	*pintire*	*pintir*	*lapintir*	'paint'
'to drink'	*bwar*		*bwar*	'drink'

Table 2. V→N conversions in Mauritian

Other French-based creoles (Table 3) exhibit similar patterns, contradicting the widespread claim of creole simplicity and revealing a recurrent fallacy in creolistic research—the tacit assumption that creoles exhibit complexity only to the extent that they manifest it in the same way as their lexifiers. As our evidence shows, the verb paradigms of French-based creoles are different from those of French; accordingly, verbs in French-based creoles participate in conversion relations that fail to parallel those of French verbs. Yet, objectively, these relations are fully as complex as those of French verbs.

	Verb		→ Noun	
	LF	SF		
'to dig'	*fouye*	*fouy*	*lafouy*	'search, dig'
'to come'	*vini*	*vin*	*vini*	'arrival'
'to beg shamelessly'	*chyente*	*chyen*	*chyen*	'beggar'
'to come'	*vini*	*vin*	*vini*	'arrival'

Table 3. V→N conversions in Guadeloupean and Haitian

References

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The Low Conditional Entropy Conjecture. *Language*.

Bakker, Peter. 2014. Creolistics: Back to square one? *Journal of Pidgin and Creole Languages.*

Good, Jeff. 2012. Typologizing grammatical complexities, or: Why creoles may be paradigmatically simple but syntagmatically average. *Journal of Pidgin and Creole Languages*.

Henri, Fabiola. 2010. *A constraint-based approach to verbal constructions in Mauritian*. Thèse de Doctorat: University of Mauritius and Université Paris Diderot.

McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology.*

Manova, Stela & Wolfgang Dressler. 2005. The morphological technique of conversion in the inflecting-fusional type. In Bauer & Valera (eds.) *Approaches to conversion/zero-derivation*. Waxmann.

Stump, Gregory & Raphael Finkel. 2013. *Morphological typology*. Cambridge University Press.

Tribout, Delphine. 2012. Verbal stem space and verb to noun conversion in French. *Word Structure.*

Vinson, Julien. 1889. Créoles. In Berthillon *et al*. (ed.), *Dictionnaires des sciences anthropologiques*. Doin.

Eric Hoekstra & Arjen Versloot

Frequency, similarity and the blocking principle

The blocking principle was defined by Aronoff (1976: 43ff, cf. Di Sciullo & Williams, 1987:10ff a.o.) as "the non-occurrence of one form due to the simple existence of another". It may be used to describe not only token blocking but also type blocking (Rainer 1988). Token blocking involves an irregular lexical form which blocks the regular attachment of a derivational suffix. For example, German **Mutigkeit* 'litt. courageness' is blocked by *Mut* 'courage'. Type blocking involves the blocking of one derivational suffix by another (German **Grotteskität* is type-blocked by *Grotteskheit*). In our paper, we will analyse an example of type blocking in a bilingual setting, involving Frisian, a minority language, and Dutch, the dominant language. Our data indicate that blocking should be viewed as a tendency. In addition, we will show that blocking within Frisian does not distinguish between the two languages, so that the frequency of word forms in Dutch affects the choice of suffix in Frisian, depending on the extent to which the Frisian word and its Dutch semantic equivalent are similar in form.

Frisian has two suffixes, "-ens" and "-heid", which are used to form nouns from adjectives. They often target the same base words and thus exhibit considerable competition (J. Hoekstra, 1990, 1998; Van der Meer, 1986, 1987, 1988; E. Hoekstra & Hut, 2003). Examples of the derivational pairs are given below:

dúdlik	'clear'	*dúdlikens* ~ dúdlikheid	'clarity'"
freedsum	'peaceful'	*freedsumens* ~ freedsumheid	'peacefulness'"
warber	'industrious'	*warberens* ~ warberheid	'industriousness'"
stom	'stupid'	*stommens* ~ stomheid	'stupidity'"

Some base words exclusively take the suffix "-ens", others exclusively take "-heid", and many base words are used with both suffixes, but mostly not with the same frequency. Semantic differences between the two suffixes were indirectly controlled for by excluding plurals en diminutives.

The morphological database of the Frisian Language Corpus (period 1980-2000) was used to establish the frequency of a large number of nouns, either ending in "-ens" or in "-heid" or both. The raw data thus obtained comprise a list of slightly more than 700 nouns. All tests and conclusions rely on that dataset. Levinshtein distance with PMI (Point Mutual Information) was used to measure the degree of similarity between a given Frisian adjectival base and its Dutch semantic equivalent.

As to blocking, 47% of all items with an item frequency > 10 has exactly either 100% or 0% tokens with "-ens". Conversely, 53% of all items violates blocking conceived of as a principle. However, it turns out that the relative distribution of "-ens" versus "-heid" per item shows a bimodal distribution, testifying to the operation of blocking as a tendency.

In addition, we will argue that the strength of the attraction between a Dutch item and its Frisian counterpart can be given by a formula relating frequency and similarity, as follows:

A ~ 2log(frequency)/LevenshteinDistance

The effect of the Dutch mental lexicon upon Frisian confirms findings of psycholinguists who have argued that forms which are similar in form or meaning are represented closely together in the brain (e.g. Dijkstra 2003). Thus similarity in form and meaning has been used by proxy to measure representational distance. Frequency has been used by proxy to measure the neural strength of a word's mental representation. The blocking effect is thus shown to constitute a tendency which derives in part from the formula given above.

References

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

Di Sciullo, A.-M. & E. Williams (1987) *On the definition of word.* Cambridge, Massachusetts: MIT Press.

Dijkstra, A. (2003). "Lexical storage and retrieval in bilinguals". In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (eds) *The interface between syntax and the lexicon in second language acquisition.* 129-150. Amsterdam: John Benjamins.

Hoekstra, J. (1990) "Adjectiefnominalisatie in het Fries". *Interdisciplinair Tijdschrift voor Taal- & Tekstwetenschap* 9:4, 273-285.

Meer, G. van der (1986) "De achterheaksels ENS en HEID yn it Frysk". *Us Wurk* 35, 108-130.

Meer, G. van der (1987a) "Friese afleidingen op ENS en HEID. Een geval van morfologische rivaliteit?" *Spektator* 17, 360-367.

Meer, G. van der (1987b) "Nominaliseringen op ENS (>NIS) en HEID (in het Fries en elders)". *Taal en Tongval* 39, 22-36.

Rainer F. (1988) "Towards a theory of blocking: The case of Italian and German quality nouns". In: G. Booij & J. van Marle (eds) *Yearbook of Morphology 1988*, 155-185. Dordrecht: Foris.

Holden Härtl

Compounds take the short way to the lexicon – A pragmatic approach to novel word-for-mation products

Word-formation products have often been claimed to be particularly prone to function as names for concepts and, as such, be stored in the mental lexicon, cf., among others, Downing (1977). In this talk, we will argue that the naming inclination of newly formed compounds is rooted in the compounds' deviation from grammatical default, that is, from the phrasal expression. Thereby, based on linguistic "markedness", a novelty effect is produced, which leads to a re-interpretation of the compound as a kind name.

We define naming as a function that establishes a lexical concept for a sub-kind. With this in mind, we examine the behavior of newly created adjective-noun (A-N) compounds in German, e.g., *Rotdach*, *Hochlampe*, in comparison to their phrasal counterparts, i.e. *rotes Dach*, *hohe Lampe*. We will start from a lexicalist stance and prove the assumption to be correct that word-formation is the preferred route for constituting a concept's name in German. Evidence will be presented from kind-sensitive environments as well as name-selecting predicates like *sogenannt* ('so-called'), with which compounds prove to be better compatible than their phrasal counterparts: *ein sogenanntes* $^{??}$*rotes Dach* / *Rotdach*. Empirical evidence for the functional difference between compounds and phrases comes from a corpus study, which reveals a systematic correlation between the use of name-indicating devices – we probed *sogenannt* as well as the rate of quotation marks – and the type of the (lexicalized) A-N expression.

Crucially, the naming inclination of A-N compounds also shows up in combinations that have elsewhere been argued to be semantically equivalent to their phrasal counterparts, as in the case of compounds containing Latin loan adjectives, see Schlücker & Hüning (2009). Consider, however, the following examples, which challenge this "equivalence assumption":

(1) a. Der Bentley hat ein optimales Design, fast schon ein Optimaldesign.
b. $^{??}$Der Bentley hat ein Optimaldesign, fast schon ein optimales Design.

We do not argue that *Optimaldesign* is interpreted to be somehow "more optimal" than *optimales Design*. Instead, we attribute the acceptability contrast in (1) to the more pronounced naming function of *Optimaldesign* as compared to *optimales Design* and, specifically, to the fact that constructions containing particles like *fast* ('almost') give rise to the implication of a degree scale. *Fast X* signifies that some property *X* of the subject referent is not fully attained on the scale in question and that its "lower" complement still holds: *almost X* → *not X*, see Rotstein & Winter (2004). Importantly, the scale involved in the felicitous example in (1a) is associated with an increase in a category match, where the expression introduced by *fast* in the second conjunct corresponds to the "stronger" category, i.e. a kind name. Viewed in this light, the functional difference between the A-N phrase and the A-N compound in (1a) can be explained based on pragmatic reasoning and, in particular, a scalar implicature: The choice of the weaker expression (*optimales Design*) indicates that, on the respective scale, a stronger classification (*Optimaldesign*) does not apply in the context of the utterance. Thus, the contrast in (1) suggests that, while their meanings are denotationally equivalent, A-N phrase and A-N compound differ in that the latter represents the stronger candidate for a kind name.

From this pragmatic angle, we argue for a systematic relation to hold between the markedness of novel A-N compounds and their interpretation as kind name as well as their affinity to be lexicalized. This relationship will be traced back, along the lines of Levinson's M-principle (see Levinson 2000), to a principle holding that deviation from a default form – here: from the phrasal expression – implies departure from the meaning of this form, thus triggering a re-interpretation as kind name and producing semantic specialization. Semantic specialization will

be argued to be effective “right from the beginning” in A-N compounds and, consequently, we will implement interpretation as kind name and semantic specialization in compounds as the cause for potential lexicalization and not as its effect, as is often proposed in the literature.

References

Downing, Pamela. 1977. On the Creation and Use of English Compound Nouns. *Language* 53-4. 810-842.

Levinson, Steven C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Schlücker, Barbara & Matthias Hüning. 2009. Compounds and Phrases: A Functional Comparison between German A+N Compounds and Corresponding Phrases. *Italian Journal of Linguistics* 21-1. 209-234.

Rotstein, Carmen & Yoad Winter. 2004. Total Adjectives vs. Partial Adjectives: Scale Structure and Higher-Order Modifiers. *Natural Language Semantics* 12-3. 259-288.

Tamara T. Irsa, Wolfgang U. Dressler & Gary Libben

Competition and cooperation between top-down and bottom-up processing of compounds

***Background**:* Psycholinguistic studies of compound processing have found that both whole-word properties as well as the semantic and morphological characteristics of compound constituents play a role in how compounds are processed in online word recognition, which suggests a dynamic interplay between bottom-up and top-down processing. Although there has been a notable trend in recent years toward considering word processing within sentence context, and not only single word processing, the effect of preceding text/discourse has not yet been investigated. This is problematic, because we rarely encounter words in isolation or within a single sentence, but within a text or conversation and research in text linguistics has shown that any text chunk creates expectations about what is to follow.

Aims: The goal of our study is to increase ecological validity by starting to investigate how top-down processing (from text to word) and bottom up processing (from constituent to word and from word to text) cooperate and compete in online and offline processing of noun-noun compounds.

Materials: In order to achieve this goal, we focused on minimal, coherent texts, which were composed of German two-sentence sequences such as the following (in translation below):

- NEUTRAL: She walked down to the cabin. The waterfall roared in the distance.
- SEMANTIC: She walked down to the river.The waterfall roared in the distance.
- LEXICAL: She walked down to the water. The waterfall roared in the distance.

All sequences consist of two sentences with six words each. The second sentence of each triplet is identical and contains a *target compound* as the second word after a definite article. The first sentences are also identical, except for the last word, which is either neutral, i.e. not related to the target compound of the second sentence (a), semantically related to one of the target compound's constituents (b), or lexically identical to one of the compound members (c). With these sets, we also investigated the role of semantic transparency by employing target compounds that are either fully transparent (*waterfall*), transparent-opaque (*jailbird*), or opaque-transparent (*strawberry*). For coherence reasons, the antecedent lexically or semantically related word is always linked to a transparent target-compound constituent. Thus, the transparent-transparent compounds are split into two subtypes according to whether the first or second constituent is primed.

***Methods**:* Twenty native German speakers were tested at Vienna University. Participants were asked to read short "stories" like the sequences shown above, in a self-paced reading task, using a bar press to uncover successive words of each sentence displayed on a computer screen. Bar pressing times were recorded for each word and constituted the primary data for analysis. After reading each sentence sequence, participants were asked to rate intuitively on a

scale, how much sense the story made to them. The stimulus materials consisted of 64 triplets (192 sentence pairs) with incoherent pairs interspersed among the coherent core sequences as a control. The experiment was conducted in three blocks with a distraction task in-between.

Results and Conclusions: We found that first-sentence differences affected the perceived coherence between the sentences: As expected, sequences were rated as most coherent for the lexical condition, less coherent for the semantic condition, and as least coherent for the neutral condition. The nonsense sequences were rated as incoherent. In processing, word reading times in the second sentence, starting with the target compound, were significantly reduced in the lexical priming condition compared to both the semantic and the neutral condition. This points to a local lexical overlap between the antecedent autonomous word and the second sentence, meaning that there is a constituent priming effect not only on the target compound but also on the processing of the whole second sentence, an effect of inter-sentence top-down processing. In addition, the study enabled us to track the word-by-word development of text coherence by probing the point at which the overall coherence ratings of the participant for a particular sequence begin to play a role in bar pressing times of the second sentence with a mixed linear effects model. We found that effects begin at the verb (third word) of the second sentence and then persist throughout the sentence. Finally, using such minimal, coherent texts lead to increased ecological validity and, thereby, also to different results than with traditional word-by-word constituent priming. That is, we found that only lexical priming lead to significantly reduced reaction times for the target compound, whereas there were no significant differences between the semantic and neutral condition.

Brian D. Joseph

Resolving morphological competition in contact situations

Competition at the morphological level can be viewed in systemic terms, looking at, for instance, what factors induce a given form that marks category X to occur when another form can potentially fill the same slot (as with different nominalization suffixes in English, and cf. Thornton 2012 on morphological overabundance), or when two morphs can potentially fill the same slot but mark different categories (as in a subset of Georgian agreement marking (Anderson 1986)). The resolution of such system-internal competition can be very revealing as to the structure of a proper morphological theory (as argued in Stewart 2001 regarding Georgian and Paradigm Function Morphology).

But what about the competition that arises between variant forms of different systems, i.e. different dialects or different languages, when they come into contact with one another? What happens in such socially induced morphological competition?

In this paper, I examine that very question through the lens of contact between and among various languages in the Balkans, with particular attention to the different kinds of resolutions seen when aspects of the morphology of these languages are seen to compete with one another for the attention of speakers, as it were. I argue here that it is not just a matter of one form or the other winning out in this socially induced competition, but rather I focus on one particular kind of outcome where elements from both competing forms are adopted. I refer to this outcome as "hybridization" and I explore the reasons behind such a dualistic outcome.

My data comes from contact between Greek and Albanian and between Turkish and several Balkan languages (Albanian, Slavic, and Greek). Of specific concern are two types of morphological mixed outcomes. First are those in which the lexical form of morphemes is partially reconfigured, yielding a hybrid form, due to the availability of resources from two separate language systems, such as southern Albania Greek *mehanikós* 'mechanic, engineer', a hybrid of expected Greek *mihanikós* and Albanian *mehaník*. The second involves blended affixes, with, e.g. both a Turkish ending indicating pluralization and a native ending, as in Albanian *efendi-ler-ë* 'gentlemen' with both Turkish *–ler* and native Albanian *–ë* marking plural, a phenomenon found also in South Slavic (Bulgarian, Macedonian, Serbian) with the same Turkish suffix.

Some of these outcomes seem to show morphological *pleonasm* (in the sense of Gardani 2015) via language contact, a possibility envisioned by Gardani but only sparsely documented. Moreover, as such, they could be interpreted as a type of morphological *overabundance*, understood, though in a somewhat different sense from the use of that notion in, e.g., Thornton 2012. It is argued here, though, that at least some instances involving Turkish plurals in the Balkans are pleonasm only from an etymological perspective, since there is some overt evidence of reanalysis of NOUN-*lar* as a stem: in Macedonian, for example, the ostensible plural *agalar* 'Turkish lords' can be the basis for the singular diminutive *agalarče,* 'O you young *aga*'. However, indisputable contact-induced pleonasm is evident in a different contact environment involving Greek, in Cappadocian Greek mixed Greek-Turkish verb endings (Janse 2009).

While some of these contact-induced pleonastic hybrid outcomes seem to be motivated initially at least by accidental phonological similarity, as Janse suggests for the Cappadocian cases, and others by opacity, since foreign elements are by definition almost opaque in the borrowing system, still others, it is argued, may be tied to extra-systemic factors such as an

ideological claiming of a form as indigenous by marking it with a native affix. The contact-related morphological competition discussed here, therefore, shows the complexity of factors — system-internal and system-external, structural and social, functional and ideological — that play a role in the resolution of such competition.

References

Anderson, Stephen. 1986. Disjunctive ordering in inflectional morphology. *Natural Language and Linguistic Theory* 4. 1-32.

Gardani, Francesco. 2015. Affix pleonasm. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer (eds.) *Word-Formation: An International Handbook of the Languages of Europe* Volume 1 (HSK 40.1), 537-550. Berlin: Mouton de Gruyter.

Janse, Mark. 2009. Watkins' Law and the development of agglutinative inflections in Asia Minor Greek. *Journal of Greek Linguistics* 9. 32-48.

Stewart, Jr., Thomas W. 2001. Georgian agreement without extrinsic ordering. In Michael Daniels, David Dowty, Anna Feldman, and Vanessa Metcalf (eds.) *Varia. Ohio State University Working Papers in Linguistics No. 56*, 107-134. Columbus: Department of Linguistics, The Ohio State University.

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

Nikolay Khakimov

'Katholik+en' and 'Muslim+*y*': Competition in plural marking of German nouns in bilingual Russian sentences

As attested in numerous corpora of bilingual speech, lone code-mixed nouns receive plural markers either from the matrix language (1) or the embedded language (2):

(1) osobenno vot *türk*-i; oni že *muslim*-y
especially PTCL Turk-PL 3PL PTCL Muslim-PL
'Especially these Turks, they are Muslims.'

(2) naprimer poljak-i oni vse *katholik-en nä,*
for.example Pole-PL 3PL all Catholic-PL PTCL
'For example, Poles, they are all Catholics.'

A third possibility, i.e. double marking, has only rarely been observed and is not attested in the Russian-German corpus under analysis. The competition of the matrix-language and embedded-language plural markers in online production has often been discussed in the literature (e.g. Backus 1996; Muhamedowa 2006; Myers-Scotton 1993, 2002), but there still remains the need to underpin the factors determining this competition, especially in a situation of contact between two fusional languages, German and Russian.

The question addressed in this paper is whether pluralised German nouns retain the plural morphology of German, i.e. the embedded language, as in (2), or receive plural markers from the Russian matrix language, as in (1). The paper examines the following factors as determinants of selecting the plural marker on inserted German noun stems: their phonological shape, the morphological case of the slot in which they are inserted, the frequency distribution of the plurals and singulars of the inserted German lexical items in German.

The data for the analysis stem from a bilingual Russian-German corpus that comprises naturally occurring conversations and informal group interviews of 20 Russian-speaking young adults living in Germany. The frequencies with which the realized German nouns are used as singulars and plurals in German were obtained from the German deWaC corpus (Baroni and Kilgariff 2006). The competition between these forms was calculated by the plural-singular ratio. A regression model model was utilized to investigate an interplay of the examined factors in determining the language of the plural marker. The final model had a high predictive capacity.

The investigation revealed three main findings: (1) The frequency with which a noun plural occurs in the embedded language (German) determines the language of the plural marker on code-mixed nouns. That is, lexical items commonly used as plurals tend to be selected as holistic units and inserted into the matrix clause retaining their German plural markers. (2) German lexical items with an accented vowel in the stem-final position cannot take Russian inflectional suffixes directly as the Russian declensional system depends on stems with consonants in the final position. If the inserted German noun has an accented vowel stem-finally, either a compromise strategy is employed, such as the use of epenthetic consonants, or German plural forms are produced. (3) When the matrix structure projects a non-core case (dative, genitive, instrumental, or prepositional) on the slot in which a German lexical item is inserted, the tendency is towards using Russian inflections, fusing plural and case. However, German nouns retain their German plural markers in slots requiring one of the core cases, nominative or accusative, owing to a structural equivalence between German

plural inflections and the Russian inflections of the nominative and accusative case *-i* (*-y*).

In sum, the paper provides compelling evidence for the frequency effect adumbrated earlier by Backus (2003) and demonstrates the role of phonological and morphosyntactic regularities in determining the outcome of the competition between the contact languages for marking plural on code-mixed nouns.

References

Backus, Ad. 1996. *Two in one: Bilingual speech of Turkish immigrants in The Netherlands*. Tilburg: Tilburg University Press.

———. 2003. Units in code switching: Evidence for multimorphemic elements in the lexicon. *Linguistics* 41 (1). 83–132.

Baroni, Marco, and Adam Kilgariff. 2006. Large linguistically-processed web corpora for multiple languages. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006). Trento, Italy.*, 87 – 90. Morristown, NJ: Association for Computational Linguistics.

Muhamedowa, Raihan. 2006. *Untersuchung zum kasachisch-russischen Code-mixing (mit Ausblicken auf den uigurisch-russischen Sprachkontakt)*. München: Lincom Europa.

Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford, England; New York: Clarendon Press; Oxford University Press.

———. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes.* Oxford: Oxford University Press.

Laila Kjærbæk & Hans Basbøll

Competition in Danish noun plural acquisition

This study focuses on competition between different morphemes in relation to Danish noun plural (PL) acquisition.

In Danish, noun PL can be formed in four different ways, departing from the singular (SG) form, by:

- adding a PL suffix (including zero suffix, Ø) (e.g. *bil* [biːˀl] 'car' – *bil-er* [ˈbiːˀlɐ] 'car-s');
- changing a phoneme of the stem (e.g. *mand* [manˀ] 'man' – *mænd* [mɛnˀ] 'men');
- combining suffix addition and stem change (e.g. *fod* [foðˀ] 'foot' – *fødder* [ˈføðˀɐ] 'feet');
- changing or adding nothing, viz. PL = SG (e.g. *mus* [muːˀs] 'mouse' – *mus* [muːˀs] 'mice').

The Danish PL markers thus consist of a combination of suffix addition (incl. Ø) and stem change (incl. no change). We consider each PL marker a single PL morpheme (although there are, of course, other logically possible interpretations). The Danish noun PL system consists of 23 competing PL markers (Basbøll et al. 2011).

For productivity, we use a three graded scale proposed by Kjærbæk et al. (2014):

1) FULLY PRODUCTIVE PL markers involve addition of the ɐ-suffix without phonemic stem change;
2) SEMI-PRODUCTIVE PL markers involve addition of the ə-suffix or Ø-suffix without phonemic stem change;
3) UNPRODUCTIVE PL markers involve phonemic stem change or addition of one of the foreign PL suffixes /s/, /a/ and /i/.

The ɐ-suffix occurs as a PL suffix in 87 % of all Danish nouns (Basbøll et al. 2011). Since the ɐ-suffix is by far the most frequent PL suffix in Danish, we will investigate which nouns do NOT take the ɐ-suffix, and why. Furthermore, we will investigate Danish children's production of noun PL and present a detailed analysis of their produced PL error forms. We will propose answers to the following questions:

1) *What types of nouns do not take a fully productive PL marker?*
2) *Why do these nouns not take a fully productive PL marker?*
3) *Which parallels can be drawn between the answers to question 1 and 2 and Danish children's production of noun PL (error) forms?*

Question 1 and 2 will be answered based on the description of the Danish noun PL system from a sound perspective presented in Basbøll et al. (2011). Question 3 will be answered based on empirical data from typically developing monolingual Danish children: 1) 160 children between the ages of 3-10 years who participated in a picture-based elicitation task inspired by Berko (1958); 2) 80 children in the ages of 3-9 years who participated in a structured interview. The children's produced forms are coded phonologically with regard to both stem and suffix, and we are therefore able to compare each child's produced PL form with the adult standard pronunciation. That way we can investigate which principles Danish children use in order to solve the problem of competing PL markers when producing a PL form.

In the talk, we set out to investigate the principles followed by Danish children when they are to select a PL marker among several competing PL markers. The children's error forms are particularly relevant to tackle this issue. Earlier studies show that overgeneralization errors are characterized by going from less productive towards more productive plural markers (e.g. Laaha et al. 2006), and we expected to see the same pattern in Danish. But we didn't. In the structured interviews 47 % of all error forms went from a FULLY PRODUCTIVE to a SEMI-PRODUCTIVE PL marker, 19 % from a SEMI-PRODUCTIVE to another SEMI-PRODUCTIVE PL marker – only 20 % went

from an UNPRODUCTIVE or SEMI-PRODUCTIVE to a FULLY PRODUCTIVE PL marker (Kjærbæk & Basbøll, submitted). We will present similar results for the picture based elicitation task.

We will take into account stem transparency, suffix predictability, gender and stem final phonology, and we will include other data material in order to shed light on 'pure zeroes' (PL = SG) – both methodologically and theoretically.

References

Basbøll et al. (2011). The Danish noun plural landscape. *Acta Linguistica Hafniensia: International Journal of Linguistics* 43, 81-105.

Berko (1958). The child's learning of English morphology. *Word* 14, 150-177.

Kjærbæk & Basbøll (submitted). Testing hypotheses on frequency effects in first language acquisition – noun plural inflection in Danish children. *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015*.

Kjærbæk et al. (2014). Sound structure and input frequency impact on noun plural acquisition: hypotheses tested on Danish children across different data types. *Nordic Journal of Linguistics* 37, 47-86.

Laaha et al. (2006). Early noun plurals in German: regularity, productivity or default? *Journal of Child Language* 33, 271-302.

Nick Kloehn

Finding productive morphology in the wild: A quantitive analysis of novel morpheme use in Swahili and Swahili-English code switching on Twitter

The aim of this study is to use new data from Swahili and Swahili-English code-switching to challenge theories of morpheme productivity that argue for a relationship between frequency and morphological decomposition (Bybee, 1995, Hay 2002, Hay & Baayen 2002). Given that these theories have investigated morphological productivity only in Indo-European languages, which themselves contain relatively small amounts of morpheme-to-word ratios, this study investigates this same relationship in a language with a higher morpheme to word ratio in order (2.55 morphemes per word for Swahili versus 1.68 morphemes per word for Modern English (Greenberg 1959)). When a language has a larger amount transparent morphological structure with a high amount of combinatorial variation, will the lexicon likewise contain a higher numbers of individual morphemes? Or is there an upper bound in the number of morphemes that are productive? Is such productivity predictable from frequency?

In order to test this question, derived and underived frequencies of word forms containing 10 inflectional and 10 derivational morphemes will be evaluated using the Helsinki Swahili Corpus (HCS) containing 13.1 million tokens gathered from literature and news sources. Examples of the tokens containing inflectional and derivational morphemes (underlined) are given in (1):

(1) a. **u- -pig-an -aji**
CL11-*hit*-RECIP-AGENT
'rivalry'

b. **u- -pig-an -o**
CL11-*hit*-RECIP-INST
'contest'

c. **m- -chez-aji**
CL1-*play*-AGENT
'player'

d. **m- -chez -o**
CL3-*play*-INST
'game'

(Mohammed 2001, TUKI 2001)

The ratio of these forms with their component morphemes, and the base lacking them will be graphed and compared for the degree to which these forms are predicted for productivity using a regression analysis of each morpheme's relative frequencies (Hay & Baayen 2002). The distance of the r^2 line (fitting the underived and derived frequencies) to the X=Y line per graph will be used as the predictor of the degree of productivity of each morpheme.

Morphological productivity will be evaluated using the novel technique of searching two corpora automatically generated from Twitter data. Via supervised machine learning, the corpora have

been created from tweets written by thousands of users from East Africa. Each corpus with around 5 million tokens will be used to find the degree to which morphemes co-occur with English words in Corpus 1, and the degree to which these morphemes occur in forms in Corpus 2 that are not found in the HCS corpus.

Examples of such combinations are in (2) where the underlined forms are English forms, and bold forms represent the morphemes with which they occur:

(1) a. 'Enda **u-** -tweet … **U-** **-na** **-ni-** -suffocate.'
IMP.*go* CL11-*tweet* ... 2SG PRES 1SG.OBJ *suffocate*
'Go and tweet … you are suffocating me.'

b. 'Ni- me- kumbuka **ku-** stuck …'
1SG PERF *remember* INF stuck
'I remember getting stuck…'

If morphological borrowing and novel formation is determined by morphological productivity, and likewise if frequency is the largest factor for productivity, then morphemes predicted to have higher degrees of productivity should occur in higher numbers in each corpus. However, if the relationship is non-significant, then the causal chain should break down at either point. Either productivity does not determine morphological borrowing and novel forms, or rather other factors such as phonological and semantic plausibility, blocking by non-complex forms, and general language change may have a large influence on morpheme productivity in Swahili.

Although such corpora are noisy, their volatility represent an ideal ground for gathering larger quantities of data on morphological productivity. Such an analysis will add to both to the growing number of methods useful to the linguist, and will crucially inform theories of morphological productivity providing either further empirical coverage, or challenging their findings.

Sources

Bybee, J. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5), 425-455.

Hay, J. 2002. From speech perception to morphology: Affix ordering revisited. *Language*, 78(3), 527-555.

Hay, J., & Baayen, H. 2002. Parsing and productivity. *Yearbook of Morphology 2001* (pp. 203-235). Springer Netherlands.

Mohamed, Mohamed Abdulla. 2001. *Modern Swahili Grammar*. East African Publishers.

Seidl, A., & Dimitriadis, A. 2003. Statives and reciprocal morphology in Swahili. *Typologie des langues d'Afrique et universaux de la grammaire*, 1.

TUKI - English - Swahili Dictionary - Kamusi Ya Kiingerza - Swahili. 2001. TUKI - The Institute of Swahili Research at the University of Dar es Salaam. Laurier Books Ltd. /AES; ISBN: 9976911297.

Katharina Korecky-Kröll

High competition in German comparatives and superlatives: Results of an online grammaticality judgment experiment

German adjective gradation is an interesting phenomenon with respect to the question of variation and competition: Comparatives and superlatives are nearly always formed synthetically from their positive forms via suffixation, and some of them also undergo a stem vowel change (umlaut), either obligatorily or optionally.

1) As to suffixation, comparatives attach an *-er* (= a-schwa) suffix. Bisyllabic adjectives ending in unstressed *-el* and *-er* (e.g. positive *eitel* 'vain') undergo deletion of the e-schwa in comparative (e.g. *eitler* 'vainer'), but not in superlative (e.g. *am eitelsten* 'the vainest'). Superlatives may either have only an *-st* suffix (e.g. *am kleinsten* 'the smallest') or show an *e* (= e schwa) epenthesis after *d, t, sk* and sibilants before the *-st* suffix (e.g. *am stolzesten* 'the proudest'). Nevertheless, for adjectives ending in shibilants and diphthongs, both *-st* and *-est* superlatives are equally correct (e.g. *am frischsten* or *am frischesten* 'the freshest', *am genausten* or *am genauesten* 'the most exact'). However, this *e* epenthesis is not very salient and most speakers will not be aware of the regularities mentioned above or the existence of competing forms, but just tend to use one form or the other.

2) Some adjectives change their stem vowel in comparative and superlative, which is not only a salient phenomenon, but also a rich source of variation and competition: Sometimes both forms just coexist (e.g. positive *fromm* 'pious'– comparative *frommer* or *frömmer* 'more pious'– *am frommsten* or *am frömmsten* 'most pious'), whereas in other cases there is only one correct form. But there are even adjectives that rhyme with each other, in which one forms the comparative and superlative via umlaut, whereas the other one does not: e.g. positive *hart* 'hard' – comparative *härter* 'harder' – superlative *am härtesten* 'the hardest' vs. positive *zart* 'tender' – comparative *zarter* 'more tender' – superlative *am zartesten* 'most tender'.

Overall, these systems of competition are opaque and difficult not only for L2 learners of German, but also for young children acquiring German as their L1 (Fürst et al. 2011, Hohaus et al. 2014).

Furthermore, comparatives and especially superlatives have very low token frequencies in spoken language (Zeldes 2011), which makes them an interesting testing ground not only for L2 learners and children, but also for adult native speakers, who also show uncertainties in their use of comparatives and superlatives, an issue that has not yet been researched in detail.

This contribution aims at investigating how 121 young adult native speakers of German (all of them university students) judge equal numbers of correct and erroneous comparatives and superlatives presented in random order if they are asked to answer as quickly as possible.

The results of the online grammaticality judgment experiment are consistent with previous studies (e.g. Fürst et al. 2011) showing that comparatives are easier than superlatives: Comparatives are judged faster and are much more likely to be correct than superlatives.

Overall, correct comparatives containing a non-umlautable stem vowel and thus having no competing forms (e.g. *tiefer* 'deeper') are judged fastest and with the highest accuracies, whereas correct comparatives containing an optional umlaut (e.g. *nässer* 'wetter') are accepted more slowly. Likewise, among correct superlatives, forms with non-umlautable stem vowels and without *e* epenthesis (e.g. *am tiefsten* 'the deepest') are judged fastest, whereas forms containing an optional umlaut and an *e* epenthesis (e.g. *am nässesten* 'the wettest') are accepted most slowly.

On the other hand, incorrect superlatives that contain an incorrect non-salient *e* epenthesis (e.g. **am starresten* instead of *am starrsten* 'the most rigid') are rejected most slowly, whereas superlatives containing a salient incorrect opacifying umlaut (e.g. **am sättesten* instead of *am sattesten* 'the fullest') are rejected faster. Furthermore, non-salient incorrect *e* epenthesis is rejected more slowly than incorrectly missing *e* epenthesis.

The results will be discussed in the framework of different morphological models dealing with competition (such as single route vs. dual route models, the Competition Model, Natural Morphology, Construction Morphology, ...), demonstrating that German adjective gradation is indeed a challenge even for adult native speakers with equally high levels of education.

References

Fürst, Bettina, Sabine Laaha, Katharina Korecky-Kröll, Gregor Laaha & Wolfgang U. Dressler. 2011. On the acquisition of adjective gradation by Austrian typically developing (TD) and SLI children. Poster, XIIth IASCL, Montreal, 21.07.2011.

Hohaus, Vera, Sonja Tiemann & Sigrid Beck. 2014. Acquisition of comparison constructions. *Language Acquisition* 21. 215–249.

Zeldes, Amir 2011. On the productivity and variability of the slots in German comparative correlative constructions. In Marek Konopka, Jacqueline Kubczak, Christian Mair, František Štícha & Ulrich H. Waßner (eds.), *Grammatik und Korpora 2009. Mannheim, 22. –24.09.2009*. 429–449. Tübingen: Narr.

Elena Kulinich, Phaedra Royle & Daniel Valois

Overregularization as competition

Morphological overregularization (OR) in child language gives rise to the coexistence of two different forms at some given period of language development. The change from irregular to regularized, and from regularized back to irregular forms is known as U-shaped development curve (Marcus *et al.* 1992). This was argued to be the (momentary) result of a failure of Blocking, which – after a short initial period where the irregular form is acquired simply via lexical learning – prevents an irregular form from falling into the regular pattern. As a result, the general rule overrides the formation of a specific usage (i.e., causing OR). For instance, English children may first use the past irregular form *broke* for the verb *to break*, then overregularize by applying the productive past tense *–ed* suffix: *braked* [brejk-t], and then return to *broke* [browk] when Blocking takes effect (Marcus *et al.* 1992).

The phenomenon is very well known but there is still a relative lack of quantitative and cross-linguistic data in support of it, and some questions remain about how different forms are distributed across verbs and across children, what the proportion of them is found at different periods of language development and how children finally arrive to the "correct" idiosyncratic usage.

We review this problem by looking at data from an elicitation task with 76 Russian-speaking children aged 3 to 4 years (M = 3;4, SD = 3.6) on present verb forms. In Russian, morphological OR errors often arise from applying the so-called "j-correlation" model (Ceytlin 2009), which has been argued to be the default pattern (Chernigovskaya & Gor 2000). This pattern relates past and non-past stems of the most productive Russian verb class as, for example, in *zna-t'/ znaj-u* 'to know'. That is, Russian children produce *plakaj-u* [plakaju] instead of *plach-u* [platʃu] (from the verb *plakat'* 'to cry') or *risovaj-u* [risovaju] instead of *risuj-u* [risuju] (from *risovat'* 'to draw'). 12 verbs of three subregular verb classes were used as stimuli: 6 verbs with suffixation, as in *ris-ova-t'*/*ris-uj-u* 'to draw' (Class 2 from Zalizniak's 2003 classification) or as in *da-va-t'*/*da-Øj-u* 'to give' (Class 13) and 6 verbs of Class 6 with morphonological alternation, as in *plaka-t'/plach-u* 'to cry'.

Results show that subregular and overregularized forms are indeed coexistent in the child speech from age 3 to 4. Non-target forms comprise 33% of all responses, and 15.35% of all productions (i.e., 46.5% of all non-target responses) are ORs in yod /j/. However, there is great variability across subjects and verbs. For instance, some children do not overregularize subregular forms at all (they varied from 0 % to 50 % or their productions).

Our results are not entirely consistent with Gor and Chernigovskaya's (2003) study in which the amount of ORs for 4-year-old Russian-speaking children is much greater than what we uncovered. More data are obviously needed to draw a clearer overall picture, but the evidence from Russian confirms that Blocking does not work in the same way in children and adults. More precisely, there is a period of coexistence of specific irregular and overregularized forms in children. With time, OR forms are driven out by specific subregular forms, but the mechanisms that help children to recover from OR errors (e.g., direct or indirect negative evidence, lexicalization processes) remain to be specified.

References:

Ceytlin, Stella N. 2009. *Ocherki po slovoobrazovaniju i formoobrazovaniju v detskoj rechi*. Moskva: Znak (Studia Philologica).

Chernigovskaya, Tatyana & Kira Gor. 2000. The complexity of paradigm and input frequencies in native and second language verbal processing: evidence from Russian. In E. Wande and T. Chernigovskaya (eds.) *Language and Language Behavior*. 20-37.

Gor, Kira & Tatyana Chernigovskaya. 2003. Generation of complex verbal morphology in first and second language acquisition: evidence from Russian. *Proceedings of the 19th Scandinavian Conference of Linguistics* 31 (6): Supplement. 819-33.

Marcus, Gary F., Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, Fei Xu & Harald Clahsen.1992. Overregularization in language acquisition. In. *Monographs of the Society for Research inChild Development* 57 (4, serial No. 228). 126.

Zalizniak, Andrej A. 2003. *Grammatičeskij slovar' russkogo jazyka*. Moskva: Russkie slovari.

Lior Laks

Hebrew agent nouns: Competing patterns and doublets formation

This talk examines the formation of Hebrew agent nouns (ANs) and their relation to the verbal system. I examine the criteria for selecting morphological patterns as well as cases of variation. I argue that pattern selection is to a great extent predictable based on morphological relations between the verb and the derived AN. In addition, semantic transparency in compound formation triggers morphological change where ANs take an additional form.

Competing patterns

Hebrew word formation relies highly on non-concatenative morphology (Berman 1978, Bolozky 1978, Schwarzwald 1981, Ravid 1990, Bat-El 1994, 2011), where words are formed in patterns. The study addresses the process of selecting pattern for verb-derived ANs (1).

(1)

Verb		**AN pattern**	**Examples**	
a. nihel	'managed'	meCaCeC	menahel	'manager'
b. ciyer	'painted'	CaCaC	cayar	'painter'
c. sixek	'played'	CaCCan	saxkan	'player'

All ANs in (1) are related to *CiCeC* verbs but are formed in different patterns. *menahel* 'manager' is formed in the participle pattern *meCaCeC* which is also the present form of the verb ('manages'). The others patterns *CaCaC* and *CaCCan* are only nominal. The selection of one pattern and not another is to some extent arbitrary. It is unclear for example, why the *saxkan* (1c) 'player' is not formed in *CaCaC* (**saxak*) or in *meCaCeC* (**mesaxek*, which only means 'plays').

However, examining the distribution of ANs reveals strong tendencies. ANs that are based on *hiCCiC* verbs are systematically formed in the participle pattern *maCCiC*. For example, the AN of *hilxin* 'composed' is formed in *maCCiC* (*malxin* 'composer') and not in *CaCaC* (**laxan*) or *CaCCan* (**laxnan*). Why is it so? I claim that pattern selection is triggered by morphological transparency between ANs and verbs. The transition from the verb to the participle form is more transparent, as it requires fewer changes. The participle formation (2a) requires only substitution of the prefix *hi-* with *ma-*, while the stem remains intact. In contrast, formation in any of the other patterns would break the consonant cluster *lx*, in addition to other changes in the base, thereby making the AN structurally less related to the verb (2b, 2c). This provides support to the claim that word formation is based on stem modification rather than the extraction of the consonantal root (Bat-El 1994, Ussishkin 2005). The morphological mechanism examines the structural properties of the verb and selects AN patterns accordingly.

(2) a. hi**lx**in → ma**lx**in
b. hi**lx**in → ***lax**an
c. hi**lx**in → ***lax**nan

When the related verb is in *CaCaC* or *CiCeC*, the transition to the none-participle patterns is simpler and requires fewer changes. For example, the formation of *cayar* based on *ciyer* (1b) requires only changing the stem vowels. The transition from the verb to the AN is more transparent and therefore there are more options for pattern selection.

Doublets formation

Because AN formation in participle patterns is morphologically more transparent, new ANs formation relies highly on these patterns, while other patterns gradually become less productive (Bolozky 1999). In addition exiting ANs that are heads of compounds demonstrate morphological variation, where the head receives an additional form of the participle. Examine *rapad* (*CaCaC*) 'upholsterer', related to the verb *riped* 'upholster'. This AN does not have a participle doublet *(meraped*), which only means 'upholsters'. However, when it is the head of the compound *rapad-safot* ('sofas-upholsterer', lit. upholsterer-sofas) it alternates with *meraped* (*meraped-safot*). Both forms exist and can be used interchangeably. More examples are presented below.

(3)

Non-participle		**Participle**			**Corresponding verb**	
sadran- sxora	CaCCan	**mesader-** sxora	meCaCeC	'stocker' (lit. arranger-merchandise'	**sider**	'arranged'
rašam-ʔamutot	CaCaC	**rošem-** ʔamutot	CoCeC	'registrar associations'	**rašam**	'registered'

Why does doublet formation occur? I argue that the presence of both the agent itself and the patient in the same NP makes the AN more semantically transparent in the sense that it corresponds to the thematic roles that the related verb assigns. Many studies pointed out the thematic relations between elements of compounds whose head is verb-derived. Theta-role relations inside such compounds are shown to be related to the properties of the argument structure of the related verb (for Hebrew, see Rappaport-Hovav&Levin 1992, Ritter, 1991, Clark & Berman 1987, Siloni 1997, Ravid 2006, Borer 2013, among others). Semantic transparency triggers a morphological change where the AN head takes additional form that is more related to the corresponding verb.

The study highlights the importance of morphological and semantic transparency in word formation and its implications for morphological competition. It enables to shed more light on the motivation for morphological changes both from morphological and semantic perspectives.

References

Bat-El, Outi. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory* 12: 572-596.

Bat-El, Outi. 2011. Semitic templates. In Mark van Oostendrop, Colin Ewen, Elizabeth Hume, and Karen Rice (eds.), *Blackwell Companion to Phonology*. Malden, MA: Wiley-Blackwell. 2586-2608.

Berman, Ruth A. 1978. *Modern Hebrew Structure*. Tel-Aviv: University Publishing Projects.

Bolozky, Shmuel. 1978. Word formation strategies in Modern Hebrew verb system: denominative Verbs. *Afroasiatic Linguistics* 5: 1-26.

Bolozky, Shmuel. 1999. *Measuring productivity in word formation: the case of Israeli Hebrew*. Leiden: Brill.

Borer, Hagit. 2013. *Taking Form*. *Structuring Sense Vol. II.* Oxford: Oxford University Press.

Clark, Eve V. & Ruth A. Berman. 1987. Types of linguistic knowledge: Interpreting and producing compound nouns. *Journal of Child Language* 14: 547-568.

Rappaport Hovav, Malka & Beth Levin. 1992. *–er* nominals: Implications for the theory of argument structure. *Syntax and Semantics* 26: 127-153.

Ravid, Dorit. 1990. Internal structure constraints on new-word formation devices in Modern Hebrew. *Folia Linguistica* 24: 289-347.

Ravid Dorit. 2006. Word-level morphology: A psycholinguistic perspective on linear formation in Hebrew nominals. *Morphology* 16: 127-148.

Ritter, Elizabeth. 1991.Two functional categories in noun phrases: Evidence from Modern Hebrew. *Syntax and Semantic* 25: 37-62.

Schwarzwald, Ora R. 1981. *Grammar and reality in the Hebrew verb.* Ramat Gan: Bar Ilan University Press. (in Hebrew)

Siloni, Tal. 1997. *Noun phrases and nominalizations: The syntax of DPs*. Dordrecht: Kluwer Academic Publishers.

Ussishkin, Adam. 1999. The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output-output correspondence. *Phonology* 16: 401-442.

Lior Laks and Evan-Gary Cohen

Competing vowels and the lexicon/syntax distinction

This talk examines vowel selection in Hebrew in the formation of adjectives and passive verbs with medial glottals. The vowel precedong the glottal consonant colloquially alternates between *o* and *u*, e.g. *men**o**hal~men**u**hal* 'is managed'. We argue that vowel selection results from the distinction between adjectives and passive verbs and, more generally, from the component of the grammar deriving them, lexicon vs. syntax.

The status of glottal consonants in Modern Hebrew is different from their status in earlier periods. Glottals used to trigger various phonological processes, like vowel lowering. However, this effect has become less stable (Faust 2005 and references therein). We examine the *meCuCaC* pattern with medial glottals. *meCuCaC* is used for two categories (Berman 1987, Doron 2003, 2013, Meltzer-Asscher 2010, among others): (i) present forms of passive verbs, e.g. *mesupaʁ* 'is told'; (ii) adjectives, e.g. and *metsujan* 'excellent'. In some cases, the same form can be ambiguous, e.g. *mesudaʁ* 'is tidied up'/'tidy'. In case of medial glottals, the vowel preceding itis lowered to *o,* e.g. *menuhal→menohal* 'is managed'. However, there are cases where such forms are still pronounced with *u*. This study examines speakers' vowel selection in both lexical categories

Experiment
18 participants (19-41) read aloud 24 sentences with *meCuCaC*: 6 verbs, 6 adjectives and 12 with 6 ambiguous forms, each form used as an adjective and as a verb (1).

(1) a. *ha-makom **menohal** al-jedej dina* (only verbal)
'The place **is managed** by Dina'
b. h*aja jeled **mevohal** fe- baxa* (only adjectival)
'There was a **frightened** boy who was crying'
c. *ha-beged **megohats** al-jedej moti* (ambiguous – verbal)
'The clothing (item) **is being ironed** by Moti'
d. *tenase limtso eze beged **megohats*** (ambiguous – adjectival)
'Try to find some **ironed** piece of clothing'

Results: Participants generally showed a preference for *o*, which was used in 66% of the sentences. Comparing the two categories, participants demonstrated a stronger tendency to use *u* in verbs. The breakdown of the results (2), with respect to ambiguous and unambiguous forms sharpens the distinction between adjectives and verbs. In forms that are only verbs, participants demonstrated the highest tendency to use *u* (55%), as opposed to forms that are exclusively adjectival (12%). This distinction is less pronounced in ambiguous forms, where the vowel *u* was used on average in 41% of cases of a verbal meaning, in comparison to only 27% of the cases with an adjectival meaning.

(2)

Vowel Category	*u*	*o*
Verbs	55%	45%
Ambiguous form: Verbs	41%	59%
Ambiguous form: Adjectives	27%	73%
Adjectives	12%	88%

What determines vowel selection?

We argue that the difference between adjectives and passive verbs, results from the component of the grammar deriving them, lexicon vs. syntax. Following Reinhart&Siloni (2005) and Horvath&Siloni (2008), we assume Hebrew passivization applies in the syntax, while the derivation of adjectives applies in the lexicon. Assuming that passive verbs are not lexically stored, their formation applies online and is automatic. Thus, speakers demonstrate a tendency to select *u* also for passive verbs with glottals conforming to the general *u-a* passive pattern.

The picture is different with respect to adjectives. In contrast to passive formation, the morphological process does not apply every time adjectives are derived. Due to the effect of glottals, some adjectives were formed with *o* and stored this way. In addition, the morphology of adjectives is less predictable than the morphology of passive verbs. Adjectives are formed in different templates with different vocalic patterns and therefore there is less motivation to form them under the *u-a* pattern. With respect ambiguous forms, the difference between the adjectival and the verbal forms is less pronounced but still lies within the lines of our argument.

This study shows that the component of the grammar deriving words has an impact on the vowel selection. Despite the strong tendency to preserve the vowel *o*, speakers demonstrate a tendency to select *u* in different degrees. The motivation for selecting *u* depends on the locus of derivation. The picture that emerges is that some vowels are more typical of the lexicon, while others are more typical of the syntax. This has direct implications on the nature of morpho-phonology with respect to the division of labor between the syntax and the lexicon. The study provides insights to a case of competition in the grammar where several types of criteria interact. It provides a general distinction between vowels associated with the lexicon and the syntax, highlighting the strong correlation between form, meaning and the relevant component of the grammar in word formation processes.

References

Berman, R.A. 1987. Productivity in the lexicon: New-word formation in Modern Hebrew. *Folia Linguistica* 21:425-461.

Doron, E. 2003. Agency and Voice: the Semantics of the Semitic Templates. *Natural Language Semantics* 11: 1-67.

Doron, E. 2013. The interaction of adjectival passive and voice in H. Borer, A. Alexiadou and F. Schaeffer (eds.) *The Roots of Syntax and the Syntax of Roots*, Oxford University Press.

Faust, N. 2005. *The Fate of Hebrew Gutturals*. MA Thesis. Tel-Aviv: Tel-Aviv University.

Horvath, J. and 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*, pp. 105-134. Amsterdam/Philadelphia: John Benjamins.

Meltzer-Asscher, A. 2010. Present participles: Categorial classification and derivation . *Lingua* 120:2211-2239.

Reinhart, T. and Siloni, T. 2005. The Lexicon-Syntax Parameter: Reflexivization and other Arity Operations. *Linguistic Inquiry* 36(3): 389-439.

Michele Loporcaro

Noun inflection and gender in Romanian

While formal (i.e. phonological or morphological) rules for gender assignment exist throughout Romance (cf. e.g. Thornton 2009), in no other language is noun inflection so reliable a predictor of gender as in Romanian. Based on this undisputable fact, some recent analyses (Bateman and Polinsky 2010, Maiden 2013, 2015) take one more step and make Romanian even more special, claiming that gender, in this language, is a property of word forms, rather than of the lexeme as a whole (or, on a somewhat different, but converging line, a property specified on a number head: Giurgea 2014). This has consequences for the analysis of the gender system, as proponents of this view support a two-gender analysis, under which Romanian has a binary system (masculine vs feminine), rather than a three-way contrast (masculine vs feminine vs neuter), as traditionally assumed both in descriptive Romanian grammar (e.g. Graur 1928) and in theoretical linguistics/linguistic typology (e.g. Jakobson 1971: 187–89, Corbett 1991: 150–53, Carstairs-McCarthy 1994: 750–52, Aikhenvald 2000: 45–6, Acquaviva 2008: 135–40, among many others).

In this paper, I argue that a three-gender analysis is the only one compatible with a set of standard assumptions on gender in cross-linguistic typological studies and in theoretical morphology (cf., first and foremost, Corbett 1991; 2012): namely, with the distinction a) between target and controller gender; b) between overt and covert gender; c) between gender and inflectional class, both understood as (features ascribed to) paradigmatic classes of lexemes, rather than (features ascribed to) specific word forms. One reviewer objected that 'this is not an argument, because the assumption that all languages work like the Latin or Russian prototype may be wrong'. Of course, I am *not* assuming this: what I am saying is that to assess whether they are alike or not as for (different aspects of) feature *x*, it is crucial that we work with a cross-linguistic definition of the feature, rather than adjust it, depending on the language-specific data to be analysed, since in the latter case we will be talking under the same heading of what is actually different things.

Back to Romanian, I also show that a three-gender analysis, unlike two-gender analyses, is compatible with (and supported by) the results of psycholinguistic studies on lexical access, which provide evidence (aphasiological, and/or from the tip-of-the-tongue effect: cf. e.g. Badecker et al. 1995, Vigliocco et al. 1997, and the review by Heim 2008, covering also neuroimaging-based studies) showing that words are stored in the mental

lexicon, in multi-layered representations, and that the morphosyntactic gender specification can be accessed independently from the phonological (and, hence, morphological) representation.

References

Acquaviva, Paolo (2008). *Lexical Plurals*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.

Badecker, William, Miozzo, Michele, and Zanuttini, Raffaella (1995). The two-stage model of lexical retrieval: evidence from a case of anomia with selective preservation of grammatical gender. *Cognition* 57: 193–216.

Bateman, Nicoleta, and Polinsky, Maria (2010). Romanian as a Two-Gender Language. In Donna Gerdts, John Moore, and Maria Polinsky (eds), *Hypothesis A – Hypothesis B. Linguistic explorations in honor of David M. Perlmutter*. Cambridge, Mass.: MIT Press, 41–77.

Carstairs-McCarthy, Andrew (1994). Inflection classes, gender, and the principle of contrast. *Language* 70: 737–88.

Corbett, Greville G. (1991). *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. (2012). *Features*. Cambridge: Cambridge University Press.

Giurgea, Ion (2014). Possible syntactic implementations of the controller vs. target gender distinction: the view from ambigenerics. In Tania Paciaroni, Anna M. Thornton and Michele Loporcaro (eds), *Exploring grammatical gender*, Special issue of *Language Sciences* 43(1): 47–61.

Graur, Alexander (1928). Les substantifs neutres en roumain. *Romania* 26: 249–60.

Heim, Stefan (2008). Syntactic gender processing in the human brain: A review and a model. *Brain and Language* 106: 55–64.

Jakobson, Roman (1971). On the Rumanian neuter. In Roman Jakobson, *Selected Writings. II. World and language*. The Hague: Mouton, 187–89.

Maiden, Martin (2013). *On the rise of genus alternans. Evidence from Romance languages*. Paper given at the International Conference on Historical Linguistics, Oslo, 2013

Maiden, Martin (2015). *The Romanian Alternating Gender in Diachrony and Synchrony*. Ms. University of Oxford

Thornton, Anna M. (2009). Constraining gender assignment rules. *Language Sciences* 31: 14–32.

Vigliocco, Gabriella, Antonini, Tiziana, and Garrett, Merrill F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science* 8: 314–17.

Varvara Magomedova

Diminutive suffixes in modern Russian

In this paper, we look at the distribution of four Russian masculine diminutive suffixes, considered to be allomorphs: *-ok, -ek, -ik* and – *t͡ɕik,* where *–ok* and *–ek* are considered to be variants of the same suffix with a vowel alternation (Polivanova, 1967). The distribution of these suffixes can be usually predicted by the stem final segment, the stress pattern of a noun and several other factors that have less significant impact (Gouskova et al., 2015).

The phonological preferences of suffixes are listed in Table 1.

Table 1. Phonological properties of nouns selected by each suffix.

suffix	stem final consonant[1]	stress position	other
-ok/-ek	velar	initial	no hiatus[2]
-ik	fricative	final, floating	
- t͡ɕik	nasal, lateral		no stem final consonant clusters

These suffixes cause the following changes to the stem they attach to:

1) *-ok*– stem-final velar consonant mutation and a stress shift (*p'orax → para∫'ok*);
2) *-ek* - stem-final velar mutation (*ar'ex → ar'eʃək);*
3) *-ik*– palatalization, may cause stem-final velar mutation[3] (*nos →n'os^{j}ik*);
4) *-t͡ɕik* – does not cause any changes of the stem(*bat'on →bat'ont͡ɕik*).

These suffixes' distribution was analyzed in the recent paper of Maria Gouskova, Sofia Kasyanenko and Luiza Newlin-Lukovicz (Gouskova et al., 2015) where they have mostly considered standard Russian nouns. However, in novel nouns much more variation occurs.

We have studied nouns that do not have established diminutive forms yet (novel, newly loaned nouns and native nouns that are rarely used in diminutive) using both web search and experiment, concentrating mostly on nouns that would have to take the *–ok* suffix.

We have searched 14 novel loanwords and 10 native words with stem-final velars rarely used in the diminutive form[4]. Additionally, we have performed three short online experiments, one to establish the difference in the meaning between *–ok* and other suffixes (80 participants, 10 items, nonce nouns only) and two other to test the factors that are supposed to trigger *–ok* (60 participants,

[1] these suffixes only attach to nouns ending in consonants
[2] this point is arguable
[3] As the *–ik* suffix is not supposed to attach to velars, we cannot know for sure. However, speakers often mutate stem-final velar segment when attaching this suffix to novel words.
[4] for the search we have used the search tool Lingui-Pingui that helps to filter out irrelevant results and doesn't count duplicates (Magomedova,2013)

20 items, 10 real, 10 nonce nouns). We have also did a paper-based survey (60 participants, 32 items, 16 real, 16 nonce noun, 16 loan or loan-like, 16 native or native-like).

Our data have shown a different suffixes' distribution from the one Gouskova, Kasyanenko and Newlin-Lukovicz got on their data. There are several differences:

1) all of three suffixes are not allomorphs anymore, the *–ok* suffix has a pejorative meaning that other suffixes do not have;
2) the *–ok* suffix loses productivity in a neutral context ;
3) *–ok* and *–ek* are not variants of the same suffix anymore, their distribution is not complementary[5];
4) the *-t͡ɕik* suffix appears with any kind of phonological context.

Our experimental data also show that the *-t͡ɕik* suffix is used significantly more often with newly loaned words when attaching to nouns that are supposed to take *–ok* (stem-final velars, initial stress). We suppose that this kind of the distribution is due to the promoted faithfulness constraints on the novel and newly loaned words: a suffix that causes most changes to the stem loses its productivity and the one that does not cause any changes to the stem is promoted. To test this hypothesis we put the experimental data into a Maxent model (Hayes et al, 2009) where we used markedness constraints (such as "no *–ik* after velars") that define the suffixes' distribution in standard Russian and faithfulness constraints (such as Id(place)) to forbid changes to the stem caused by suffixes *–ok* and *–ik.* The weights of faithfulness constraints we got appear to be as much (or almost as much) as of markedness ones.

References

Gouskova, Maria, Sofya Kasyanenko, and Luiza Newlin-Łukowicz. 2015. Selectional restrictions as phonotactics over sublexicons. *Ms., NYU.*

Hayes, Bruce, Colin Wilson, and Benjamin George. 2009 Maxent grammar tool. *Java program downloaded from http://www. linguistics. ucla. edu/people/hayes/Maxent-GrammarTool*.

Magomedova V. 2013. Lingui-Pingui script collection for linguistic search, Ms., St.Petersburg State University.

Polivanova, Anna K. 1967. Obrazovanie umen'shitel'nyx suschestvitel'nyx muzhskogo roda. *Russkij jazyk v natsional'noj shkole.* (reprinted, 2008. *Obscheee russkoe jazykoznanie: Izbrannye raboty, volume 4, 8–22. Moscow: RGGU.)*

[5] The *ik/ek* difference in the internet data is not only a matter of spelling. The difference between two suffixes may be tested: the *–ek* suffix drops it's vowel in certain inflection forms while the *–ik* one does not. We did the relevant search.

Martin Maiden

Inflexional morphology and gender in Romanian.
Why Romanian only has two genders, and how it might yet acquire a third

Romanian occupies a prominent place in the typological literature on gender by displaying extensive and productive *genus alternans*: a subclass of nouns denoting non-mortals (incorrectly called 'inanimate') displays exclusively masculine gender agreement in the singular and exclusively feminine gender agreement in the plural. These nouns are traditionally labelled 'neuter' and thereby distinguished from masculine and feminine. In the theoretical literature (notably Corbett 1991:150-52), and in the light of Hockett's classic definition of genders as 'classes of nouns reflected in the behavior of associated words' (1958:231), such nouns are described as constituting a third 'controller' gender, since their agreement patterns are different both from masculine and from feminine nouns, even though agreement 'targets' only distinguish masculine and feminine. In reality, invocation of a 'third gender' actually obfuscates the real mechanisms underlying Romanian alternating gender, by focusing on an epiphenomenon of those mechanisms. Romanian alternating gender can only be understood by postulating just two genders.

A fundamental problem with all accounts asserting a Romanian 'third gender' is that membership of this class appears aleatory, and the impossibility of 'reversed' alternating gender (singular feminine and masculine plural) is unexplained. In fact the behaviour of this class is demonstrably, and exclusively, a function of inflexional morphology. Certain kinds of inflexional structure (a bare stem, or -u, in the singular, for masculine; -e in the plural, for feminine) necessarily select masculine or feminine agreement respectively. In Romanian, gender is, to a much higher degree than in other Romance languages, determined by inflexional form. So-called 'neuters' are nothing more than nouns the inflexional properties of whose singulars are such that they can do no other than select masculine agreement, and the inflexional properties of whose plurals are such that they can do no other than select feminine. This is essentially the synchronic insight of Bateman and Polinsky (2010), but I show that it receives compelling support from diachrony. Otherwise general morphological changes involving the introduction of inflexional structures which would not unambiguously signal gender are either prevented from applying in the case of such nouns or, if they apply, can be shown to entail immediate loss of gender alternation gender. Apparent counterexamples (e.g., introduction of the gender-ambiguous ending *-i* into the plural) turn out systematically not to be so, because they only occur in the presence of types of root -allomorphy uniquely distinctive of plural feminines (most perspicuously, where the plural has a root-allomorph in final *..ur-*).

Other apparent counterexamples (particularly involving the characteristically femininine plural agreement behaviour of coordinated *singulare tantum* nouns, or the assignment of recent, morphologically invariant, lonwords to *genus alternans*) will be shown to be attributable to a default plural agreement strategy: there is overwhelming evidence of various kinds that in Romanian the only possible agreement in the case of coordinated non-mortal referents, in the absence of any morphological clue to the contrary, is feminine.

I believe, however, that claims that Romanian (especially as articulated by Corbett, or Loporcaro and Paciaroni 2011:319) has a 'third gender' are premature, not 'incorrect'. First, it is in principle possible for speakers to have 'covert knowledge' of the agreement patterns of nouns, independently of their inflexional structure, and this

is perhaps shown by the fact that speakers can 'block' morphological changes which would otherwise obscure the link between form and gender. I also adduce evidence from some Romanian dialects for a 'nascent' third, controller, gender: alternating gender persists, where inflexional signalling of gender has been compromised by sound change. The difficulty is that the examples are extremely few, and may constitute an 'inquorate' (and unproductive: cf. Gardani 2013) gender. In conclusion, however, I consider some southern Italian dialects, having *genus alternans* nouns historically cognate with those of Romanian, but where extensive sound change has neutralized inflexional distinctions apparently without destroying the gender alternation. I shall suggest that for there to arise unambiguous evidence of a third, 'controller', gender of the alternating type claimed for Romanian, it may be universally necessary to go through the stages of, first, inflexionally conditioned agreement and, second, non-morphological obscuration of inflexional gender distinctions.

References

Bateman, Nicoleta and Polinsky, Maria 2010. Romanian as a two-gender language. *Festschrift for David Perlmutter* ed. by Donna Gerdts, John Moore & Maria Polinsky, Cambridge: MIT Press, 41-77.

Corbett, Greville 1991. *Gender*. Cambridge: CUP.

Gardani, Francesco 2013. *Dynamics of Morphological Productivity: Evolution of Noun Classes from Latin to Italian*. Leiden: Brill.

Hockett, Charles 1958. *A Course in Modern Linguistics*. New York: Macmillan.

Loporcaro, Michele and Paciaroni, T. 2011. Four-gender systems in Indo-European. *Folia linguistica* 45:389-334.

Francesca Masini

Morphologically complex words and multiword expressions: Competition or cooperation?

Competition in morphology is generally viewed as a relation holding between different strategies (typically different affixes / WFRs, or a complex word / WFR vs. simple word) that compete to realize the same grammatical or lexico-conceptual meaning within a specific subfield (either word-formation or inflection). Competition has also been used to refer to the relationship between morphology and syntax in the formation of linguistic expressions (cf. e.g. Ackema & Neeleman 2004). In this paper I would like to explore a third type of competition that has to do with word formation, and that is far less investigated (but see Hüning & Schlücker 2015), i.e. the competition between the products of word formation, namely morphologically complex words (MCWs), and so-called *multiword expressions* (MWEs), namely items that are larger than a morphological word and are nonetheless stored into our lexicon.

MWE is a widely known umbrella term that refers to a pretty large set of objects (see, a.m.o., Sag et al. 2002 and Baldwin & Kim 2010), including idioms (*spill the beans*, *pull the wool over someone's eyes*), complex nominals (*weapons of mass destruction*, *hall of residence*), verb-particle constructions (*wrap up*, *give up*) and other complex predicates (*take advantage* (*of*), *give rise* (*to*)). In the theoretical morphology literature, some relevant MWEs have been analyzed and discussed, especially verb-particle constructions, defined as a case of 'periphrastic word formation' by Booij (2002), and so-called phrasal names/nouns (cf. Booij 2009, Masini 2009).

In this paper I propose a first characterization of the relationship between MCWs and MWEs by illustrating a series of cases from various languages, with special attention to Italian.

First, I show that we can think of the competition between MCWs and MWEs as operating at different levels of abstraction/specificity, since it may involve either specific lexical items, or lexeme formation patterns. For instance, we may think that a possible compound like °*capostato* in Italian is blocked by the presence of a MWE such as *capo dello stato* 'head of state'. At the same time, we may suppose that the competition lies at a more abstract level, that of NN compounding vs. NPN phrasal nouns (cf. e.g. Delfitto & Melloni 2009).

Second, I illustrate that the MCW-MWE relation may be investigated both synchronically – see e.g. the above-mentioned formation of NN compounds vs. NPN phrasal nouns in Romance, or the formation of AN phrasal nouns vs. AN compounds in German (cf. Schlücker & Hüning 2009) – and diachronically – see e.g. the emergence of verb-particle constructions in Italian, which superseded the locative prefixal system typical of Latin (cf. Iacobini & Masini 2007).

Third, I discuss the actual nature of the relation between MCWs and MWEs: is it really *competition*, or is it rather a form of *cooperation*? Data from Italian will be used to discuss this point.

Finally, I frame the picture sketched above against a constructionist view of the language architecture, such as Construction Grammar (Hoffmann & Trousdale 2013) or Construction Morphology (Booij 2010), where 'constructions' (i.e. form-meaning pairings) of different size, complexity and schematicity are organized into a hierarchical 'lexicon' called constructic-on.

References

Ackema, Peter & Ad Neeleman. 2004. *Beyond morphology*: *Interface conditions on word formation.* Oxford: OUP.

Booij, Geert. 2002. Constructional idioms, morphology and the Dutch lexicon. *Journal of Germanic Linguistics* 14(4). 301-329.

Booij, Geert. 2009. Phrasal names: a constructionist analysis. *Word Structure* 2(2). 219-240.

Booij, Geert. 2010. *Construction Morphology*. Oxford: OUP.

Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkhya & Fred J. Damerau (eds.), *Handbook of Natural Language Processing*, 267-292. Boca Raton: CRC Press.

Delfitto, Denis & Chiara Melloni (2009). Compounds don't come easy. *Lingue e Linguaggio* VIII(1). 75-104.

Hoffmann, Thomas & Graeme Trousdale (eds.). 2013. *The Oxford handbook of Construction Grammar*. Oxford: OUP.

Iacobini, Claudio & Francesca Masini. 2007. The emergence of verb-particle constructions in Italian: locative and actional meanings. *Morphology* 16(2). 155-188.

Masini, Francesca. 2009. Phrasal lexemes, compounds and phrases: a constructionist perspective. *Word Structure* 2(2). 254-271.

Hüning, Matthias & Barbara Schlücker. 2015. Multi-word expressions. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation. An International Handbook of the Languages of Europe*, Vol. 1, 450-467. Berlin: De Gruyter.

Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. *Proceedings of CICLing 2002*, 1-15.

Schlücker, Barbara & Matthias Hüning. 2009. Compounds and phrases. A functional comparison between German A+N compounds and corresponding phrases. *Italian Journal of Linguistics* 21(1). 209-234.

Dimitra Melissaropoulou

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

The aim of this paper is to offer further insights on the notion of "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.

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). Different phenomena are put under scrutiny such as:

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

A detailed analysis of the above mentioned phenomena reveals that, while in a broader perspective all might ultimately be seen as instances of loss of morphological complexity, they should not be treated under a unified account. On the contrary, 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, (e.g. ii and iv), while others (e.g. iii and v) 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.

Generalizing, our data seem to suggest that complexity or loss of it may arise as the outcome of historical and socio-linguistic circumstances and various sorts of restructurings (cf. Anderson 2015), arguing in favor of the idiosyncratic and language specific nature of morphological complexity especially in situations of language contact. This presentation is meant as a contribution to the discussion on morphological complexity in situations of language contact, arguing that in assessing complexity we should always bear in mind cross-linguistic divergence.

References

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

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

Janse, Mark forthcoming. Cappadocian. In Tzitzilis, C. (ed.), *Η ελληνική γλώσσα και οι διάλεκτοί της.* Thessaloniki: Institouto Neoellinikon Spoudon (Manolis Triantafyllides Foundation).

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

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*. Amsterdam: Benjamins, 111–150.

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

Anne Merkuur, Arjen Versloot & Willem Visser

Distribution and productivity of changing past participles in Frisian

This paper evaluates a particular change in the inflection of Frisian past participles, attested in various former dialect surveys (Van den Berg, 2003; Boelens & Van der Woude, 1955). Frisian has two classes of regular verbs (class I with an infinitive ending in *-ə*, class II with an infinitive ending in *-jə*), both with their own inflectional paradigms, and quite a range of strong and irregular verbs. Past participles of weak class I verbs are marked with the weak suffix *-t* or *-d*, as in (1) (norm). Alternatively, participles of this kind with a stem already ending in *-t* or *-d* get extended with the suffix common for strong verbs: *-ən* (alternative):

1)	norm	alternative
prat-ə	praat-**t** > praa**t**	prat-**ən**
talk-INF	talk.PTCP	talk-PTCP

Vice versa, monosyllabic past participles of strong verbs, which normally end in *-(ə)n* , sometimes get extended with the weak class I suffix *-t* or *-d*, as shown in (2):

2)	norm	alternative
sjen	sjoe**n**	sjoen-**d**
see.INF	see.PTCP	see-PTCP

Both of these alternative forms can be analysed as the result of the process of reanalysis, whereby the normative forms are analysed as not having a suffix (Tiersma, 1999). This results in these forms getting a new suffix, which makes them more transparently inflected. Both forms have been present as dialectical forms for quite a long period in, roughly, the south-western part of Fryslân (Tiersma, 1999: 63-64, 67), but, according to several dialect surveys, may be spreading over the rest of the Frisian language area (Van der Veen, 1980).

Since the use of these forms has not been thoroughly investigated recently, an online survey was conducted to find out whether the use of these forms is indeed increasing and whether a change is going on. We also conducted a *wug* test, to further investigate how people use and interpret these suffixes. The survey, which was completed by 280 speakers of Frisian, showed that the alternatives of the type in (1) have spread and are used in the majority of the cases all across the language area, whereas the alternatives of the type in (2) are still marginal dialectical forms.

On the basis of this outcome, we argue that reanalysis alone is not a sufficient explanation for the emergence and spread of the alternative forms. While reanalysis may have led to the initial emergence of the forms, it does not immediately explain why only the type 1 alternatives spread and the others did not. Because analogy and levelling tendencies are known to play a large role in the processes of verbal inflection, we propose that the reason for the spread of type 1 is to be found in its position in the verbal system and the associations with and across other forms and paradigms. In order to determine where precisely this analogical support for the type I alternatives comes from, we try and test the goodness of fit of three different approaches which all formalize these analogical associations and the way in

which they are processed in different ways. Hence we will see which one of them is able to predict the emergence of both forms and the spread of the type 1 forms best. The particular approaches are three computational models: Analogical Modelling (Skousen et al, 2002), the Minimal Generalization Learner (Albright and Hayes, 2003) and the computational implementation of Yang's Tolerance Principle (Yang, 2005; Yang, 2012).

References

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, 90(2), 119-161.

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

Boelens, Krine & Gosse van der Woude. 1955. *De dialectatlas van Friesland*. (Reeks Nederlandse Dialectatlassen). Antwerpen: De Sikkel, vol 15.

Skousen, Royal, Deryle Lonsdale & Dilworth B. Parkinson (ed.). 2002. *Analogical modeling: An exemplar-based approach to language* (Human Cognitive Processing vol. 10). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Tiersma, Pieter Meijes. 1999. *Frisian reference grammar*. Leeuwarden: Fryske Akademy.

Veen, Klaas F. van der. 1980. Praat my net fan praten! In Ph.H. Breuker, K. Sikkema & K.F. van der Veen, (eds.), *Coulonnade; Twa-en-tweintich FAriaasjes oanbean oan mr. dr. K. de Vries*, 122-130. Leeuwarden: Fryske akademy.

Yang, Charles. 2005. On productivity. *Linguistic variation yearbook*, 5(1), 265-302.

Yang, Charles. 2012. *Knowledge and learning in natural language*. Springer US, 1680-1681.

Ingo Plag, Lea Kawaletz &Marios Andreou

Meaning in word-formation: The frame-semantic hypothesis

Approaches to this problem have a long tradition and have been tackled using numerous different frameworks. Currently the most elaborate and explicit theory (including formalization) is Lieber (2004 et seq.), which, however, does not have a straightforward answer to the question of which kinds of meaning extensions are possible and which ones should be impossible for a given derivative. This is all the more so for deverbal derivation, where Lieber explicitly leaves open "exactly what the verbal body looks like" (Lieber 2004: 72).

In Lieber's theory, polysemy chiefly emerges through the mechanism of coindexation and violations of coindexation. The problem is, however, that in addition to coindexation some further mechanisms are needed, which are not clearly spelled out. In the analysis of some suffixes, semantic features of the affixes are introduced rather ad hoc (e.g. 'collective') and these features then interact with the meaning of the base to arrive at a particular interpretation. Metonymy is explicitly mentioned as part of this process, but the process itself is not formally modeled, but assumed as a given. In more general terms, Lieber postulates highly abstract skeletal features that often do not straightforwardly translate into the specification of the particular readings of individual derivatives. To spell out the mechanisms of this process would, however, be crucial for a better understanding of the semantic processes at work in the interpretation of complex words.

This paper tackles this problem by putting forward a new formal approach to derivational semantics, i.e. frame semantics. In frame theory (Barsalou 1992a, 1992b), frames are complex structures which model mental representations of concepts. These representations are typed, recursive attribute-value structures, where the attributes are functional relations, assigning unique values to the concept they describe (see Petersen 2007). Using the apparatus of this framework, we hypothesize that the semantics of a derivational process is describable as its potential to perform certain operations (such as metonymic shifts) on the frames of its bases. It is shown that these operations are subject to specific restrictions which are either inherent in the frame architecture (e.g. the condition of bidirectional functionality), or which reflect preferences of the derivational process at hand (see Kawaletz & Plag 2015 for a case study).

We will illustrate the frame-based approach with two derivational processes, both of which show interesting semantic interactions between affix and base: nominalizations with *-ment* (beyond Kawaletz and Plag 2015), and the prefix *un-* as a stereotype negator.

Using a dataset of contemporary neologisms of English we systematically model the semantics of the bases and derivatives in order to find patterns in the transition from input semantics to output semantics. By decomposing the semantics of the base verbs as well as that of the resulting nominalizations using the tools of frame theory, we are able to identify clear generalizations and restrictions concerning possible meaning shifts. For example, with PSYCH verbs as bases we find that *-ment* derivatives show semantic shifts to most elements of the decomposed meaning of the base (STIMULUS, EFFECT, RESULT STATE, CAUSE, TRANSPOSED EVENT), but that some shifts are impossible (EXPERIENCER and INITIAL STATE). The former restriction can be interpreted as a general animacy constraint of *-ment* derivatives, the latter is precluded by the frame architecture.

Thus, the frame-based approach allows us to spell out important details of the mechanisms underlying affix polysemy (e.g. metonymic shifts, lexical rules) that are unaccounted for in other frameworks (such as Lieber 2004).

References

Barsalou, Lawrence W. 1992a. *Cognitive psychology: An overview for cognitive sciences*. Hillsdale and NJ: Erlbaum.

Barsalou, Lawrence W. 1992b. Frames, concepts, and conceptual fields. In Adrienne Lehrer & Eva Feder Kittay (eds.), *Frames, fields and contrasts*, 21–74. Hillsdale and NJ: Erlbaum.

Kawaletz, Lea & Ingo Plag. 2015. Predicting the semantics of English nominalizations: A frame-based analysis of -ment suffixation. In Laurie Bauer, Lívia Körtvélyessy & Pavol Štekauer (eds.), *Semantics of complex words*, vol. 3 Studies in morphology, Springer.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*, vol. 104 Cambridge studies in linguistics. Cambridge and New York: Cambridge University Press.

Petersen, Wiebke. 2007. Representation of concepts as frames. In Jurgis Skilters (ed.), *The Baltic international yearbook of cognition, logic and communication*, vol. 2, 151–170. Kansas: New Prairie Press.

Angela Ralli

On borrowing compounds

In this presentation, I propose to deal with the lexical and structural aspects of borrowing compounds, in a language-contact situation involving several Greek-based dialectal varieties affected by Turkish and Romance.

Data drawn from the Asia Minor Aivaliot (Sakaris 1940), Pontic, Cappadocian and Pharasiot (Dawkins 1916) as well as from Heptanesian (Frangopoulou 2015), Cretan (Chairetakis 2015) and Griko (Rohlfs 1933) show the following properties: (a) Different dialects display varying degrees of productivity into creating native compounds, depending on the extent of socio-cultural contact with the dominant language. For instance, Pharasiot and Griko which have been heavily influenced by Turkish and Romance, respectively, do not show as many compounds as other Greek varieties affected to a lesser extent, such as Cretan and Heptanesian. (b) Full bilingualism may often lead to the adoption of foreign compound structures, see Cappadocian and Pharasiot, but it is not always the case if the dialect preserves an archaic character, as for instance, Pontic. (c) Typical features of Greek compounding, such as the linking element, are usually preserved even in innovative structures driven by contact. (d) Beside a possible phonological adjustment, compound loans or loan constituents of compound loanblends often display morphological changes, e.g. stem allomorphy, according to native morphological requirements. (d) In the rare case where a foreign compound is adopted as a whole, it is reanalyzed into a stem in order to receive Greek inflection.

I argue that although compounding is a very productive word-formation process in Greek, it is not 'change-proof' in a language-contact situation, Nevertheless, it is less prone to change as compared to derivation, since Greek compound structures are very different from those in both Turkish and Romance. As asserted in Ralli (2013a,b) and Bagrlaçlk & Ralli (2015), Greek compounds are right headed, combine stems, have a compulsory linking element between the two constituents and are subject to lexical integrity. As such, they are created in morphology. On the contrary, on the one hand Turkish compounds are clearly phrasal objects built in syntax (Ralli 2013a), and on the other hand, native Romance compounds are predominantly left-headed structures (Scalise 1992), with no linking element, involving combination of full words forms. Assuming that morphological congruence is a prerequisite for the borrowing of morphological structure (Myers-Scotton 2002), Greek compounding seems to resist radical change.

Selected References

Bagrlaçlk, Metin & Angela Ralli 2015. Phrasal vs. Morphological Compounds: Insights from Modern Greek and Turkish. *Language Typology and Universals* 68 (3), 282-323. Special issue Phrasal Compounds from a Typological and Theoretical Perspective, ed. by Carola Trips and Jaklin Kornfilt.

Chairetakis, Georges 2015. *Ta sintheta tis Kritikis* [Compounds in Gretan]. MA Thesis, University of Patras.

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

Frangopoulou, Katerina 2015. *I sinthesi stin Eptanisiaki dialekto* [Compounding in the Heptanesian Dialect]. MA Thesis, University of Patras.

Myers-Scotton, Carol 2002. *Contact linguistics*. Oxford: Oxford University Press.

Ralli, Angela 2013a. Compounding and its locus of realization: evidence from Greek and Turkish. *Word Structure* 6 (2), 181-200.

Ralli, Angela 2013b. *Compounding in Modern Greek*. Dordecht: Springer.

Rohlfs, Gerhard 1933. *Scavi linguistici nella Magna Grecia*. Collezione Meridionale. *Studi e ricerche su lingue e dialetti d'Italia.* Milano: Sansoni.

Sakkaris, Georges 1940. *I glossa ton Kydonion* [The language of Kydonies]. Athens: Academy of Athens.

Scalise, Sergio 1992. Compounding in Italian. *Rivista di Linguistica* 4(1), 175-200.

Dorit Ravid, Orit Ashkenazi, Levie Ronit, Galit Ben Zadok, Tehila Grunwald, Ron Bratslavsky, Efrat Raz & Steven Gillis

Semantic competition in Hebrew morphological families: A developmental analysis

This study examines the learning of Hebrew morphological verb families in view of root semantic opacity, showing how competition between different meanings of the same structural root skeleton is mediated by changing distributions across linguistic development and the pragmatic context.

The verb system is the most tight-knit derivational mechanism in Hebrew, combining two discontinuous morphemes - *binyanim* (literally: buildings) and consonantal roots. *Binyanim* are prosodic templates of vowels (often prefixed), specifying slots for the insertion of the Semitic root radicals (Ravid, 2012; Schwarzwald, 2002). Together, they create derivational verb families based on the same root in combination with different *binyanim,* such as *nixnas* 'enter', *hixnis* 'insert', *kines* 'gather, Trans', and *hitkanes* 'gather, Intr', all based on root *k-n-s*. The Hebrew verb lexicon is entirely organized by this root-and-pattern structure, with about 1,500 different roots and the seven *binyan* conjugations responsible for all verbs and verb families (Ravid et al, in press). Obviously, gaining command of the root-and-*binyan* system is a significant achievement by children (Berman, 1993).

Verb learning is however challenged by semantic opacity in the Semitic root, said to convey the lexical essence of verbs (Bolozky, 1999). Some derivational families are related by a semantically transparent root -- e.g., *k-t-b* in *katav* 'write', *nixtav* 'be written', *hixtiv* 'dictate', *kitev* 'cc' and *hitkatev* 'correspond'. In other cases the root is less semantically coherent, as in *xašav* 'think', *nexšav* 'be considered', *hexšiv* 'take into consideration', *xišev* 'calculate', and *hitxašev* 'act considerately' (root *ħ-š-b*). In yet other cases, the semantic link is rather iffy, e.g. *maxar* 'sell', *nimkar* 'be sold' and *hitmaker* 'become addicted' (root *m-k-r*). Finally, there are many cases of the same root skeleton conveying completely different meanings, e.g., *xilxel* 'percolate' and *hitxalxel* 'become horrified'; *gazam* 'prune' and *higzim* 'exaggerate'; or *ne'evak* 'struggle' and *ibek* 'dust'. Priming evidence shows that verbs with structurally but non-semantically related roots obtain reliable facilitation (Frost, Deutsch & Forster 2000); moreover, adults consistently judge such families as related, despite the obviously different semantics (Glucker, in progress). But how do children learn root-related morphological verb families, given the often competing semantics of structurally similar roots?

We propose that the system 'starts small' (Elman, 1993), that is, children first encounter semantically coherent morphological families with reduced morphological entropy in one sub-system of the *binyanim*. With age and schooling, derivational verb families increase in size and decrease in coherence. By the time families gain root members with competing semantics, learners have come to possess a robust notion of the root as a morphological entity.

We examined root-based morphological families in four databases of spoken and written Hebrew: (i) parental interaction with toddlers aged 1;8-2;2; (ii) children 2-8 years in spontaneous peer talk; (iii) storybooks targeting children aged 2-8; and (iv) written texts produced by school-aged writers. All verbs in the four corpora were counted and analyzed into their root and *binyan* components, with root families rated by adults on a scale of semantic relatedness. Findings indicate that semantic competition is a late-emerging phenomenon, as young children mostly encounter verbs with no morphological siblings. Root families start small with transparent semantics, where one member of the pair has a much larger frequency, while a second, root-related verb with a complex meaning modulation makes a smaller appearance. Larger families with opaque root semantics occur mostly in adults' speech and in written corpora. This developmental progress ensures that the structure-semantics relationship in roots is established early on, enabling the gradual introduction of structurally related yet semantically competing members. Analyses of families with competing root members shows that they are interpreted in supportive, age- and modality-appropriate communicative contexts.

References

Berman, R.A. 1993. Marking of verb transitivity by Hebrew-speaking children. *Journal of Child Language*, 20, 641-669.

Bolozky, S. 1999. *Measuring productivity in word formation*. Leiden: Brill.

Elman, J.L. 1993. Learning and development in neural networks: the importance of starting small. *Cognition*, 48, 71-99.

Glucker, A. In progress. Semantic opacity in Hebrew verb roots. Doctoral dissertation, Tel Aviv University.

Frost, R., Deutsch, A. & Forster, K.I. 2000. Decomposing morphologically complex words in a nonlinear morphology. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26, 751-765.

Ravid, D. 2012. *Spelling morphology: the psycholinguistics of Hebrew spelling*. New York: Springer.

Ravid, D., O. Ashkenazi, R. Levie, G. Ben Zadok, T. Grunwald, R. Bratslavsky & S. Gillis. In press. Foundations of the root category: analyses of linguistic input to Hebrew-speaking children. In R. Berman (Ed.) *Acquisition and Development of Hebrew: From Infancy to Adolescence*. Amsterdam: Benjamins.

Schwarzwald, O.R. 2002. *Modern Hebrew morphology*. Tel Aviv: The Open University.

Enzo Santilli

The puzzling case of overabundance in Italian comparatives

Italian language commonly realizes comparison of majority by using periphrastic – also called 'analytic' – forms composed of the adverb *più* 'more' placed before the qualified adjective. For example, we might have a sequence like the following one:

Positive degree adjective		Comparative	
alto	>	*più*	*alto*
tall.M.SG		more	tall.M.SG
'tall'		'taller'	

Nevertheless, some synthetic forms expressing comparison and deriving from Latin remained in usage, and these are nowadays used along with the analytic ones. Traditionally, four pairs of synthetic and analytic comparatives are recognized in Italian, and Italian grammaticography has always considered the elements of these pairs as being synonymous forms. The pairs in question are: MIGLIORE/PIÙ BUONO 'better'; PEGGIORE/PIÙ CATTIVO 'worse'; MAGGIORE/PIÙ GRANDE 'bigger', 'older'; MINORE/PIÙ PICCOLO 'smaller', 'younger'. The aim of this work is to understand if, and if so, to what extent, overabundance can be realized between the forms of analytic and synthetic comparatives by following the canonical approach firstly described by Corbett (2005).

Overabundance is a kind of non-canonical behavior of paradigms discussed for the first time by Thornton in 2011. According to the author, it occurs when "a cell in a paradigm is filled by two or more synonymous forms which realize the same set of morpho-syntactic properties" (Thornton 2011: 360). These multiple forms filling a single cell may differ in three ways: showing different composition/structure, showing different lexical material, showing different inflectional endings (Thornton 2011: 361). For what concerns Italian comparatives, if we take, for instance, the forms of the pair MIGLIORE/PIÙ BUONO, we have two synonymous forms that realize the same set of morpho-syntactic features which show different composition/structure (the latter is a periphrastic form, the former is not). Once verified if the different forms appear in the same context with no conditions (no conditions = the two forms appear in the same context without any semantic or syntactic restrictions), the more our forms will respect a frequency ratio close to 1:1, the more overabundance will be canonical (Thornton 2012: 460).

In Santilli (2014), a study has been done in order to see to what extent analytic and synthetic comparatives realize overabundance. Here, data has been drawn from *la Repubblica* corpus (1985-2000), (≈ 330 million of word-tokens), with a research focused on the forms of the pair PIÙ BUONO/MIGLIORE when they express comparison of majority.

By simply searching the forms of the two elements in the corpus, the forms and shapes of MIGLIORE widely outnumbered the ones of PIÙ BUONO (72198:554, ratio 130.32:1), 300 out of 554 forms of PIÙ BUONO were comparatives of majority, 147 were relative superlative, 107 were not involved in comparison. Forms and shapes of PIÙ BUONO appeared mainly along with NP heads in the semantic areas of food (e.g. *Il mio vino è più buono del tuo* 'My wine is tastier than yours') or kindness (e.g. *l'umanità diventerà sempre più buona* 'humankid will become kinder'). The work consisted in verifying if a given synthetic form appeared in the same semantic area along with the same NP head to which the corresponding analytic comparative referred to. If so, with the two forms being

synonymous and in free variation, overabundance was realized. The results do not seem to be encouraging. In fact, both analytic and synthetic comparatives appeared only along with 17 out of the 173 different NP heads, with an average ratio PIÙ BUONO/MIGLIORE equal to 1:12.41, meaning that, for what concerns comparison of majority, analytic and synthetic forms of Italian language realize overabundance in very few occasions, and in a considerably non-canonical way.

One might wonder if overabundance can be more canonical when these two very same forms are relative superlative instead of comparatives of majority. By querying the same corpus, this is what I am going to investigate in this study.

References

Corbett, Greville G. 2005. *The canonical approach in typology*. In Frajzyngier, Zygmunt, Adam Hodges and David S. Rood (eds.), *Linguistic Diversity and Language Theories*. Amsterdam / Philadelphia: Benjamins.

la Repubblica corpus (1985-2000) - http://sslmit.unibo.it/repubblica

Santilli, Enzo 2014. *Italian Comparatives: a Case of Overabundance?*

Thornton, Anna M. 2011. *Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology*. In Maiden, Martin, John Charles Smith, Maria Goldbach and Marc-Oliver Hinzelin (eds.), *Morphological autonomy: perspectives from Romance inflectional morphology*. Oxford: Oxford University Press. 358-381.

Thornton, Anna M. 2012. *Reduction and maintenance of overabundance. A case study on italian verb paradigms.* In *Word Structure*, vol. 5, issue 2. 457-468.

Paraskevi Savvidou

Issues of competition in evaluative morphology: A corpus-based study of Modern Greek intensifying affixes and compounding components

The present paper deals with three main topics: a) a corpus-based description of the system of Modern Greek intensifying affixes and compounding components, b) the investigation of the issue of competition in the system of these morphological elements and its potential relation with the "peculiar" character of evaluative morphology, which in this study is identified mainly in regard to its semantic/pragmatic characteristics, and c) the exploration of the potential contribution of corpus linguistics in a function-based approach of competition in evaluative morphology.

The intensifying sub-lexical units are examined in the data of the Corpus of Greek Texts (CGT), which consists in approximately 30 million words from various genres of written and spoken texts. The corpus-based methodology is designed on the basis of two main methodological principles: (a) the extension of the notion of co-occurrence in two levels: the word formation level (namely, various characteristics of the bases or compounding components which the elements under examination tend to combine with) and the (con)text level and (b) the combination of qualitative and quantitative analysis on the study of every aspect of the behavior of the sub-lexical units under examination, including function identification, combinatoriality, productivity etc. The theoretical framework combines and extends certain principles of existing morphological models which focus on the meaning and/or function of sub-lexical units, as well as models which focus on the pragmatic aspects of word formation (see Dressler and Merlini Barbaresi 1994), adapting them to the proposed corpus-based methodology.

The results offer a comprehensive account of the behavior of each individual affix and compounding component, which includes a full description of their semantic/pragmatic contribution in their derivatives and compounds, the characteristics of the bases they tend to select (lexical categories, semantic characteristics, connotations and/or semantic prosody, register), their frequency and productivity, as well as a description of the derivatives and compounds they form, such as their lexical categories, meaning, connotations or semantic prosody, register, distribution to individual genres of spoken and written texts etc. The contribution of the present description in comparison with the existing important research on the system of intensifying affixes of Modern Greek (see Gavriilidou 2014, Efthymiou 2015), consists in its corpus-based methodology, which allows the examination of the dynamic relation between the formation and the use level of the elements under examination, offering a perspective which cannot be in view without the use of corpus data.

The methodological choice of studying the formation and the use level in interaction contributes significantly in the study of competition as well. As was expected, the findings show that there is competition in the system under examination. The difference in comparison with morphological competition in general (in *non*-evaluative morphological elements and rules) is not in its extent but in the factors that motivate it, as in the system under examination one of these factors is the pragmatic function of the competing elements. The kind of the communicative aims that are served by the pragmatic intensification favours the occurrence of morphological

competition as an acceptable pleonasm of the system, which allows the speaker to "intensify the intensification".

Corpus-based analysis gives support to the above assumptions. A set of characteristics which occurs in a particular sub-category of competing elements and concerns several aspects of their behavior, including productivity, frequency, co-occurrence in both word and (con)text level, register etc, seems to be associated with the evaluative character of their function and the intention of the speaker to intensify it. The role of morphopragmatics of evaluative morphology as a motivation of competition needs further examination and, if confirmed, it may have significant implications in the study of the potential resolution of the phenomenon, as it introduces a new parameter to the traditional approach of morphological competition as a transitory stage which is motivated by borrowing and similar factors (see Bauer 2006).

References

Bauer, Laurie 2006. Competition in English Word Formation. In van Kemenade, A. & B. Los (eds.), *The Handbook of the History of English*. Oxford: Blackwell Publishing Ltd., 177-198.

Dressler, Wolfgang U. & Lavinia Merlini Barbaresi 1994. *Morphopragmatics: Diminutives and Intensifiers in Italian, German and Other languages*. Berlin: Mouton de Gruyter.

Efthymiou, Angeliki 2015. Modern Greek diminutive and augmentative adjectives (in a cross-linguistic perspective). *SKASE Journal of Theoretical Linguistics* 12(1), 57-71.

Gavriilidou, Zoe 2014. The system of intensifying prefixes in Modern Greek. *Lingvisticæ Investigationes* 37(2), 240–255.

Marcel Schlechtweg & Holden Härtl

The memorization of complex constructions from a cross-linguistic perspective

The current paper asks whether German, French and English AN/NA constructions differ in terms of memorization and, if they do, what this difference tells us about the distinction between compounds and phrases (cf. also Kotowski, Böer & Härtl 2014 for a recent study on German). Native speakers of these languages participated in an auditory memorization experiment on three days. On each day, the experiment consisted of a memorization and a recall phase. In the memorization phase, subjects were asked to memorize non-lexicalized complex items (e.g. *Jungtourist*/*jeune touriste*/*YOUNG tourist*/*young TOURist*) and, as a baseline, existing nouns (e.g. *Architekt*/*architecte*/*architect*/*architect*) of their native language. The German AN constructions examined here were compounds and the French AN/NA constructions were phrases as the adjective only agreed with the noun in French. In English, due to the absence of inflectional markers, AN constructions with initial stress or non-initial stress could only be speculated to represent a possible compound or phrase respectively (Chomsky & Halle 1968). In the recall phase, subjects responded to items that they memorized in the memorization phase (response = yes) as well as to items that they did not memorize (response = no). We focus now on the dependent variable of RESPONSE TIME and the independent variables LANGUAGE (German, French, EnglishA (complex items with initial stress), EnglishB (complex items with non-initial stress)), ITEM TYPE (complex items (AN/NA), existing nouns) and DAY (1, 2, 3).

Existing nouns were not hypothesized to exhibit a difference across languages. For the complex items, we hypothesized a memorization advantage of morphological products. Two distinct reflections as to how a *memorization advantage* becomes evident were taken into consideration, i.e., first, compounds show faster response times on all three days together (MemoA) or, second, compounds show a greater improvement than phrases in that they are slower than phrases on day 1 but not on day 2 and/or 3 (MemoB).

While the existing nouns in German and French did not differ, the German AN compounds showed a memorization advantage in comparison to the French AN/NA phrases (MemoA). MemoA, however, was not applicable for the English data as the advantage caused by non-initial stress simply resulted from its status as the "normal 'unmarked' stress pattern" (Giegerich 1992: 252) in English AN constructions rather than from memorization (cf. also Bybee 1995 referring to Greenberg 1966; Schiller, Fikkert & Levelt 2004). As a consequence, we used MemoB but did not find a significant result.

Since almost all complex items used in the study were semantically compositional, we developed a follow-up study to investigate whether the interaction of stress and semantic compositionality has an impact on the memorization of AN constructions in English because both morphology and lexicalization/memorization seem to be connected to initial stress and semantic non-compositionality (cf. Bakken 2006; Bauer 2004; Chomsky & Halle 1968; Downing 1977). Our analysis of the memorization experiment of this study revealed a memorization advantage (MemoB) of semantically non-compositional items with initial stress (e.g. *HARD shirt*) in comparison to semantically compositional items with non-initial stress (e.g. *short BRUSH*). We will discuss our results against the background of a cognitive distinction between compounds and phrases as well as its implications across languages.

References

Bakken, Kristin. 2006. Lexicalization. In Brown, Keith & Anderson, Anne H. & Bauer, Laurie & Berns, Margie & Hirst, Graeme & Miller, Jim (eds.), *Encyclopedia of language & linguistics*, 106-108. Oxford: Elsevier.

Bauer, Laurie. 2004. Adjectives, compounds, and words. *Nordic Journal of English Studies* 3(1). 7-22.

Bybee, Joan. 1995. Diachronic and typological properties of morphology and their implications for representation. In Feldman, Laurie Beth (ed.), *Morphological aspects of language processing*, 225-246. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chomsky, Noam & Halle, Morris. 1968. *The sound pattern of English.* New York, NY: Harper and Row.

Downing, Pamela. 1977. On the creation and use of English compound nouns. *Language* 53(4). 810-842.

Giegerich, Heniz J. 1992. *English phonology: An introduction*. Cambridge: Cambridge University Press.

Greenberg, Joseph. 1966. *Language universals*. The Hague: Mouton.

Kotowski, Sven & Böer, Katja & Härtl, Holden. 2014. Compounds vs. phrases: The cognitive status of morphological products. In Rainer, Franz & Gardani, Francesco & Luschützky, Hans Christian & Dressler, Wolfgang U. (eds.), *Morphology and meaning: Selected papers from the 15th International Morphology Meeting, Vienna, February 2012* (Current Issues in Linguistic Theory 327), 191-203.Amsterdam: John Benjamins.

Schiller, Niels O. & Fikkert, Paula & Levelt, Clara C. 2004. Stress priming in picture naming: An SOA study. *Brain and Language* 90. 231-240.

Natalia Slioussar

Singular nouns looking like plurals cause more agreement attraction than genuine plurals

Much work has been devoted to number attraction errors in subject-verb agreement, as in (1). Among other things, it was noted that only plural attractors cause a significant effect. In languages with morphological case, attraction is much stronger when the form of the attractor coincides with nominative plural, like in the German example (2a) as opposed to (2b) (Hartsuiker et al. 2003). Attraction was observed both in production (triggering more errors) and in comprehension (making errors less noticeable).

(1) **The key to the cabinets were rusty* (*key* = head noun, *cabinets* = attractor).

(2) a/b. *die Stellungnahme gegen die Demonstrationen / zu den Demonstrationen* 'the position against the$_{\text{ACC.PL(=NOM.PL)}}$ demonstrations / on the$_{\text{DAT.PL(≠NOM.PL)}}$ demonstrations'

We show that singulars can cause attraction too – if they look like Nom.Pl forms. Gen.Sg forms of some Russian nouns coincide with Nom.Pl (and Acc.Pl) forms: e.g. *večerinki* from *večerinka* 'party'. We compared them to genuine plurals in production and comprehension.

Exp.1 was run on a PC using *Presentation* software. 32 participants. In every trial, participants saw a predicate, like (3a), then a subject, like (3b-c), and were asked to produce a complete sentence. Half of the predicates did not agree with the subjects in number, so participants had to modify these predicates. Eight protocols included 80 target items with Acc or Gen attractors in one of the 8 conditions (Sg/Pl head, attractor and predicate) and 120 fillers.

(3) a. *byla krasivoj / byli krasivymi* 'was beautiful$_{\text{SG}}$ / were beautiful$_{\text{PL}}$'

b. *doroga/dorogi čerez pole/polja* 'road$_{\text{NOM.SG / PL}}$ across field$_{\text{ACC.SG(≠NOM.PL) / ACC.PL(=NOM.PL)}}$'

c. *komnata/komnaty dlja večerinki/večerinok* 'room$_{\text{NOM.SG / PL}}$ for party$_{\text{GEN.SG(=NOM.PL) / GEN.PL(≠NOM.PL)}}$'

Agreement errors occurred only with Sg heads and three attractor types: 49 errors (22.3%) with Acc.Pl, 13 errors (5.9%) with Gen.Sg, and 2 errors (0.9%) with Gen.Pl. A mixed-effects logistic regression model shows that the main effects of case and number and their interaction are significant ($p < 0.01$). Thus, looking like a Nom.Pl was more important than carrying a Pl feature.

Exp.2 was run on a PC using *Presentation* software and self-paced reading method**.** 32 (different) participants. We took sentences from Exp.1 ('N1 P N2 was/were Adj/Part') and added four words to them (PPs modifying the predicate). There were eight protocols with 80 target sentences (half ungrammatical) and 150 fillers (grammatical). Average RTs are presented in Fig. 1 and 2. RM ANOVA analysis of RTs shows that there are significants effects ($p < 0.05$ for F1 and F2) in region 5 (Acc and Gen groups) and region 7 (Gen group). Like in production, Gen.Sg (=Nom.Pl) nouns cause attraction (the effect of the number error on the verb is diminished), while Gen.Pl (≠Nom.Pl) nouns have alsmot no attraction effect, but a later slow down indicates that the readers come back to these errors and revise them, unlike in the Acc.Pl case.

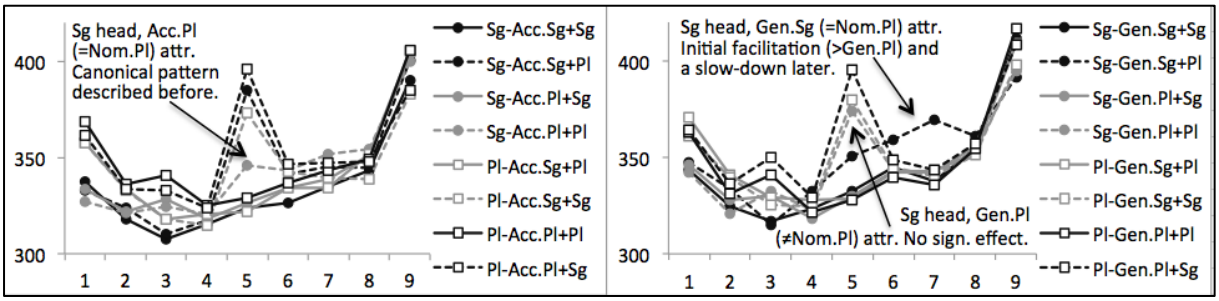


Fig. 1 & 2. Average RTs per region (in ms) in the Acc and Gen groups. Regions: $N1_1$ $Prep_2$ $N2_3$ was/were$_4$ Adj/Part$_5$ + four-word PPs. Template for condition names: 'head-attractor+predicate'.

Conclusions

- There are two major approaches to attraction: representational (errors arise due to illicit feature percolation from the attractor or similar mechanisms) and retrieval-based (errors arise when subject DPs are accessed to determine/recheck the number on the verb). Our data are problematic for both because both depend on the presence of a Pl feature: this is what should percolate or be retrieved.
- We did not see ungrammaticality illusions (delays in grammatical sentences due to interference from the attractor, see e.g. (Wagers et al. 2009) for discussion), so we will argue that the error arises at the retrieval stage, but our data shed new light on the nature of the representation that is retrieved.
- Our study offers new insights on grammatical ambiguity processing. Unlike in many cases discussed before, in this study, at the stage when we see or produce an ambiguous form we are certain about its case (defined by the preposition). Still, alternative feature sets associated with it get activated to the extent they can influence agreement.
- We found preliminary evidence that retrieval errors may be revised – if there is no Pl feature.
- We found attraction with Gen.Sg forms in production, so our data are harder to explain in non-lexicalist frameworks saying that syntax operates with sublexical units and word forms are inserted at the last stage.

References

Hartsuiker, R., & al. (2003). *Mem Cognition, 31*, 1316-1326.
Wagers, M.W., & al. (2009). *J Mem Lang* 61: 206-223.

Natalia Slioussar & Anton Malko

Gender agreement attraction in Russian: Different profiles in production and comprehension

Agreement attraction errors like (1) have been extensively studied in the last 20 years (Bock & Miller 1991; Eberhard et al. 2005; Wagers et al. 2009, among others). The phenomenon was observed in production (attraction errors are more frequent than the ones without attraction) and in comprehension (attraction errors cause smaller delays in reading times), and the results were largely parallel. In particular, only plural attractors were found to cause a significant effect.

(1) **The key to the cabinets were rusty* (*key* = head noun, *cabinets* = attractor).

Two major approaches have been suggested: i) attraction arises due to erroneous feature percolation from the embedded DP or similar mechanisms, i.e. because the syntactic structure is built incorrectly; ii) attraction arises because memory retrieval is noisy, and sometimes a wrong noun is selected during dependency resolution.

With a few exceptions, previous studies focused on number. So it is interesting whether the observed patterns would be confirmed for other features, e.g. for gender, especially in the systems with more than two genders. Badecker and Kuminiak (2007) and Malko and Slioussar (2013) report experiments on gender agreement attraction in production in Slovak and Russian. Gender systems in these two languages are similar (M, F and N genders, frequency: M>F>N, but N is used as default, e.g. in impersonal sentences). The results of the studies were also similar: F attractors caused more errors than M ones, M ones more than N ones. We ran three comprehension (self-paced reading) experiments on Russian, and, unlike in number agreement attraction studies, the results were not parallel to production ones.

In Exp.1, 40 Russian speakers read sentences in which a past tense predicate matched or mismatched in gender with a subject noun ('head') or a noun embedded inside a PP complement ('attractor'), following the scheme in (2). Head NPs were Nom.Sg, attractor NPs were Acc.Sg (we chose NPs for which Acc.Sg=Nom.Sg because this is known to boost attraction).

(2) NP_{HEAD} – P – NP_{ATTR} – was – Adj/Part – four additional words modifying the predicate

We manipulated grammaticality and attraction (2x2) in the following combinations of heads, attractors and verbs: MMM, MFM, MMF, MFF; MMM, MNM, MMN, MNN; FFF, FMF, FFM, FMM; NNN, NMN, NNM, NMM (every target sentence appeared in four conditions, exemplified in (3)). We had four experimental lists with 48 target and 120 filler sentences. One third of sentences were followed by forced choice comprehension questions.

(3) *Recept na porošok/maz' byl/byla…* 'prescription$_{M.NOM.SG}$ for powder$_{M.ACC.SG=NOM}$ / ointment$_{F.ACC.SG=NOM}$ was$_{M.SG}$ / was$_{F.SG}$'

We found a classical attraction profile in the sentences with M attractors and F and N heads (significantly smaller delays in FMM and NMM ungrammatical sentences), see Fig. 1, but no attraction in the sentences with F and N attractors and M heads, see Fig. 2 (RM ANOVAs were used, for all reported significant differences, $p<0,05$ for F1 and F2). Two subsequent experiments with analogous materials and design confirmed this pattern. In Exp.2 (N=32), we

looked at MMM, MFM, MMF, MFF and MMM, MNM, MMN, MNN conditions again, in Exp.3 (N=36), we looked at FFF, FNF, FFN, FNN and NNN, NFN, NNF, NFF conditions and found a classical attraction effect (significantly smaller delays in FNN and NFF ungrammatical sentences)

These findings motivate a modified retrieval-based account of attraction (gender, unlike number, is specified for each noun in the lexicon, which also supports the noisy memory retrieval account). We propose that (i) while N behaves as the unmarked gender in production, M appears to behave as the unmarked gender in comprehension. (ii) Only heads with marked gender (F/N) initiate predictions, as only these are associated with overt verb agreement. (iii) Retrieval is initiated at the verb if there is either no prediction or an unmet prediction. (iv) The retrieval mechanism engages differently with unmarked (M) and marked (F/N) attractors. The parser engages a shallow check for the presence of an M feature, which allows it to mistake structurally inappropriate attractors for grammatical licensors. But the more prominently represented F/N nouns engage more thorough processing, leading to recognition of the structural mismatch and ensuing conflict-induced slowdown.

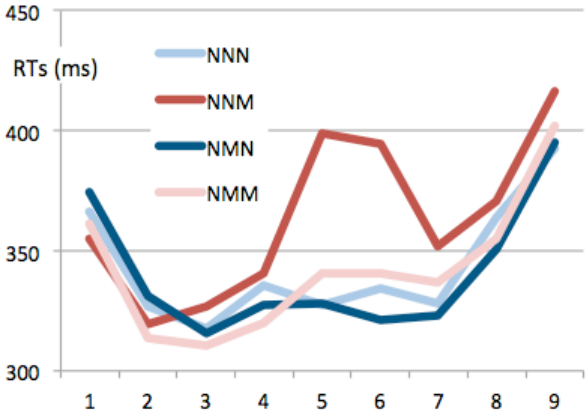


Fig. 1. Average RTs: N heads and M attractors.

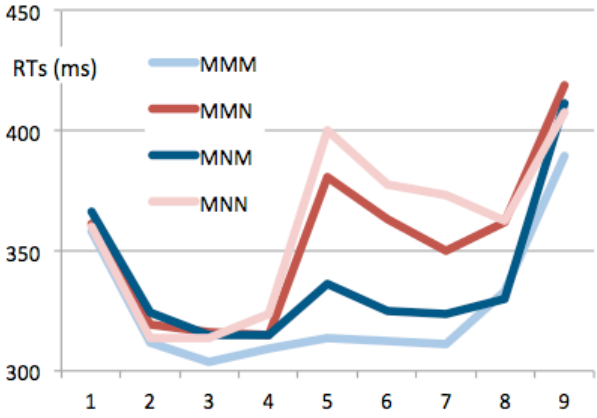


Fig. 2. Average RTs: M heads and N attractors.

References

Badecker, W., & Kuminiak, F. (2007). *J Mem Lang* 56: 65–85.

Bock, J.K., & Miller, C.A. (1991). *Cogn Psychol* 23: 45–93.

Eberhard, K.M., & al. (2005). *Psychol. Rev.* 112: 531–559.

Malko, A., & Slioussar, N. (2013). *FASL 2012 proceedings* (pp. 162-175). Ann Arbor, MI: Michigan Slavic Publications.

Wagers, M.W., & al. (2009). *J Mem Lang* 61: 206–223.

Alexandra Soares Rodrigues

Multiple affixation in Portuguese: Structural restrictions and processing conditions

In recent years, different studies have focused on affix combination (Cf. Hay & Plag 2004; Plag & Baayen 2009; Talamo 2015; Manova & Aronoff 2010). Independently of the difference between the approaches, none of them is dedicated to the analysis of the syntagmatic extension of affixation.

This paper aims to analyse the syntagmatic extension of multiple affixation that involves suffixes of Portuguese productive patterns (Rio-Torto et alii 2013). We base our theoretical claims on the analysis of corpora (Reference Corpus of Contemporary Portuguese and Linguateca) and of experiments (lexical decision and recall tasks) with native speakers.

Under the perspective of the description of structural constraints, multiple affixation should be possible, as long as structural constraints between affixes are preserved.

However, how far does this multiple combination maintain semanticality and grammaticality?

Structural constraints predict that words found in corpora such as *comercializabilidade* and *materializabilidade* are possible forms. In fact, they are the result of multiple affixation that manifests existent derivational patterns in Portuguese:

matéria$_{N}$>*materi*$_{N}$-*al*$_{ADJ}$>*material*$_{ADJ}$-*iz*$_{V}$>*materializ*(*a*)$_{V}$-*vel*$_{ADJ}$>*materializabil*$_{ADJ}$-*idade*$_{N}$

In spite of the structural well-formedness of the derivatives that result from those patterns, forms such as *materializabilidade* are considered to be odd by Portuguese speakers. We claim that processing conditions may explain this. In the processing of this kind of words, in a parallel dual-route model (Baayen, Dijkstra & Schreuder 1997; Schreuder & Baayen 1997), the decomposed route would be favoured due to the conjugation of two factors: the very low-frequency of the word and its morphological complexity degree (Niswander-Klement & Pollatsek 2006).

However, this route encounters one obstacle: Considering that the goal of morphological analysis made by the speaker/listener is to create meaning (Libben 2015) and that the presence of morphemes in a word expresses a conceptual category (either we consider morphemes as lexical items or as spell-outs of paradigms or rules), the quantity of morphological segments with no referential semantics makes morphological processing difficult. Another obstacle goes against a whole-word processing: Since this combination of suffixes has a very low frequency, it does not correspond to predictable/expected combinations in the speaker's mind (Hawkins & Blakeslee 2004; Plag & Baayen 2009).

Two series of multiple affixation contrast with those cases:

I- Words such as *ornamentalização,* although containing four suffixes (*orna-men-al-iz*(*a*)-*ção*), are accepted. Two reasons may explain this: 1) the first suffix (-*ment*-) has not a visible derivational role here, since the word *ornamento* occurs as a concrete and not as an event noun; 2) the sequence -*al-iz*(*a*)-*ção* is very frequent in contemporary Portuguese, which turns it predictable and easily processed (Schreuder & Baayen 1997).

II- Multiple affixation in diminutives/augmentatives. Affix combinations that produce nouns/adjectives such as *cas-inh-oto-zinho, pequen-in-inho-zinho* are frequent and easily processed via the decomposed route. The hypothesis is the following: 1) evaluative affixes have more transparent semantics than suffixes such as -*al*-, -*bil*- have; 2) in evaluative formations each one of the suffixes repeats the semantic information of its precedent. In this case, multiple affixation is reinforcing information that is needed to process the word.

References

Baayen, Harald, Ton Dijkstra & Robert Schreuder. 1997. Singulars and plurals in Dutch: evidence for a parallel dual-route model. *Journal of Memory and Language*. 37. 94-117.

Hawkins, Jeff & Sarah Blakeslee. 2004. *On intelligence*. New York: Henry Holt and Company.

Hay, Jennifer B. & Ingo Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural language and linguistic theory*. 22. 565-596.

Libben, Gary. 2015. Word-formation in psycholinguistics and neurocognitive research. In Peter Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation. An international handbook of the languages of Europe*, 203-217. Berlin: Mouton de Gruyter.

Manova, Stela & Mark Aronoff. 2010. Modeling affix order. *Morphology*. 20. 109-131.

Niswander-Klement, Elizabeth & Alexander Pollatsek. 2006. The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Mem Cognit*. 34(3). 685-702.

Plag, Ingo & Harald Baayen. 2009. Suffix ordering and morphological processing. *Language*. 85(1). 109-152.

Rio-Torto et alii. 2013. *Gramática derivacional do português*. Coimbra: Imprensa da Universidade de Coimbra.

Schreuder, Robert & Harald Baayen. 1997. How complex simplex words can be. *Journal of Memory and Language*. 37. 118-139.

Talamo, Luigi. 2015. Suffix combinations in Italian: Selectional restrictions and processing constraints. In Stela Manova (ed.), *Affix ordering across languages and frameworks*, 175-204. Oxford: Oxford University Press.

Pavel Štichauer

Morphomic splits within periphrasis: Mixed paradigms in Italo-Romance

In this paper, following mainly Corbett 2013: 182-185; 2015a; 2015b, I intend to explore the nature of some of the mixed paradigms (where there is auxiliary alternation *esse/habere* (E/H) within one and the same paradigm) attested in some Italo-Romance varieties.

In standard Romance languages, compound tenses can exhibit one generalized auxiliary or they can present auxiliary selection on the basis of the active/stative split. However, this clear-cut situation where, on the one hand, one perfective auxiliary is generalized, and, on the other hand, auxiliary selection follows the split intransitivity criterion, is far from being the only pattern of alternation attested in the Romance languages (Loporcaro 2007; 2014: 53; Ledgeway 2012: 321). The crosslinguistic variation in Romance operates also at the level of other categories or features which can be involved in the auxiliary alternation: in various Italo-Romance varieties the alternation follows mood distinctions, tense, or even person and/or number.

The *person-based systems* have recently come to the fore precisely because the auxiliary selection follows here apparently unpredictable distributions within one and the same paradigm, as in the dialect of L'Aquila in which E/H alternate according to the pattern EEH-EEH. This type of *mixed paradigm* presents an important theoretical challenge because the attested patterns exhibit distributions which in some cases resist a unified functional explanation. In a majority of cases, the alternations are considered to involve some syntactically or semantically relevant features, but in others they require a purely morphological approach, namely the one which interprets some patterns as *morphomic*. This hypothesis comes very close to Cruschina's (2013: 274-275) proposal that *„morphomic distribution is also found at the level of periphrasis."*

There are clearly splits which can be said to be *motivated*, such as the one already introduced above which is the most common (Loporcaro 2007: 182-183). However, there are other combinations, such as HEE-HHH, HHE-HHH, HEH-HHH (see Loporcaro 2007: 184). Moreover, apart from the cases where one and the same pattern runs across all compound tenses, there are also patterns limited to just the present perfect or instead to counterfactual; and there are also cases of the so-called *free variation* where specific cells of the paradigm allow freely for either auxiliary.

For example, the Abruzzi variety of Campli, displays in the present perfect the common pattern EEH-EEH, but shows free variation E/H in the counterfactual. But the free variation involves only 1st pers. sg. and 3rd pers. pl. exhibiting thus **E/H**.E.E.-E.E.**E/H,** which parallels the U-distribution of the stem alternation defined by Maiden (2011: 223-241)**.** Similarly, in the dialect of Viticuso, the counterfactual displays free variation in all the cells of singular and only in the 3rd plural, yielding theoretically **H**.**H.H**.-E.E.**H** (coming thus very close to Maiden's N-pattern). This is not to claim that auxiliary selection and stem alternation are one and the same thing; I simply wish to argue that the distribution of both phenomena can be of *morphomic* nature.

The aim of this paper is therefore to describe the well-known patterns (Manzini & Savoia 2005) according to the *gradient* nature starting with the motivated splits and arriving at the strictly *morphomic* ones.

References

Corbett, Greville C. 2013. Periphrasis and possible lexemes. In Marina Chumakina & Greville C. Corbett (eds.), *Periphrasis. The Role of Syntax and Morphology in Paradigms*, 169-189. Oxford: Oxford University Press.

Corbett, Greville C. 2015a. Morphosyntactic Complexity: A Typology of Lexical Splits. *Language* 91(1). 145-193.

Corbett, Greville C. 2015b. Morphomic splits. In Ana R. Luís & Ricardo Bermúdez-Otero (eds.), *The morphome debate. Diagnosing and analysing morphomic patterns*, to appear. Oxford: Oxford University Press.

Cruschina, Silvio. 2013. Beyond the stem and inflectional morphology: an irregular pattern at the level of periphrasis. In Silvio Cruschina, Martin Maiden & John Charles Smith (eds.), *The Boundaries of Pure Morphology*, 262-283. Oxford: Oxford University Press.

Ledgeway, Adam. 2012. *From Latin to Romance*. Oxford: Oxford University Press.

Loporcaro, Michele. 2007. On triple auxiliation in Romance. *Linguistics* 45(1). 173-222.

Loporcaro, Michele. 2014. Perfective auxiliation in Italo-Romance. In Paola Benincà, Adam Ledgeway & Nigel Vincent (eds.), *Diachrony and Dialects. Grammatical Change in the Dialects of Italy*, 48-70. Oxford: Oxford University Press.

Maiden, Martin. 2011. Morphophonological innovation. In Martin Maiden, John Charles Smith & Adam Ledgeway (eds.), *The Cambridge History of the Romance Languages. Volume I. Structures*, 216-267. Cambridge: Cambridge University Press.

Manzini, Maria R. & Leonardo M. Savoia. 2005. *I dialetti italiani e romanci: Morfosintassi generativa* (3 voll.). Alessandria: Edizioni dell'Orso.

Rossella Varvara & Roberto Zamparelli

Polysemy and event nominals competition in Italian

In this talk, we focus on the competition and productivity difference between two forms of nominalization in Italian: deverbal nouns (ex.1, henceforth "DN") and nominal infinitives (ex.2, henceforth "NI"), which are infinitive forms preceded by a definite determiner.

(1) La *costruzione* di una relazione richiede impegno
The construction of a relation requires dedication

(2) Il *costruire* una relazione richiede impegno
The build-INF a relation requires dedication

Our aim is twofold: first, we discuss issues related to the frequency of nominal infinitives vs. the frequency of the corresponding deverbal nouns; secondly, we analyze the semantic differences which allow the co-existence of both forms in some cases.

Previous studies focused on the productivity of various nominalizing suffixes (for italian: Thornton 1990/1991, Gaeta 2002, Melloni 2007, among others), without taking into consideration NIs as an alternative way to generate nominals. Moreover, no exhaustive work exists that explains the semantic difference between NI and morphological nominalizations: Zucchi (1993) and Gaeta (2002) proposed some hypothesis, which, in our opinion, do not cover all aspects of this phenomenon and are based on very few examples.

Relying on corpus evidence from Itwac (Baroni et al.2009), we propose that, when both the NI and the DN exist, they are not true synonyms but they show complementary meanings, each of them inheriting a particular sense of the base verb.

Using the CWB-CQP query language, we extract and analyze the frequencies and the syntagmatic patterns in which the forms under investigation occur. We consider a sample of 30 pairs of DN-NI, extracted from the most frequent nominal infinitives, focusing on the syntactic pattern "DN or NI + *di* ('of') + noun".

A preliminary result is that, despite being completely productive, NIs are far less frequent than DNs and also than bare infinitives (3) used as subjects:

(3) Costruire una relazione richiede impegno
Build-INF a relation requires dedication

Unlike NIs and DNs, bare infinitives cannot take external arguments or appear under the prepositions *con* ('with'), *su* ('over'), *in* ('in'), being as consequence not directly comparable to Nis and DNs. This means that NIs and DNs cover the same range of uses and suggests that the frequency difference noted above comes from the fact that DNs bleeds

the occurrence of NIs.

In cases in which the two forms occur, we observe an interesting effect with polysemous verbs: the two forms refer to different meanings of the base verb, especially in cases in which one of them narrows its meaning to a specific domain.

Consider as an example the NI *"il passare"* ('the pass$_{INF}$') and the corresponding DN *"il passaggio"* ('passage, transfer'). Among the different senses of the base verb *"passare"*, the NI covers only the sense of *"passing of time"*, while the DN refers to all the others (with also a locative meaning of "way, passage"). This is particularly clear if we look at the nouns occurring as arguments of those nominalizations, i.e. nouns in the pattern "NI or DN + preposition *di* ('of') + NOUN":

Nouns arguments of NI	Nouns arguments of DN
ora ('hour')	consegna ('delivery')
tempo ('time')	proprietà ('property')
anno ('year')	mano ('hand')
giorno ('day')	livello ('level')
minuto ('minute')	pacchetto ('package')
settimana ('week')	denaro ('money')

Similarly, the NI *"il fiorire"*('to flower') is used mainly in its metaphorical interpretation, occurring with nouns such as *iniziativa* ('initiative'), *proposta* ('proposal'), *dichiarazione* ('declaration'), *attività* ('activity'), while the DN *"fioritura"* preserves mainly the meaning linked to flowers.

After presenting further evidence, we will discuss the reason of this specific configuration. We will first show how NIs are generally preferred for abstract and metaphorical meanings, while DNs occur more with concrete entities and literal meanings. Finally, we will discuss the hypothesis made by Gaeta (2002), who recognizes a more imperfective and atelic meaning to NIs.

References

Baroni, Marco, Bernardini, Silvia, Ferraresi, Adriano and Eros Zanchetta. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation* 43(3): 209-226.

Gaeta, Livio. 2000. *Quando i verbi compaiono come nomi*. Milano: Franco Angeli.

Melloni, Chiara. 2007. *Polysemy in Word Formation: The Case of Deverbal Nominals*. PhD Thesis. University of Verona.

Thornton, Anna Maria. 1990-1991. Sui deverbali italiani in -mento e -zione. *Archivio glottologico italiano*, LXXV-LXXVI, 169-207, 79-102.

Zucchi, Alessandro. 1993. *The Language of Propositions and Events*. Dordrecht: Kluwer.

Madeleine Voga & Hélène Giraudo

Lexical competition within the derivational and inflectional paradigm: Some evidence from masked priming with French stimuli

According to the lexeme-based view of morphology (e.g. Booij, 2002) it is an abstract stem form of a lexeme, which never surfaces as a concrete word form that constitutes the basis for morphological operations. Languages are nevertheless made of words, i.e., stems and inflected or derived words exist as free word-forms, sharing different relations which can influence morphological processing, despite the fact that they are not of morphological nature: masked priming data (Grainger, Colé & Segui, 1991) showed that orthographic similarity of the prime (e.g., *mûrir* 'ripen') inhibits lexical access of the morphologically complex target (e.g., *MURAL* 'wall$_{ADJ}$'). This inhibition is interpreted in terms of preactivation of lexical representations interfering with target processing, demonstrating that processing a morphologically complex word can be subject to competition exerted by word-forms morphologically unrelated to the word to be identified.

Competition is central for interactive activation models (e.g. McClelland & Rumelhart, 1981; Bowers; Davis & Hanley, 2005a) and is the notion on which the variable 'pseudo-family size', studied here, is based. This variable is designed at the other end of the scale of Morphological Family Size and it reflects the number of lexical items functioning as "antagonists" at the lexical level, thus delaying the identification of the prime (Voga & Giraudo, 2009). The variable is defined as follows: A prime like *portons* ('we carry') can potentially activate all words sharing its initial letters, i.e., it has numerous "pseudo-relatives" at the lexical level, among others *portail* 'portal', *porte* 'door', *port* 'harbour', etc. Our assumption is that all these pseudo-relatives behave like competitors at the lexical level. A verb like *mourir* 'die', on the other hand, is almost a hermit, since the only pseudo-relative it has is the rare *mouron* 'scarlet pimpernel', and therefore will receive a very small amount of competition in the lexical – orthographic level. A word can belong to the pseudo-family of another word even if they don't share their stem, e.g. *portugais* 'Portuguese' is a pseudo-relative of *portons* because its stem is a part of the superset *portugais* (see studies on lexical co-activation, e.g., Bowers, Davis & Hanley, 2005b).

The results of three experiments with French native speakers verify our hypotheses: Experiment 1 shows that derivations of low PsFamSize verbs prime their infinitive form targets, whereas those of large pseudo-families fail to prime. Experiment 2 replicates and extends this finding to opaque derivations for verbs having no competitor at the lexical level. Experiment 3 replicates the results of Exp. 1 under slightly different conditions. Results are interpreted within the supra-lexical approach of morphological processing (e.g., Giraudo & Grainger, 2001), in which abstract morphemic representations receive activation from whole-word form representations, so that word recognition enables the activation of the morphological level, and not the other way round. Among the models postulating the absence of whole-word representations, the Naïve Discrimination Learning model (Baayen, Milin, Filipović Đurđević, Hendrix & Marelli, 2011), based on competition processes during acquisition and learning, can be compatible with the data presented here. On the contrary, our results are incompatible with decompositional accounts in which activation and/or inhibition exerted on a morphologically complex word by its pseudo-family should not play any role.

References

Booij, G. 2002. *The Morphology of Dutch*. Oxford: Oxford University Press.

Bowers, J.S., Davis, C.J., & Hanley, D. 2005a. Interfering Neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition,* 97(3), B45-B54.

Bowers, J.S., Davis, C.J., & Hanley, D. 2005b. Automatic semantic activation of embedded words: Is there a "hat" in "that"? *Journal of Memory and Language,* 52, 131-143.

Baayen, H., Milin, P., Filipović Đurđević, D., Hendrix, P. & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. Psychological review, 118(3), 438-481.

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.

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.

Voga, M. & Giraudo, H. 2009. Pseudo-family size influences processing of French inflections: Evidence in favor of a supralexical account. In F. Montermini, G. Boyé & N. Hathout (eds.) *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 148-155. Somerville, MA: Cascadilla Proceedings Project.

Verena Wecker

Competition in the German plural system: A product-oriented perspective

Plural in German is a complex system, as there are several forms that mark plurality on the noun. Depending on the theoretical background, five to nine markers are assumed (*-(e)n*, -(*e*), *-er*, *-s* that can partly be combined with the Umlaut). Additionally, the definite article *die* can be seen as a syntactic plural marker as for masculine and neuter nouns the article changes from *der*/*das* in the nominative singular to *die* in the nominative plural (*der Mann* (sg.) – *die Männer* (pl.)). The German plural system thus provides a core example of competition in inflectional morphology.

The morphological models describing and explaining this system can be grouped in two approaches: On the one hand, rule- or source-based approaches focus on the regularities underlying the distribution of the plural markers to the nouns. In this perspective, plural forms are derived from the singular form and are not assumed to have an own representation in the mental lexicon. All are equally well suitable to signal the meaning of plurality and the choice of one marker depends entirely on characteristics of the singular form. E.g., in the Dual-Mechanism Model, the plural marker *-s* is analyzed as the default and the only regular plural marker in German (Clahsen 1999). In the productivity approach (Laaha et al. 2006) it is assumed that different degrees of productivity account for the choice of one marker. In analyses in the frame of Optimality Theory the choice of a plural marker is explained by different constraint rankings (Elgersma/Houseman 1999).

On the other hand, the product-oriented approach describes schemas for plural forms that are abstracted from the input and are grouped around a prototype (Köpcke 1998). From this point of view, different plural markers are more or less suitable for signaling the function plurality, depending on their proximity to the prototypical plural schema. Thus, characteristics of the plural form itself are crucial for the choice of a plural marker.

An experimental study was conducted with 65 children from six to ten years old with Turkish or Russian as their L1 and German as L2. In addition, 20 children of the same age with German as L1 were tested. Nonce words (accompanied by the definite article) were presented orally and in written form to the children who had to decide if the form refers to one or several objects. As Turkish and Russian belong to different morphological types of languages, it was thus possible to test whether the theoretical assumptions are valid for different learner types (German as L1, German as L2 and Turkish as L1, German as L2 and Russian as L1). Results show that

- The article is the most important feature for the speakers' interpretation of whether the word form refers to one or several objects; if the nonce word is accompanied by the definite article *der* or *das*, i.e. forms that appear almost exclusively in singular contexts (except for *der* which occurs also in the genitive plural), it is very seldom perceived as a plural form, independent of the following word form.

- If the article is *die* and can thus not be clearly associated with one function (sg. or pl.), speakers interpret the items gradually more often as plural forms the more they resemble the prototypical plural schema.
- If the word form is neither strongly associated with a plural nor with a singular meaning, items with Umlaut are more often interpreted as a plural form than items without Umlaut. This, however, is not true for learners with L1 Turkish, probably due to influences form their agglutinative L1.

These results support the product-oriented approach as they strongly suggest that speakers make use of abstract schemas in interpreting a given word form as singular or plural and that these schemas can be placed on a continuum between a prototypical singular and a prototypical plural schema.

References

Clahsen, Harald. 1999. Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences* 22. 991–1013.

Elgersma, Diana & Houseman, Paul. 1999. Optimality theory and natural morphology: An analysis of German plural formation. *Folia Linguistica* XXXIIl (3-4). 333–353.

Köpcke, Klaus-Michael. 1998. The acquisition of plural marking in English and German revisited. *Journal of Child Language* 25. 293–319.

Laaha, Sabine; Ravid, Dorit; Korecky-Kröll, Katharina; Laaha, Gregor & Dressler, Wolfgang. 2006. Early noun plurals in German: regularity, productivity or default? *Journal of Child Language* 33. 271–302.

Zheng Xu

Chinese adjective-noun combinations in a framework combining Construction Morphology with Realization Optimality Theory

This paper proposes an approach to the formation of Chinese adjective-noun combinations which combines Construction Morphology (Booij 2010) with Realization Optimality Theory (Xu 2007, Aronoff and Xu 2010, Xu 2011, Xu and Aronoff 2011a, b, Xu to appear). This approach is superior to alternative frameworks such as Optimal Interleaving (Wolf 2008) and Construction Morphology.

In Chinese, adjective-noun combinations with *de* ([A *de* N]) can alternate with adjective-noun combinations without *de* ([A N]), e.g. *huang (de) chenshan* 'yellow shirt'. This poses a problem to frameworks such as Optimal Interleaving, which claims that lexical insertion takes place in the phonological component of the grammar. This framework predicts when two forms compete to express a meaning, the shorter form will win because the longer form will cause more violations of phonological markedness constraints. It therefore wrongly predicts that Chinese [A N]'s would win over Chinese [A *de* N]'s, which contain *de* and therefore are longer.

In many contexts only one construction, either [A N] or [A *de* N], is selected because of various types of constraints, either universal or language-specific:

- [A *de* N]'s tend not to immediately occur after *de* because of the OCP constraint, e.g. *Lisi de huang (*de) chenshan* 'Lisi's yellow shirts'.
- [A *de* N]'s cannot name entities while [A N]'s can because the former are phrases while the latter are words (Zhu 1956, Duanmu 1998), e.g. *Zhe jiao huang (*de) chenshan* 'This is called "yellow shirt".' (cf. Paul 2005)
- [A *de* N]'s tend to be used if the adjective is complex, e.g. *hong-tongtong *(de) lian* (red-MODIFYING SUFFIX face) 'reddish face'.
- [A *de* N]'s are used if a selectional restriction forbids the adjective to directly combine with the noun, e.g. **duan he* 'short river' is bad while *duan de he* is good because *duan* 'short' cannot directly modify nouns expressing large objects (cf. *duan chi* 'short ruler', *duan dao* 'short knife').

It is hard to imagine how the above competitions between [A N] and [A *de* N] can be accounted for under Construction Morphology, whose major task is to abstract constructions from their specific instantiations and organize them into inheritance hierarchies. CM does not have a formal mechanism of modeling the above type of competition for realization of meanings.

The above cases can be accounted for under a framework that combines Construction Morphology with Realization OT. Constructions assumed in CM can be easily converted into realization constraints in Realization OT. We can posit two realization constraints which require the meaning 'N(oun) with the property of being A(djective)' to be expressed by either [A N] or [A *de* N]. They can be violated by a null output. In addition to OCP and language-specific selectional restrictions, we can

posit the markedness constraint NOPHRASENAMING, which forbids a phrase to name an entity. This constraint predicts if a language has phrases that can name entities, it should also have phrases that cannot name entities, but not vice versa. We can also propose the markedness constraint $[A\ N]_N$ (A = μ), which requires the adjective in a [A N] compound to be monomorphemic. This constraint also applies to languages such as German and Dutch (Booij 2010).

The above competitions between [A N] and [A *de* N] can be explained as follows. [A *de* N] will be ruled out after *de* by OCP. To express a meaning which names an entity, [A *de* N] will be ruled out by NOPHRASENAMING so that [A N] is selected. When a meaning needs to be expressed by a complex adjective, [A N] will be ruled out by the monomorphemic constraint. When a meaning needs to be expressed by, for example, the adjective *duan* 'short' and its modified noun that expresses a large object, [A N] will be ruled out by the selectional restriction that forbids *duan* to directly combine with a noun denoting a large object. If none of these constraints is violated, [A N] and [A *de* N] will be equally optimal, assuming that the realization of [A N] and [A *de* N] does not take place in the phonological component of the grammar and is not subject to phonological constraints banning longer forms.

References

Aronoff, Mark and Zheng Xu. 2010. A Realization Optimality-Theoretic approach to affix order. *Morphology* 20: 381-411.

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

Wolf, Matthew. 2008. *Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model*. Doctoral dissertation. University of Massachusetts at Amherst.

Xu, Zheng. 2007. *Inflectional morphology in Optimality Theory*. Doctoral dissertation. Stony Brook University.

Xu, Zheng. 2011. Optimality Theory and morphology. *Language and Linguistics Compass* 5.7: 466-484.

Xu, Zheng and Mark Aronoff. 2011a. A Realization Optimality Theory approach to blocking and extended morphological exponence. *Journal of Linguistics* 47.3: 673-707.

Xu, Zheng and Mark Aronoff. 2011b. A Realization Optimality Theory approach to full and partial identity of forms. In Martin Maiden, John Charles Smith, Maria Goldbach, and Marc-Olivier Hinzelin (eds.) *Morphological autonomy: Perspectives from Romance Inflectional Morphology*, 257-286. Oxford: Oxford University Press.

Xu, Zheng. To appear. The role of morphology in Optimality Theory. In Andrew Hippisley and Gregory T. Stump (eds.) *Cambridge handbook of morphology*. Cambridge: Cambridge University Press.

Zhu, Dexi. 1956. Xiandai hanyu xingrongci yanjiu [A study of modern Chinese adjectives]. *Yuyan yanjiu* 1: 83-112.

Julia Zimmermann

Morphological status and acoustic realization: Is there a difference between Bra[d] Pitt and a grille[d] cheese omelet, or between Kate Mo[s] and killer robot[s]?[1]

Recent research on lexemes has shown that homophonous lexemes show striking phonetic differences (e.g. Gahl 2008, Drager 2011). This poses a challenge to traditional models of speech production which locate frequency information at the level of the phonological form, and which postulate that phonetic processing does not have access to morphological information (e.g. Levelt & Wheeldon 1994, Levelt et al. 1999).

These findings induce the question of whether similar problems also hold for allegedly homophonous affixes (instead of free lexemes). Plag, Homann & Kunter (2015) recently found that in English, S (that is [s] or [z]) as markers of plural, genitive, genitive plural, 3rd person singular and the cliticized forms of *has* and *is* systematically differ in length. Furthermore, non-morphemic S was found to be systematically longer than morphemic S.

The present study extends the research on S in two dimensions. First, another group of English affixes and clitics is investigated in order to see whether other homophone suffixes also show systematic phonetic differences. Second, I test whether the phonetic differences between different kinds of S go beyond duration. The parameter I look at is center of gravity.

One set of English morphemes that is standardly assumed to share some of the same allomorphs at the phonological level are past tense, past participle and adjectival marker *-ed*, as well as cliticized forms of *had* and *would*. In this study, I focus on the absolute duration of the allomorph D (that is [d] or [t]), using more than 300 pertinent items from the Buckeye Corpus (Pitt et al. 2007), also including non-morphemic instances of D. Linear mixed effects regression models with a number of pertinent covariates (such as frequency, speaking rate, phonetic environment, etc.) demonstrate that there are significant acoustic differences between at least some of these morphemes. For example, the *had*-clitic is significantly longer than some other D morphemes, including the *would*-clitic. These results mirror Plag, Homann & Kunter's findings for S in so far as there are systematic length differences between the morphemes. However, the distribution of these differences does not pattern with those for S: durations of *has* and *is* clitics were found to be at the same (i.e. shorter) end of the scale, while I find *had* and *would* clitics to be the longest and one of the shortest D's, respectively.

With regard to the center of gravity of non-morphemic S and of the different morphemic S's I used the same set of 644 items from the Buckeye Corpus (Pitt et al. 2007) as Plag, Homann & Kunter (2015). It turned out that *has* has a significantly lower center of gravity than most other morphemic S and a marginally significantly lower center of gravity then non-morphemic S. This means that the spectral properties of final S are at least partially dependent on its morphological

[1]This research was funded by the Deutsche Forschungsgemeinschaft (Forschergruppe 2373 'Spoken Morphology', TP3 VAR, PL 151/8-1 and ZP PL 151/7-1) and by Strategischer Forschungsförderfonds der Heinrich-Heine-Universität Düsseldorf, which I gratefully acknowledge.

status. This finding provides additional evidence for the presence of morphological information in the speech signal.

At the theoretical level, my findings on the duration of D and on the center of gravity of S further challenge standard assumptions in morphological theory, Lexical Phonology and models of speech production.

References

Drager, Katie. 2011. Sociophonetic variation and the lemma. *Journal of Phonetics* 39(4). 694-707.

Gahl, Susanne. 2008. *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3). 474-496.

Levelt, William J. M. 1989. *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

Levelt, William J. M. & Linda R. Wheeldon. 1994. Do speakers have access to a mental syllabary? *Cognition* 50(1). 239-269.

Plag, Ingo, Julia Homann & Gero Kunter. 2015. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*. Available on CJO 2015 doi:10.1017/S0022226715000183

Discriminative perspectives on morphology

Workshop, 17th International Morphology Meeting

Convenor: Jim Blevins

The extreme variation exhibited by morphological systems cross-linguistically has discouraged any serious attempt to challenge the central importance that learning from primary input plays in the acquisition of morphology. However, many of the debates about the acquisition of these systems have been conducted in the narrow context of associative learning models. These models assume that learners gradually build up associations between expressions and concepts or meanings, and tend to disagree about the strategies or mechanisms that support this process of association.

The past decade has seen a resurgence of interest in discriminative approaches based on the Rescorla-Wagner learning model and their application to morphological acquisition and analysis. The key intuition that underlies these models is that learners come to discriminate between objects through a gradual disassocation of noncontrastive properties. This shift in perspective has consequences for the interpretation of a range of morphological issues, including 'U-shaped' learning curves, patterns of suppletion, syncretism and compounding, the organization of gender systems, and the design of general comprehension and production models.

The aim of the workshop is to explore the convergence of current work, assess the cumulative impact of this research on traditional problems and issues, and clarify the general model of language acquisition, structure and use that emerges from this paradigm. The workshop will bring together a number of the linguists who have been instrumental in developing this approach, with the goal of facilitating inter-action and at the same time providing clear expositions of the goals, methods and results of current models, all addressed to a general morphological audience.

Naïve Discriminative Learning of English Compound Stress

Melanie J. Bell and R. Harald Baayen
Anglia Ruskin University and University of Tübingen

A significant body of research has shown that accent placement in English noun-noun compounds is probabilistically related to semantic and distributional variables, including the identities of the constituent nouns, the semantic relation between them and the semantic class of the modifier (e.g. Plag et al. 2007, 2008). Analogical algorithms, implemented using TiMBL or AM::Parallel, can be used to predict accent placement on the basis of these variables, with a high degree of success (Arndt-Lappe 2011). However, as models of compound production by humans, such algorithms assume a very high memory load of stored exemplars, and it is unclear to what extent this assumption is justified. Using the ndl package in R, we show that these probabilistic patterns can emerge through naïve discriminative learning without the need for such massive storage. On a sample of about 10,000 compounds extracted from the spoken section of the BNC, ndl achieves state-of-the-art accuracy in classification of compound stress. This is further evidence that a wide range of the probabilistic distributional patterns found in language, including prosodic patterns, can be learnt through a process of naïve discrimination (cf. Baayen 2011).

References

Arndt-Lappe, Sabine. 2011. Towards an exemplar-based model of stress in English noun-noun compounds. *Journal of Linguistics* 47(3). 549-585.

Baayen, R. Harald. 2011. Corpus linguistics and naïve discriminative learning. *Brazilian Journal of Applied Linguistics* 11. 295-328.

Plag, Ingo, Gero Kunter & Sabine Lappe. 2007. Testing hypotheses about compound stress assignment in English: a corpus-based investigation. *Corpus Linguistics and Linguistic Theory* 3(2). 199–233.

Plag, Ingo, Gero Kunter, Maria Braun & Sabine Lappe. 2008. The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language* 84(4). 760-94.

Discriminative abstraction

James P. Blevins Petar Milin Michael Ramscar
Cambridge Tübingen/Novi Sad Tübingen

Work within the contemporary Word and Paradigm tradition over the past decade has lent a measure of support to a number of basic claims of the classical WP model:

(1) a. Languages exhibit predictive dependencies that permit the deduction of full systems from subsets of forms; individual forms do not in general exhibit properties that facilitate their disassembly and reassembly.
b. Minimal contrasts are not individually meaningful, but serve to distinguish larger units with more stable forms and meanings/functions.
c. The multidimensional space of variation defined by contrasts within sets of forms cannot be reduced to contrasts between the morphotactic or derivational structure of individual forms without loss of information.

Clarification of these issues has made it possible to address deeper questions about the source of patterns and the mechanisms that determine them.

(2) a. Why do morphological systems exhibit predictive dependencies?
b. How do minimal contrasts function collectively to distinguish forms?
c. What is the locus of 'competition' and other paradigmatic contrasts?
d. How is structure represented without the decomposition of units?

At the system level, the analogical generalizations expressed within classical WP models exploit the implicational dependencies exhibited by morphological systems. But these models provide no insight into why the dependencies exist in the first place. Information-theoretic approaches (Milin et al. 2009a,b; Ackerman et al. 2009; Ackerman & Malouf 2013) are successful atquantifying dependencies in terms of uncertainty and uncertainty reduction. But the applicability of these measures does not clarify why they, like classical models in general, work as well as they do.

At the level of individual analyses, the advantages of WP approaches largely derive from the that fact that 'abstractive' analyses (Blevins 2006) can describe the composition and structure of forms that cannot be 'constructed' from their minimal parts (or at least not without the aid of diacritic features that encode 'assembly instructions'). Yet apart from the general observation that word-sized units appear more stable and informative that smaller units, the basis for the descriptive advantages of abstractive analyses of morphological systems remain mostly unexamined.

A discriminative learning approach offers answers to these types of questions. The point of departure for this perspective is the discriminative models developed

over the past decade, largely in parallel with the development of WP models. In a series of papers, Ramscar associates (Ramscar & Yarlett 2007; Ramscar & Dye 2010; Ramscar et al. 2010, 2013; Ramscar 2013) elaborate a general discriminative approach based on the learning rule of Rescorla & Wagner (1972) and Rescorla (1988).

From a discriminative learning standpoint, the predictive dependencies exhibited by morphological systems are not due to abstract economy principles. Instead, interpredictability serves a very practical purpose: it is a prerequisite for the use and propagation of language, given the sparse, biased input that speakers encounter. Interpredictability is a variety of regularity, and regularity aids generalization. The descriptive success of abstractive analyses derives from the fact that the form variation in the systems they are describing serves a discriminative, not an associative (or taxonomic) function. Competition and other types of paradigmatic effects arise in the process of language learning, rather than functioning as strategies for correcting rules or constraints that overgeneralize attested patterns of form variation. Morphological structure remains implicit in networks that represent a speaker's morphological knowledge in terms of Rescorla-Wagner-rule-mediated mappings between systems of form contrasts and systems of lexical contrasts (Baayen et al. 2015).

By offering answers to these questions, a discriminative learning perspective provides a frame of reference that shifts the focus from **how** WP models can be applied to the analysis of morphological systems to **why** they apply so generally.

References

Ackerman, F., Blevins, J. P. & Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In Blevins, J. P. & Blevins, J. (eds.), *Analogy in Grammar: Form and Acquisition*, Oxford: Oxford University Press, 54–81.

Ackerman, F. & Malouf, R. (2013). Morphological organization: The Low Conditional Entropy Conjecture. *Language* **89**, 429–464.

Baayen, H., Shaoul, C., Willits, J. & Ramscar, M. (2015). Comprehension without segmentation: A proof of concept with naive discrimination learning. *Language, Cognition, and Neuroscience.* , (in press).

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

Milin, P., Filipović Đurdjević, D. & Moscoso del Prado Martín, F. (2009a). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language* **60**, 50–64.

Milin, P., Kuperman, V., Kostić, A. & Baayen, R. H. (2009b). Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and

derivation. In Blevins, J. P. & Blevins, J. (eds.), *Analogy in Grammar: Form and Acquisition*, Oxford: Oxford University Press, 214–253.

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija* **46(4)**, 377–396.

Ramscar, M. & Dye, M. (2010). Learning language from the input: Why innate constraints can't explain noun compounding. *Cognitive Psychology* **62**, 1–40.

Ramscar, M., Dye, M. & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of *mouses* in adult speech. *Language* **89(4)**, 760–793.

Ramscar, M. & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognititive Science* **31**, 927–960.

Ramscar, M., Yarlett, D., Dye, M., Denny, K. & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognititive Science* **34**, 909–957.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist* **43(3)**, 151–160.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. & Prokasy, W. F. (eds.), *Classical Conditioning II*, New York: Appleton-Century-Crofts, 64–99.

One word or two? Discriminative effects of word entrenchment and competition on processing compounds and pseudo-compounds

Marcello Ferro*, Claudia Marzi*, Vito Pirrelli*, Christina Gagné°, Thomas Spalding°
*Institute for Computational Linguistics, CNR
°Department of Psychology, University of Alberta

Speakers are sensitive to morphemic structure and semantic transparency during the typing of English compounds. For example, there is an elevation in typing latency for the initial letter of the second constituent (start-C2) relative to the latency for the final letter of the first constituent. The size of this elevation is affected by semantic transparency (Libben & Weber 2014), as well as by exposure to a prime that was semantically related to C1 (Gagné & Spalding 2014). The increase in latency was larger with unrelated primes than with semantically related primes for compounds with transparent heads, but not with opaque heads. Crucially, this elevation at the boundary is less prominent in pseudo-compounds (e.g. *carpet*).

We replicate some of these effects by running a Temporal Self-Organising Map (TSOM) on the same set of test-words. A TSOM is a grid of memory nodes, fully connected to the input stimulus and to all other nodes through Hebbian connections (Ferro et al. 2011; Marzi et al. 2014; Pirrelli et al. 2015). During training, nodes get sensitive to time-bound stimuli. Inter-node connections mimic neuron synapses whose strength determines the amount of influence the activation of one node has on other nodes at a one-tick delay. The level of activation of each node at time *t* is a function of the node's sensitivity to the current stimulus, and of the re-entrant contribution from the activation state of the map at *t-1* through its presynaptic connections. The map's response to a sequence of input symbols is the chain of the most highly firing nodes at each time tick (BMUs).

In training a TSOM with words, connection weights are adjusted to the frequency distribution of the input symbols according to Hebbian principles of discriminative learning. Words sharing substrings compete for the same memory resources because they tend to activate the same node chains. High token-frequency of input strings develops specialised node chains and decreases competition.

This dynamic can account for some effects of sensitivity to English compound structure. Trained on both compounds and pseudo-compounds, a TSOM gets sensitive to structure by turning chains of loosely connected, general-purpose nodes into entrenched specialised sub-chains of BMUs. Compounds not only tend to occur less frequently than their C1/C2 constituents (Ji et al. 2011), but they present, in our training set, lower-frequency bigrams at the C1-C2 boundary than pseudo-compounds do. In compounds, low-frequency bigrams at morpheme boundaries are likely to fire loosely connected nodes with low activation levels. This explains a deeper drop in activation for BMUs responding to the start-C2 in compounds than in pseudo-compounds. The effect is magnified by competition, when TSOMs are trained on an increasing number of C1/C2 constituents (isolated words) (Figure 1).

We propose an explanatory memory-based model of inter-letter typing latencies as the combined effect of predictive specialisation and competition between concurrently memorised compounds and their constituents. Data analysis shows different quantitative correlates between behavioural and simulative data and illustrates the effect of adding compound families (based on either C1 or C2) to the map's training set.

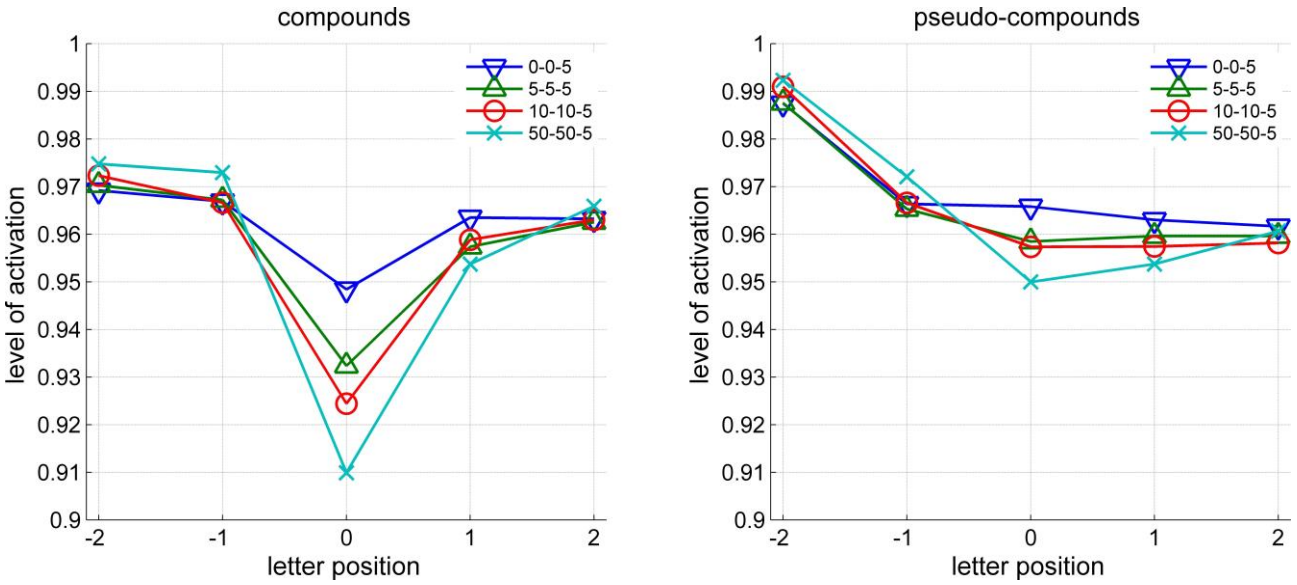


Figure 1 - Node activations by letter-position relative to morpheme boundary. Position 0 corresponds to the start-C2 letter. Different plots illustrate four different training regimes: C1 and C2 constituents are input 0, 5, 10 and 50 times each per regime; (pseudo)compounds 5 times each.

References

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.

Gagné, Christina L. & Thomas L. Spalding. 2014. Typing time as an index of morphological and semantic effects during English compound processing. *Lingue e Linguaggio* XIII(2). 241-262.

Ji, Hongbo, Christina L. Gagné & Thomas L. Spalding. 2011. Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language* 65. 406-430.

Libben, Gary & Silke Weber. 2014. Semantic transparency, compounding, and the nature of independent variables. In Rainer F., Dressler W.U., Gardani F., Luschützky H.C. (eds.) *Morphology and Meaning*. 205–221. Amsterdam: Benjamins.

Marzi, Claudia, Marcello Ferro & Vito Pirrelli. 2014. Morphological structure through lexical parsability. *Lingue e Linguaggio* XIII(2). 263-290.

Pirrelli, Vito, Marcello Ferro & Claudia Marzi. 2015. Computational complexity of abstractive morphology. In Baerman M., D. Brown & G. Corbett (eds.), *Understanding and Measuring Morphological Complexity.* 141-166. Oxford: OUP.

Grammatical cases as cued by prepositions: Entropy effects in processing of isolated prepositions

Dušica Filipović Đurđević[1,2,3] & Isidora Gatarić[2]

[1]Department of Psychology, Faculty of Philosophy, University of Novi Sad

[2]Laboratory for Experimental Psychology, Faculty of Philosophy, University of Novi Sad

[3]Laboratory for Experimental Psychology, Faculty of Philosophy, University of Belgrade

There is a growing body of research revealing that cognitive processing is affected by Information Theory based descriptions of various segments of language. In case of morphology processing, among other things, it has been demonstrated that processing of inflected forms is affected by inflectional entropy, that is entropy of inflected form probability distribution (Tabak, Schreuder, & Baayen, 2005), as well as with relative entropy (KL divergence), that is divergence of probability distribution of paradigm from that of class (Milin, Filipović Durdević, & Moscoso del Prado Martín, 2009). These effects have been interpreted in light of the model that is based on basic principles of learning (NDL; Baayen, Milin, Filipovic Durdevic, Hendrix, & Marelli, 2011). In this research, we aim to test for the effects of entropy derived from the distribution that is not manifested in the word form itself, but rather in a word's syntactic potential – in the forms of its co-occurring words. We selected prepositions as an ideal candidate, having in mind that they do not undergo morphological transformations, whereas, at the same time, they contribute to resolving of syntactic roles of various forms of other word categories (e.g. nouns, adjectives). Previous research conducted in Serbian language found that processing of prepositions is affected by information load derived from average frequency per number of congruent cases, and hence suggested that prepositions do capture syntactic information (Radojčić, & Kostić, 2003). In terms of NDL model (Baayen, et al, 2011) we see prepositions as cues, and grammatical cases as outcomes.

Based on subsample of literary prose from Corpus of Serbian Language (Kostić, 1965) we selected 42 prepositions that were not homophones/homographs with any other word category. Corpus of Serbian Language has the property of being manually annotated at the level of grammatical categories, which means that for each noun, adjective, number or pronoun there is information on its word category, gender, case and number. We used this information to build frequency distributions of prepositions occurring in the context of various grammatical cases. We did so by traversing the corpus and collecting information on grammatical case of the noun, adjective, number or pronoun that is preceded by a given preposition. This procedure revealed that 24 of selected prepositions appeared in a context of a single case, whereas 18 prepositions appeared in multiple case context taking from two up to five different cases of the nominal word that followed. Based on co-occurrence frequencies of the given preposition with different cases of the following nominal word, for each of the multifunctional prepositions we calculated Shannon's Entropy (Shannon, 1948).

We presented 39 native speakers with the selected prepositions in a visual lexical decision task. The processing latencies were analysed in linear mixed-effect regression, with participants and items as random effects, and the additional correction of trial order slope for each participant. After controlling for the effects of trial order, word length in letters/phonemes, and word frequency, for those prepositions that were multifunctional (i.e. that occur in the context of multiple cases), the

analysis revealed significant effect of Shannon's Entropy. Processing latencies were negatively correlated with preposition entropy, revealing that increase in preposition entropy led to shorter processing latencies. This increase in processing speed as a reaction to an increase in complexity is in accordance with previous related findings (Baayen, et al, 2011; Milin, et al, 2009; Tabak, et al, 2005), and will be discussed in terms of NDL model.

Baayen, R. Harald, Petar Milin, Dušica Filipović Đurđević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118. 438-482.

Milin, Petar, Dušica Filipović Đurđević & Fermin Moscoso del Prado Martín. 2009. The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language* 60. 50–64.

Radojčić, Vesna, & Aleksandar Kostić. 2003. Razlike u kognitivnoj obradi predloga i imenica. *Psihologija* 36 (3). 379-394.

Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27. 379-423.

Tabak, Wieke M, Robert Schreuder & R. Harald Baayen. 2005. Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. In Reis Marga & Stephan Kepser (eds), *Linguistic Evidence*, 529-555. Berlin: Mouton de Gruyter.

Unsupervised learning of morphology: how to measure unpredictability?

Ákos Horváth
Eötvös Loránd University, Budapest

Dániel Czégel
Institute for Cross-Disciplinary Physics and Complex Systems, Palma de Mallorca
czegel_d@yahoo.com

Csaba Oravecz
Research Institute for Linguistics, Hungarian Academy of Sciences

Csaba Pléh
Central European University, Budapest
vispleh@ceu.hu

Introduction

When dealing with a morphologically complex language, speakers have to segment the word forms into morphemes. Applications of information theory to language, from the early proposals of Antal (1962) for Hungarian tried to base segmentation on entropy fluctuations. In this paper the feasibility of strictly information theory based approaches to morphological segmentation on a given large corpus of Hungarian are presented. We address the question whether fluctuations of entropy over word forms could be used to predict segmentations that are given by more traditional linguistic means.

Methods

Morphological boundaries are supposed to correspond to peaks in the unpredictability of the next grapheme, given that the previous part of the word is known. The three main approaches to measure unpredictability (Hammarström & Borin, 2011):

a) Letter Successor Variety (LSV): number of possible continuations.

b) Letter Successor Entropy (LSE): determines the Shannon entropy of the next character based on corpus statistics

c) Letter Successor Max-Drop (LSM): only takes into account the continuation having maximal probability, since, as Golcher (2006) argues, one highly predictable continuation is necessary and sufficient to signal a non-break.

Which unpredictability measure is the most applicable for morphological segmentation? Is there any other measure defined in a similar manner that outperforms all of these three? In order to answer these questions, we generalize these three unpredictability measures in a quantity called Rényi entropy (Rényi, 1961), defined as

$$H_\alpha^{\rightarrow} = \frac{1}{1-\alpha}\log_2 \sum_i p_i^\alpha$$

where $\alpha \in [0, \infty)$ is a parameter.

Figure 1 illustrates the $H_\alpha^{\rightarrow}$ Rényi entropies in case of the Hungarian word *rokonságban* (relative-ITY-IN, i.e., 'in relation with'), computed from the Hungarian National Corpus(Oravecz, Váradi, & Sass, 2014).

Usually entropy starts from a high value at the first character and then decreases slowly over the word, independently of the morphological structure of the word. In order to eliminate this baseline, Hammarström & Borin (2011) proposed to compute the backward entropies $H_\alpha^{\leftarrow}$ as well.

Figure 2 shows the $H_\alpha^{\leftarrow}$ values in case of the word *rokonságban*.

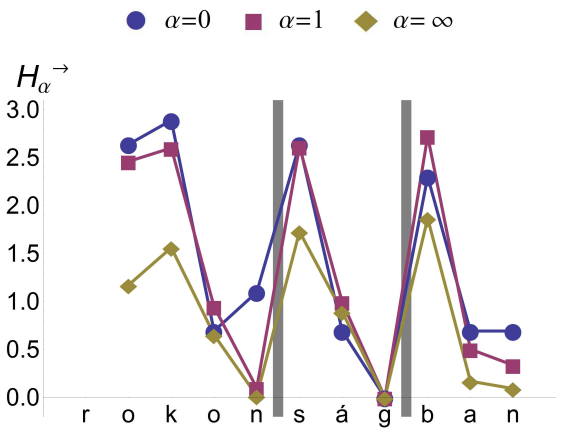


Figure 1: The $H_\alpha^{\rightarrow}$(m) values in case of the Hungarian word *rokonságban*. The real morpheme boundaries are indicated by gray vertical lines.

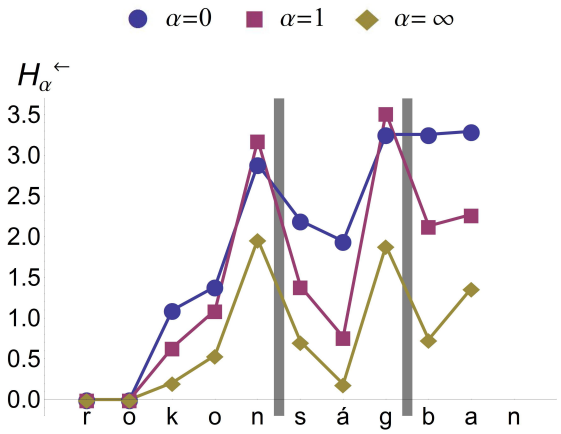


Figure 2: The $H_\alpha^{\leftarrow}$(m) values in case of the Hungarian word *rokonságban*. The real morpheme boundaries are indicated by gray vertical lines.

Having the $H_\alpha^{\rightarrow}$ and $H_\alpha^{\leftarrow}$ values for a given word, our algorithm determines a morpheme boundary between the mth and $m+1$th character if

$$H_\alpha^{\rightarrow}(m+1) \geq K^{\rightarrow} \text{ and } H_\alpha^{\leftarrow}(m) \geq K$$

with some pre-defined threshold parameters $K^{\rightarrow}$ and $K^{\leftarrow}$.

Data and evaluation

For our experiments we compiled a 7.2 million word sample from the new version of the Hungarian National Corpus, which now provides a detailed morpheme level analysis for each wordform, including morpheme segmentation. The method used in the preparation of this corpus for the resolution of segmentation ambiguities is described in (Oravecz et al., 2014).

In order to determine the efficiency of our algorithm, we used the *Jaccard index* to quantify the similarity of the set of morpheme boundaries found by our algorithm b_a and the set of the real morpheme boundaries b_r for a given word, defined as $J = \frac{|b_a \cap b_r|}{|b_a \cup b_r|} \in [0,1]$, i.e., the number of real morpheme boundaries found by our algorithm divided by the number of morpeheme boundaries that are either real or found by our algorithm or both. Then, the overall efficiency E is computed as the average of the Jaccard indices $J(w_i)$ over all of the *multimorphemic* word types w_i weighted by the token frequencies $f(w_i)$ of the types:

$$E = \frac{\sum_i f(w_i) J(w_i)}{\sum_j f(w_j)}$$

Since the morpheme boundaries found by our algorithm are depending on parameters α, $K^{\rightarrow}$ and $K^{\leftarrow}$, the efficiency E also depends on them. Our aim here is to compare the efficiency of the different entropy measures parametrized by α, by choosing the $(K^{\rightarrow}, K^{\leftarrow})$ parameter pair such that we get the maximal efficiency for that given α. This maximal value is denoted by $E_{\max}(\alpha)$.

Results

The maximal efficiency $E_{\max}(\alpha)$ is shown in Figure 6. Although the difference between the efficiency of the different entropy measures is moderate, Shannon entropy H_1 clearly outperforms both H_0 and H_∞ suggesting that information on the probability of the most likely continuation or the mere number of the possible continuations alone is not enough to achieve optimal performance in morphological segmentation. Moreover, the highest efficiency is corresponding to $\alpha = 1/2$ justifying the usage of the Rényi entropy H_α as a generalization of LSV, LSE and LSM.

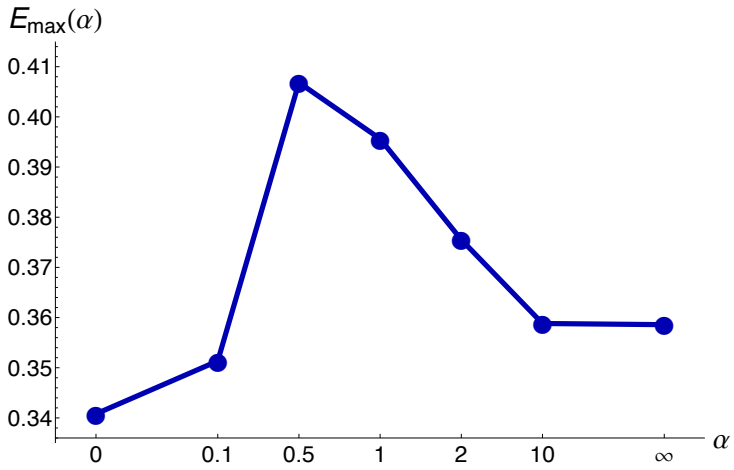


Figure 3: The maximal efficiency $E_{\max}(\alpha)$.

Conclusions

- The highest efficiency can be achieved with $H_{1/2}$ which is not amongst the standardly used entropy measures.
- LSE (i.e., H_1 Shannon entropy) clearly outperforms both LSV (corresponding to H_0) and LSM (corresponding to H_∞).
- By using a simple threshold parameter-pair to indicate morpheme boundaries, the efficiency cannot go beyond 40% (measured as the weighted average of Jaccard indices over the multimorphemic words in the corpus).
- Questions for further analyses: i) Does the efficiency of the segmentation depend on the derivational or inflectional nature of the suffix and its linear position? ii) How does stem or suffix allomorphy influence the results? iii) Are these results universal, at least in case of morphologically complex, agglutinative languages?

References

Andreev, N. (1967). *Statistiko-kombinatornyje metody v teoretičeskom i prikladnom jazykovedenii [statistical-combinatorial methods in theoretical and applied linguistics].* Leningrad: Nauka.

Antal, L. (1962). A megnyilatkozások tagolása morfémák szerint [segmentation of utterances into morphemes]. *Magyar Nyelvőr*, *86*, 189–202.

Golcher, F. (2006). Statistical text segmentation with partial structure analysis. *Proceedings of KONVENS 2006*, 44–51.

Hammarström, H., & Borin, L. (2011). Unsupervised learning of morphology. *Computational Linguistics*, *37*(2), 309–350.

Harris, Z. S. (1970). *From phoneme to morpheme*. Springer.

Oravecz, C., Váradi, T., & Sass, B. (2014). The Hungarian Gigaword Corpus. In *Proceedings of lrec* (Vol. 2014).

Rényi, A. (1961). On measures of entropy and information. In *Fourth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 547–561).

Beyond Rescorla-Wagner in discriminative morphology

Emmanuel Keuleers and Paweł Mandera
Department of Experimental Psychology, Ghent University

In this talk, we will discuss two fundamental limitations of the Rescorla-Wagner model and their implications for the theoretical questions addressed by the NDL model (Naive Discriminative Learning, Baayen et al. 2011) in morphology. First, we explore the limitations posed by the use of binary representation for the cues and outcomes in the model. We argue that this pushes NDL in the direction of using linguistically informed representations rather than naive ones, for instance in the representation of morphological contrasts. We argue that this problem could be solved by using the Delta rule, a more general version of the update rule in the Rescorla-Wagner model which can deal with real vector values. The advantages of this approach are that instead of arbitrary binary representations outcome vectors can be based on distributional semantical analysis and that a priori morphological representations can be dispensed with altogether as there is evidence that we call morphological cases such as plurality and gender can also be derived by vector-arithmetic operations on these distributional semantic representations (Mikolov et al., 2013). Second, we discuss the limitations posed by the single layer architecture of the Rescorla-Wagner model. We will shortly address the misconception that introducing hidden layers in a model increases the complexity and/or computational efficiency of the model and will show that introducing hidden layers in the model is not less cognitively plausible and that it can dramatically reduce the number of connections in the model, thereby increasing computational efficiency.

It's a matter of weights: Enriching a NDR network with FRACSS-based semantic information

Marco Marelli, Marco Baroni

Center for Mind/Brain Sciences, University of Trento

The FRACSS (Functional Representations of Affixes on Compositional Semantic Space) model provides a functional characterization of the semantic dynamics in affixed words (Marelli and Baroni, 2015). FRACSSs are based on the tenets of distributional semantics (Turney and Pantel, 2008), stating that word meanings can be modelled through lexical co-occurrences; however, they make a further step in the morphology domain by approximating affixes as functions that map stem meanings into derived-word meanings. The present work aims at using the FRACSS system to introduce semantic information into the Naïve Discriminative Reader (NDR; Baayen et al., 2011). This latter architecture is able to explain a large number of morphological-processing effects in terms of statistically-reliable patterns between orthographic and semantic units, as captured through the Rescorla-Wagner learning equations (1972). However, the semantic layer included in the NDR is relatively underspecified, with no inclusion of specific characterizations for different morphological units, nor explicit description of the morpheme combination process. Simply, in the NDR a fixed lower activation weight was assigned to affix representations (as opposed to free forms).

The FRACSS system was used to generate weights to include in the computation of unit activations in the NDR. Semantic representations of derived words were obtained by applying the relevant FRACSS to stem meanings (e.g., *-less* FRACSS was applied to *name* to obtain the meaning of *nameless*). The distance between the resulting representation and the original stem was taken as an index of the semantic modification brought upon by the affixation procedure, and included as a stem-specific weight in the NDR system. This modification led to new model-predictions that were tested against the same derived-word dataset used in the NDR paper (Baayen et al., 2011). The FRACSS-enriched version of NDR obtained a better performance than the one of the original NDR implementation. Moreover, the former also outperformed a baseline model in which semantic weights were obtained through a traditional (whole-word based) distributional-semantic system.

The present results indicate that complementing the NDR with semantic information improves model performance, and suggest that NDR predictions may suffer for its lack of semantic specification. In particular, the introduction of FRACSS-based weights leads to a more accurate estimate of the stem frequency effect, whose overestimation was considered a misstep of the original NDR (Baayen et al., 2011). Importantly, the fact that the FRACSS-based approach also outperforms a traditional distributional-semantics baseline confirms that our method is not only improving on the original NDR because of the extra information, but also because of the *nature* of this extra information. However, these observations only apply to the original NDR implementation (Baayen et al., 2011). More recent developments, in which NDR is lexeme-based (as opposed to morpheme-based) and word-to-word associations are included as a further layer, may overcome these issues.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*(3), 438-481.

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review, 122*(3), 485-515.

Rescorla R. A., Wagner A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In *Classical Conditioning II: Current Research and Theory* (Eds Black A. H., Prokasy W. F.). New York: Appleton Century Crofts.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. Journal of *Artificial Intelligence Research, 37*(1), 141-188.

Discriminative morphomes in Kiranti

GÉRALDINE WALTHER
DDL (CNRS, Université Lumière Lyon 2)
geraldine.walther@cnrs.fr

This paper reexamines the structure of Kiranti verbal paradigms from a discriminative perspective (Ramscar *et al.*, 2013). Kiranti languages constitute a sub-branch of the Tibeto-Burman family and are mainly spoken in Eastern Nepal. The study focuses on two Kiranti languages, Khaling (Jacques *et al.*, 2012) and Chatthare Limbu (Tumbahang, 2007). One of the main characteristics of Kiranti languages is that they are highly inflectional with very large verbal paradigms displaying complex patterns of stem allomorphy and form syncretisms spanning cells whose distribution does not follow homogeneous feature splits like person, number, or direct/inverse marking.

In previous work on Khaling, we showed that in a static realisational description, a large part of the descriptive structural complexity (in terms of description length) resides in the non-homogeneous feature expression across the paradigm (Walther *et al.*, 2014). Using a Minimum Description Length (MDL, (Rissanen, 1984)) evaluation measure (Sagot and Walther, 2011), we confirmed quantitatively that although a direct/inverse system still largely structures the paradigm and constitutes a key element for understanding and economically encoding the paradigmatic patterns, the majority of the syncretisms across the paradigm are most economically rendered by highlighting the paradigm's morphomic structure in the sense of (Aronoff, 1994).

This study goes beyond the previous study in that it focuses on the discriminative nature of the paradigms' morphomic properties. This discriminative nature is reflected in both the exact distributions of regularities and form contrasts at full-form level and across corresponding parts of forms. We will show how the distributions in those paradigms implement a complex system of layered overlapping contrasts (prefix contrasts, stem contrasts, suffix contrasts) with partial regularities facilitating the learning of the system and high overall discriminability between verb-forms likely to appear in similar syntactic contexts, enhancing form processing. We compare the word-form structure exuberantly observable in Kiranti with previous findings on prefixing and suffixing preferences within work on discriminative approaches to morphology (Ramscar, 2013). As a result, this study will show how discriminative contrasts lead to the emergence of fundamentally discriminative morphological units whose interaction structures the overall inflectional system.

Acknowledgments This work was supported by the LabEx ASLAN (ANR-10-LABX-0081) operated by the French National Research Agency (ANR).

References

ARONOFF, MARK. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Linguistic Inquiry Monographs. MIT Press.

JACQUES, GUILLAUME, AIMÉE LAHAUSSOIS, BOYD MICHAILOVSKY, AND DHAN BAHADUR RAI. 2012. An overview of Khaling verbal morphology. *Language and linguistics*, **13.6**, 1095–1170.

RAMSCAR, MICHAEL. 2013. Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija*, **46**(4), 377–396.

RAMSCAR, MICHAEL, MELODY DYE, AND STEWART M. MCCAULEY. 2013. Error and expectation in language learning: The curious absence of *mouses* in adult speech. *Language*, **89**(4), 760–793.

RISSANEN, JORMA. 1984. Universal coding, information, prediction, and estimation. *IEEE Transactions on Information Theory*, **30**(4), 629–636.

SAGOT, BENOÎT AND GÉRALDINE WALTHER. 2011. Non-canonical inflection : data, formalisation and complexity measures. In C. MAHLOW AND M. PIOTROWSKI, Eds., *Systems and Frameworks in Computational Morphology*, volume 100, p. 23–45, Zurich, Switzerland: Springer.

TUMBAHANG, GOVINDA BAHADUR. 2007. *A descriptive grammar of Chhatthare Limbu*. PhD thesis, Tribhuvan University, Kathmandu, Nepal.

WALTHER, GÉRALDINE, GUILLAUME JACQUES, AND BENOÎT SAGOT. 2014. The Opacity-Compactness Tradeoff: Morphomic features for an economical account of Khaling verbal inflection. Presentation at the 16th International Morphology Meeting, Budapest, Hungary.

IMM 2016 Workshop

Computational Methods for Descriptive and Theoretical Morphology

Organisers:

Olivier Bonami
U. Paris-Diderot & Lab. de Linguistique Formelle, Paris
olivier.bonami@linguist.univ-paris-diderot.fr

Benoît Sagot
Inria & ALPAGE, Paris
benoit.sagot@inria.fr

Although computational morphology is a respected and well-established subfield of computational linguistics with important applications in NLP, until recently there had long been a lack of cross-fertilisation with work in descriptive and theoretical morphology. This led to situations of mutual misunderstandings (see for example the discussions of theoretical approaches to morphology in Karttunen (2003) and Roark and Sproat (2007)) and missed opportunities. The situation has changed radically over the last decade, with important work in 4 directions.

1. The implementation of morphological fragments provides a means of confirming the validity of analyses. This approach, which is familiar from work in syntax and semantics since the mid-1980s, has started becoming more common, in particular within Network Morphology (Corbett and Fraser, 1993; Brown and Hippisley, 2012) and Paradigm Function Morphology (Stump (2001); see the *Cats CLAW* online tools),[1] but also through ad hoc implementations not tied to a specific theoretical framework.
2. Quantitative explorations have started to uncover previously unstudied aspects of the structure and properties of morphological systems. Three main lines of research can be identified, focusing on implicative structure (Finkel and Stump, 2007; Ackerman, Blevins and Malouf, 2009; Sims, 2010; Ackerman and Malouf, 2013; Stump and Finkel, 2013; Bonami and Beniamine, 2015), on the inference of inflection classes from raw paradigms (Brown and Evans, 2012; Lee, 2014) and on the relative information-theoretic compactness of alternate descriptions of a given system (Walther and Sagot, 2011; Walther, Jacques and Sagot, 2014).
3. Such studies rely on the availability of large-scale electronic morphological lexica, which can be developed using lexicographic and/or corpus-based approaches. Such lexica constitute a way to formalise lexical knowledge, enable quantitative linguistic studies of morphology and the lexicon, and pave the way for natural language processing applications. When freely available, they allow for a better mutualisation of efforts and reproducibility of the experiments — see for example Le*fff* (Sagot, 2010) and Flexique (Bonami, Caron and Plancq, 2014) for French.
4. Computational morphology can also provide techniques useful in the development of

[1] http://www.cs.uky.edu/~raphael/linguistics/claw.html

large-scale resources, especially when dealing with under-resourced languages. One approach applies unsupervised learning of morphology (e.g. Goldsmith (2001); see Hammarström and Borin (2011) for a recent overview) to bootstrap morphological descriptions (Hammarström, 2009). Another line attempts to automatically derive implemented grammars and lexica from existing resources (Bender, Schikowski and Bickel, 2012; Bender, Crowgey *et al.*, 2014).

This workshop is devoted to work using computational methods to address descriptive or theoretical issues in morphology, with a particular focus on these four areas.

Invited Speaker

Rob Malouf (San Diego State University)
Title: *Deep learning for abstractive morphology*

Accepted Papers

Blake Allen. *Sublexical Morphology: learning and generalizing probabilistic inflectional morphology*

Sebastian Bank, Daniela Henze, Jochen Trommer and Eva Zimmermann. *Assessing the typology of person portmanteaux*

Gilles Boyé. *Small world inflection morphology: a fragment for French conjugation*

Berthold Crysmann. *Reduplication in an implemented HPSG of Hausa*

Atticus Harrigan, Lene Antonsen, Antti Arppe, Dustin Bowers, Trond Trosterud and Arok Wolvengrey. *Learning from the Computational Modeling of Plains Cree Verbs*

Nabil Hathout and Fiammetta Namer. *Enriching the Démonette morpho-semantic network: computational and linguistic issues*

Roland Mühlenbernd, Dankmar Enke and Igor Yanovitch. *Modeling the Grammaticalization Path of Functional Morphemes. A Game-Theoretic Analysis*

Sabine Stoll, Jekaterina Mažara and Balthasar Bickel. *Measuring the acquisition of morphology in big longitudinal corpora*

Programme Committee

Delphine Bernhard (Université de Strasbourg, France)
Emily Bender (University of Washington, USA)
Olivier Bonami (Université Paris-Diderot, France)
Dunstan Brown (University of York, United Kingdom)
Roger Evans (University of Brighton, United Kingdom)
Nabil Hathout (CNRS, France)
Rob Malouf (San Diego State University, USA)
Fiammetta Namer (Université de Lorraine, France)
Benoît Sagot (Inria, France)
Andrea Sims (Ohio State University, USA)
Gregory Stump (University of Kentucky, USA)
Géraldine Walther (CNRS, France)

References

Ackerman, Farrell, James P. Blevins and Robert Malouf. 2009. “Parts and wholes: implicative patterns in inflectional paradigms.” In: *Analogy in Grammar*. Ed. by James P. Blevins and Juliette Blevins. Oxford: Oxford University Press, pp. 54–82.

Ackerman, Farrell and Robert Malouf. 2013. “Morphological organization: the low conditional entropy conjecture.” In: *Language* 89, pp. 429–464.

Bender, Emily M., Joshua Crowgey, *et al.* 2014. “Learning Grammar Specifications from IGT: A Case Study of Chintang.” In: *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages, ACL 2014*.

Bender, Emily M., Robert Schikowski and Balthasar Bickel. 2012. “Deriving a lexicon for a precision grammar from language documentation resources: a case study of Chintang.” In: *Proceedings of COLING 2012*.

Bonami, Olivier and Beniamine, Sarah. 2015. “Implicative structure and joint predictiveness.” In: *Word Structure and Word Usage. Proceedings of the NetWordS Final Conference*, Pisa. Ed. by Vito Pirelli and Claudia Marzi and Marcello Ferro, pp. 4–9.

Bonami, Olivier, Gauthier Caron and Clément Plancq. 2014. “Construction d’un lexique flexionnel phonétisé libre du français.” In: *Actes du quatrième Congrès Mondial de Linguistique Française*. Ed. by Franck Neveu et al., pp. 2583–2596.

Brown, Dunstan and Roger Evans. 2012. “Morphological complexity and unsupervised learning: validating Russian inflectional classes using high frequency data.” In: *Current Issues in Morphological Theory: (Ir)regularity, analogy and frequency*. Ed. by F. Kiefer, M. Ladányi and P. Siptár. Amsterdam: John Benjamins, pp. 135–162.

Brown, Dunstan and Andrew Hippisley. 2012. *Network Morphology: a defaults-based theory of word structure*. Cambridge: Cambridge University Press.

Corbett, Greville G. and Norman M. Fraser. 1993. “Network Morphology: a DATR account of Russian nominal inflection.” In: *Journal of Linguistics* 29, pp. 113–142.

Finkel, Raphael and Gregory T. Stump. 2007. “Principal parts and morphological typology.” In: *Morphology* 17, pp. 39–75.

Goldsmith, John A. 2001. “Unsupervised Learning of the Morphology of a Natural Language.” In: *Computational Linguistics* 27, pp. 153–189.

Hammarström, Harald. 2009. “Unsupervised learning of morphology and the languages of the world.” PhD thesis. Chalmers University of Technology and University of Gothenburg.

Hammarström, Harald and Lars Borin. 2011. “Unsupervised learning of morphology.” In: *Computational Linguistics* 27, pp. 309–350.

Karttunen, Lauri. 2003. “Computing with realizational morphology.” In: *Computational linguistics and intelligent text processing*. Ed. by Alexander Gelbukh. Heidelberg: Springer, pp. 205–216.

Lee, Jackson. 2014. *Automatic morphological alignment and clustering*. Tech. rep. University of Chicago.

Roark, Brian and Richard Sproat. 2007. *Computational approaches to morphology and syntax*. Oxford: Oxford University Press.

Sagot, Benoît. 2010. “The *Lefff*, a freely available and large-coverage morphological and syntactic lexicon for French.” In: *Proceedings of LREC 2010*.

Sims, Andrea. 2010. “Probabilistic paradigmatics: Principal parts, predictability and (other) possible particular pieces of the puzzle.” Fourteenth International Morphology Meeting, Budapest.

Stump, Gregory T. 2001. *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge:

Cambridge University Press.

Stump, Gregory T. and Raphael Finkel. 2013. *Morphological Typology: From Word to Paradigm*. Cambridge: Cambridge University Press.

Walther, Géraldine, Guillaume Jacques and Benoît Sagot. 2014. "The Opacity-Compactness Trade-off: Morphomic Features for an Economical Account of Khaling Verbal Inflection." Paper presented at the Sixteenth International Morphology Meeting, Budapest.

Walther, Géraldine and Benoît Sagot. 2011. "Modélisation et implémentation de phénomènes flexionnels non-canoniques." *Traitement Automatique des Langues* 52.2, pp. 91–122.

Blake Allen (University of British Columbia)
Sublexical Morphology: learning and generalizing probabilistic inflectional morphology

This presentation introduces Sublexical Morphology, a computationally implemented, language-general formalism for encoding probabilistic knowledge about inflectional paradigms, as well as its associated learning algorithm. Sublexical Morphology builds on the Sublexical Phonology framework (Becker and Gouskova 2013, Allen and Becker in review), which draws on research in surface-oriented word-and-paradigm/realizational morphology (Stump 2001, Sims 2006, Ackerman, Blevins, Malouf, and Blevins 2009) as well as probabilistic models of grammar, especially Maximum Entropy harmonic grammar (Goldwater and Johnson 2003, Hayes and Wilson 2008, Wilson 2010). While Sublexical Phonology focuses on modeling the morpho-phonological relationship between some base cell (e.g. English plural nouns) and some derivative cell (e.g. English plural nouns) in order to predict responses in *wug* tests (Berko 1958, Becker, Nevins, and Levine 2012), Sublexical Morphology generalizes this approach to the domain of entire inflectional systems.

Concretely, the Sublexical Morphology framework allows a probabilistic grammar governing all implicational relationships in the morpho-phonology of an inflectional paradigm to be learned from minimally annotated training data. The learned grammar can be used to generate probabilistic predictions given arbitrary queries, e.g. the predicted third-person plural present form (or distribution over forms) of some novel verb lexeme given its first-person and second-person singular present forms. In addition to discussing the internal structure of a Sublexical Morphology grammar, I will present evidence that its predictions maintain a high degree of accuracy even when tested on a typologically diverse (Stump and Finkel 2013) range of inflectional systems. Moreover, because Sublexical Morphology models the Paradigm Cell Filling Problem (Ackerman et al. 2009, Malouf and Ackerman 2010), the framework is also useful for investigating related morphological phenomena, such as paradigmatic gaps (Sims 2006), paradigm entropy (Malouf and Ackerman 2010), and paradigm leveling (Albright 2010). Subject to time and audience interest, I will detail ways that Sublexical Morphology can facilitate research in these areas.

Proposal: Sublexical Morphology generalizes the existing model of Sublexical Phonology (Becker and Gouskova 2013, Allen and Becker in review) and its learning algorithm to the domain of entire inflectional systems. In brief, Sublexical *Phonology* deals with a special case of the Paradigm Cell Filling Problem (Ackerman et al. 2009, Malouf and Ackerman 2010) in which there is only a single base cell at play. Sublexical Phonology proposes a particular structure to this inference problem: lexemes are organized into groups (*sublexicons*) such that each group has the same morphological operation (or set of operations) for transducing its base form into its derivative form. For example, taking regular English verb (non-3rd person singular) present and past tense forms as bases and derivatives, respectively:

(1)

	operation	*example lexemes*
Sublexicon 1	add [d] at right edge	love, seize, free
Sublexicon 2	add [t] at right edge	hope, work, snap
Sublexicon 3	add [ɪd] at right edge	greet, seed, plot

In Sublexical Phonology, rather than these three groups of forms being unified by an underlying morpheme for the plural marker, they are left distinct as sublexicons. A Maximum Entropy harmonic grammar (Goldwater and Johnson 2003, Hayes and Wilson 2008, Wilson 2010) is then created for each sublexicon, which provides a wellformedness score for any arbitrary base (e.g. present tense) form *as a member of that sublexicon*, as opposed to as a member of the others. In other words, this score, stated as a probability, indicates the likelihood that the base form "belongs" to that sublexicon. Because each sublexicon is associated with a morphological operation (like "add [d] to the right edge of the base") which allows the generation of a novel derivative (e.g. past tense) form, a probability distribution over sublexicons for the base can be used as the predicted probability distribution over derivative candidates for that word, with each candidate taking its probability from the sublexicon that generates it.

Whereas Sublexical Phonology was primarily developed as a way of solving "phonological" problems like opacity and the encoding of gradient morpho-phonological subregularities into a grammar, Sublexical *Morphology* is designed to solve a "morphological" problem: the Paradigm Cell Filling Problem. While the two domains are irrefutably tightly intertwined, the new term is intended to suggest that the two models are fully consistent with each other in their "sublexicalism" while focusing on different types of linguistic phenomena.

Sublexical Morphology takes the concept of a sublexicon and scales it up such that information from multiple base forms can be included. As a result, the sublexicons in Sublexical Morphology can often resemble the inflectional classes of traditional conceptions of morphology. As a simple example of these multi-base sublexicons, again consider the task of predicting an English past tense verb form, but rather than relying only on the non-3rd person singular present tense forms as bases, also allow the use of 3rd person singular present forms. To distinguish these sublexicons from the single-base sublexicons used in Sublexical Phonology, I refer to multi-base sublexicons as *paradigm sublexicons*. Note that the sublexicon groupings of lexemes from each base form are not necessarily identical, despite being so in this example case.

(2)

	operation from non-3rd sg. bases	*operation from 3rd sg. bases*	*example lexemes*
Sublexicon 1	add [d] at right edge	mutate final [z] to [d]	love, seize, free
Sublexicon 2	add [t] at right edge	mutate final [s] to [t]	hope, work, snap
Sublexicon 3	add [ɪd] at right edge	mutate final [s/z] to [ɪd]	greet, seed, plot

Sublexical Morphology uses the same mechanism as the Sublexical Learner to identify paradigm sublexicons in first inferring the sublexicons for each base-derivative pair, but then then uses a novel process to combine them into a set of paradigm sublexicons. Sublexical Morphology also fits grammars to these sublexicons to encode their phonological subregularities in the same way as in Sublexical Phonology, but does so while considering all base forms together. After learning, the procedure for wug testing is the same as the Sublexical Learner, but again with a twist: generate derivative candidate for each sublexicon (the various bases' operations in a single sublexicon should all generate the same derivative candidate), and then use the sublexicons' grammar to predict a distribution over sublexicons given *all known bases simultaneously*. These probabilities correspond to the predicted probabilities of the various derivative candidates. The extent of "interaction effects" between different base forms is highly restricted; initially, no such interactions (dependencies, in stats language) are posited, but a process of error-driven constraint induction allows a search through a small space for useful interaction constraints.

Evaluation: I will present two types of evidence for the accuracy of algorithmically learned Sublexical Morphology grammars. The most direct method of assessing a model's accuracy is to quantify its prediction error on a set of testing data not given to the learning algorithm. I follow this convention using two types of testing data: items from the training data set that have been "held out", i.e. using cross-validation; and *wugs*, novel words that could not conceivably be in the testing data. In the cross-validation case, I show that for a wide set of inflectional systems including Spanish verbs, Icelandic nouns, and Polish nouns, accuracy on held-out training items is high as measured by the KL-divergence (Kullback and Leibler 1951) of the predicted distribution over forms from the observed distribution. Assessment of model prediction accuracy on wugs is based on recent experimental studies by the author and a co-author on Icelandic and Polish noun paradigms, in which participants were provided subsets of a novel lexeme's inflectional paradigm and then asked to select their preferred inflected forms for a particular derivative cell. Error in the Sublexical Morphology models' prediction accuracy for these cases is again assessed using KL-divergence.

Bibliography

Ackerman, Farrell, James P. Blevins, Robert Malouf, and Juliette Blevins. 2009. Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In *Analogy in grammar: Form and acquisition*, 54–82. Oxford Scholarship Online Monographs.

Albright, Adam. 2010. Base-driven leveling in Yiddish verb paradigms. *Natural Language & Linguistic Theory* 28:475–537.

Allen, Blake, and Michael Becker. in review. Learning alternations from surface forms with sublexical phonology.

Becker, M., A. Nevins, and J. Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88:231–268.

Becker, Michael, and Maria Gouskova. 2013. Source-oriented generalizations as grammar inference in Russian vowel deletion. Ms. lingbuzz/001622.

Berko, J. 1958. The child's learning of english morphology. Doctoral Dissertation, Radcliffe College.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*, 111–120.

Hayes, B., and C. Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Kullback, Solomon, and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 79–86.

Malouf, Rob, and Farrell Ackerman. 2010. Paradigm entropy as a measure of morphological simplicity. In *Workshop on Morphological Complexity: Implications for the Theory of Language, Harvard*.

Sims, Andrea D. 2006. Minding the gaps: inflectional defectiveness in a paradigmatic theory. Doctoral Dissertation, Ohio State University.

Stump, Gregory, and Raphael A Finkel. 2013. *Morphological typology: From word to paradigm*, volume 138. Cambridge: Cambridge University Press.

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

Wilson, C. 2010. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive science* 30:945–982.

Assessing the typology of person portmanteaux

Sebastian Bank, Daniela Henze, Jochen Trommer & Eva Zimmermann – University of Leipzig

There is a widespread intuition in the literature that the distribution of portmanteaux affixes in transitive agreement paradigms is at least partially non-arbitrary and exhibits a strong cross-linguistic preference for 'local' contexts, i.e. clauses where both arguments are 1st or 2nd person (Heath 1991, 1998, Wunderlich 2006, Georgi 2012, Nevins 2012): the *Local Portmanteau Hypothesis*. A similar, but orthogonal claim is made by Lakämper & Wunderlich (1998) and Woolford (2010) who argue that person portmanteaux have a special affinity to 'direct' contexts, i.e., clauses where the subject is higher than the object on the person hierarchy $1 \succ 2 \succ 3$: the *Direct Portmanteau Hypothesis*. Woolford, Lakämper and Wunderlich draw their evidence for the Direct Portmanteau Hypothesis from only a handful of languages, but the most extensive empirical studies on the Local-Portmanteau Hypothesis, Heath (1991, 1998), cite an impressive inventory of local person portmanteaux from different languages. However, to establish the cross-linguistic affinity of portmanteau agreement and local scenarios, it is obviously not sufficient to demonstrate that this context exhibits many portmanteaux, but necessary to show that it exhibits significantly *more* portmanteaux than other paradigmatic contexts. In this talk, we report results of a typological pilot study which tests both hypotheses and related claims against a small, but crosslinguistically balanced language sample. As the identification of portmanteaux depends on understanding complete inflectional paradigms, which is underdetermined by the empirical data and prone to interference from alternating analytical biases, we use automatic computational procedures to segment and analyze unsegmented affix-paradigms. With this novel approach, the criteria for identifying (non-)portmanteaux are fully transparent and ensured to be uniform. Our results indicate that there is no significant cross-linguistic effect of the Local and Direct Portmanteau Hypothesis. We examine alternative learning strategies, and show that our results are independent of the specific evaluation metrics for affix segmentation.

Hypotheses on the distribution of portmanteaux Heath (1991, 1998) argues that due to the pragmatic awkwardness of transitive predications involving both the speaker and the hearer of a speech act, there is a strong crosslinguistic tendency to disguise this situation semantically, and more concretely to make it morphologically less transparent (opaque) in transitive pronominal agreement paradigms, e.g. by the neutralization of number features in this context, or by using allomorphs for the participants not employed elsewhere in the language. Under the assumption that portmanteaux are also morphosyntactically less transparent than marking a $1 \leftrightarrow 2$ predication by distinct affixes for subject and object, he concludes that using portmanteau affixes is also a strategy to the same end. Basically the same claim is made by Wunderlich (2006: 4) who states: "To express the combination I $\rightarrow$ you is a special communicative task, so it does not wonder that a portmanteau morpheme adapted to this special task is found in several languages."

(1) **The Heath-Wunderlich Prediction**

	1	**2**	**3**
1		P	
2	P		
3			

Woolford (2010) advocates a related claim on the distribution of person portmanteaux which implies that only specific local scenarios – those with 1st person subjects and 2nd person objects – show a tendency to exhibit portmanteaux. In fact she argues that the preference for portmanteaux in transitive agreement crosscuts the local/non-local distinction and reflects a person hierarchy in

the sense that there portmanteaux affixes are preferred in constellations where the subject is higher in person than the object. This assumption is implemented by the constraint in (2):

(2) **Person Restriction on Portmanteau Agreement Formation:** (Woolford 2010: 24-25) In a portmanteau agreement form, the person of the subject must be higher than or equal to the person of the object.

Under the assumption that there is a crosslinguistically uniform person hierarchy $1 \succ 2 \succ 3$, (2) predicts that $1 \rightarrow 2$ forms should exhibit more portmanteaux than $3 \rightarrow 2$ forms, but also $2 \rightarrow 1$ forms, and that the preference for portmanteaux should also be observable in $1 \rightarrow 3$ and $2 \rightarrow 3$ forms. Whereas Woolford's explicit claim thus goes beyond local scenarios, the evidence she cites to establish (2) is restricted to local forms. She argues based on the sample of languages cited in Heath (1998) that there are languages which have $1 \rightarrow 2$ portmanteaux, but no $2 \rightarrow 1$ portmanteaux, and languages which have portmanteaux in both contexts, but no languages with $2 \rightarrow 1$ portmanteaux that lack $1 \rightarrow 2$ portmanteaux, resulting in the typology in (3):

(3) **Woolford's Typology of Local Person Portmanteaux**

		2 → 1 Portmanteaux	
		+	−
1 → 2 Portmanteaux	+	attested	attested
	−	non-attested	attested

The same empirical predictions follow from a proposal by Lakämper & Wunderlich (1998). Strictly speaking, the Heath-Wunderlich prediction and the Woolford-Lakämper-Wunderlich prediction are not logically incompatible. Local contexts could show a preference for person portmanteaux over non-local contexts, and direct contexts over inverse ones predicting a distribution as in (4) where "PP" marks a particular preponderance of portmanteaux:

(4) **The Woolford-Lakämper-Wunderlich Prediction and its combination with (1)**

	1	**2**	**3**
1		P	P
2			P
3			

	1	**2**	**3**
1		**PP**	P
2	P		P
3			

Methodology Methodologically our study applies a novel approach to one of the principle problems of morphological typology: the analytic underdetermination of the categories under investigation by empirical data. Thus it is by no means obvious whether a string of segments employed in transitive agreement morphology is a portmanteau or a combination of separable affixes. We tackle this problem by applying automatic decomposition algorithms to full affixal paradigms for regular verb stems. Thus the criteria for identifying (non-)portmanteau affixes become fully explicit and are applied homogeneously across the languages of our sample. Crucially, they are blind with respect to the local/non-local or hierarchical distinctions.

As the combination of segmentation options and meaning assignments results in combinatorial explosion even for smaller paradigms, the space of possible analyses is too large for a naive search algorithm. The procedure in (5) uses a greedy approach such that the analysis of a paradigm is conducted in individual learning steps. At each step, the learner searches for the best form-meaning-pair hypothesis that can be identified in the current state of the paradigm according to an OT-like optimization. Different analytic biases can be explored by manipulating the ranking

of the optimization criteria. After the optimal hypothesis is found, it is added to the lexicon, its instances are removed from the paradigm and the algorithm proceeds to the next step until all material has been learned and the paradigm is empty:

(5) **Iterative Algorithm for Segmentation and Analysis**

a. (i) Build the set P of perfect hypotheses, i.e. all ⟨form, meaning⟩ pairs combining all free affix strings that do not have a free affix as substring with every meaning specification such that the meaning subsumes only and all the paradigm cells where the form occurrs (free or bound)

(ii) Identify the best marker hypotheses $O \subseteq P$ having ($\alpha > \beta > \gamma > \delta$):
- α maximal number of (free or bound) true positives
- β non-portmanteau > portmanteau
- γ minimal number of blind cells
- δ maximal number of segments

b. If $O = \emptyset$

(i) Build the set M of marker hypotheses with minimum 50% precision, i.e. all ⟨form, meaning⟩ pairs combining all free affix strings with every meaning specification such that at least one half of the cells subsumed by its meaning contain an occurrence of the form (free or bound)

(ii) Identify the best marker hypotheses $O \subseteq M$ having ($\alpha > \beta \ldots > \theta$):
- α maximal number free true positives
- β non-portmanteau > portmanteau
- γ maximal number of bound true positives
- δ minimal number of free false negatives
- ε minimal number of false positives
- ζ maximal number of cells
- η minimal number of blind cells
- θ maximal number of segments

c. (i) Add some $\langle form, meaning \rangle \in O$ to the lexicon, let $O = \emptyset$ and remove a single occurrence of $form$ from all paradigm cells subsumed by *meaning*

(ii) If any paradigm cell has a free occurrence of $form$, goto step b.

d. If any paradigm cell has a (non-empty) affix string, goto step a, else end.

Whenever the algorithm fails to identify a perfect hypothesis in (5a), it falls back to (5b) which also learns less accurate form-meaning pairs. This is crucial to heuristically distinguish portmanteaux from non-portmanteaux with imperfect surface distributions that often result from being blocked, possibly by portmanteaux. Consider for example the algorithm run result in (6):

(6) **Hixkaryana transitive agreement** (Derbyshire 1985), **categorized by (5)**

			1s	**1pe**	**1pi**	**2s**	**2p**	**3s**	**3p**
1s	k_1-	**1s**				k_1-	k_1-	ø-	ø-
1pe	n_2-	**1pe**				o-	o-	n_1-	n_1-
1pi	t-	**1pi**						t-	t-
2s	m-	**2s**	m-					m-	m-
2p	m-	**2p**	m-					m-	m-
3s	n_1-	**3s**	ro-	*	k_2-	o-	o-	n_1-	n_1-
3p	n_1-	**3p**	ro-	*	k_2-	o-	o-	n_1-	n_1-

t-	↔	SA[+1 +2]
ø-	↔	[+1 +sg]A→P[+3]
m-	↔	SA[–1 +2]
n_1-	↔	SP[+3]
n_2-	↔	S[+1 –2 +pl]
o-	↔	P[–1 +2]
ro-	↔	[+3]A→P[+1 +sg]
k_1-	↔	SA[+1 +sg]
k_2-	↔	P[+1 +2]

While *t*-:SA[+1 +2] and *m*-:SA[–1 +2] are 100% accurate, n_1-:SP[+3] has a 50% accurate person distribution such that it only occurs in 8 of the 16 cells that match its meaning possibly due to blocking by the perfect markers *m*-:SA[–1 +2], *t*-:SA[+1 +2], and ø-:[+1 +sg]A→P[+3]. For ø- on the other hand, neither [+1 +sg] nor [+3] make good enough meanings, both cannot easily reproduce the distribution by being blocked, so the portmanteau-meaning [+1 +sg]A→P[+3] remains the best choice.

Results For the pilot study, the learning procedure(s) have been applied to an areally and genealogically diverse sample of 26 languages. Only languages with obligatory agreement with A and P arguments on the verb were considered. Only in three of the languages marker occurrence locality interacts with portmanteau-status (Fisher's exact $p \leq .05$, cf. asterisks below), in only one (Tepehuan) in the predicted direction. As the number of cells and markers vary, and all languages have more non-local than local paradigm cells, we counted for every paradigm cell the ratio of portmanteau affixes from the total affixes in the cell. The Direct Portmanteau Hypothesis predicts that the mean portmanteau/affix ratio for local cells tends to be higher than the mean portmanteau/affix ratio for non-local cells. As shown in Figure 1, Ket, Maricopa, Reyesano, and Sahu where categorized to have no portmanteaus at all and therefore contribute no non-local vs. local distinction. The remaining languages did not show a significant correlation of non-local/local with low/high portmanteau ratio (point biserial correlation $p \not\leq .05$). In fact, 7 are analyzed to have portmanteaux only in non-local cells but only Lakhota to have them only in local cells.

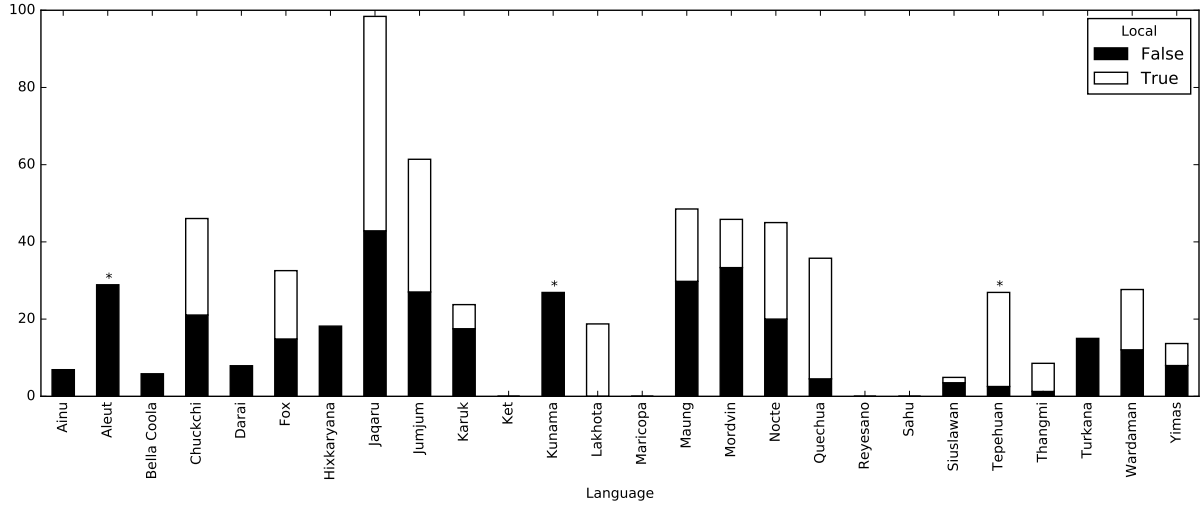


Figure 1: Mean percentage of portmanteau affixes per cell: non-local (black) vs. local (white)

References

Derbyshire, Desmond C. 1985. *Hixkaryana and Linguistic Typology*. Dallas: Summer Institute of Linguistics.

Georgi, Doreen. 2012. Deriving the Distribution of Person Portmanteaux by Relativized Probing. In *Proceedings of the North-Eastern Linguistic Society 42 (2011, Toronto)*, .

Heath, Jeffrey. 1991. Pragmatic disguise in pronominal-affix paradigms. In Frans Plank (ed.), *Economy of Inflection*, 75–89. Berlin: Mouton de Gruyter.

Heath, Jeffrey. 1998. Pragmatic Skewing in 1-2 pronominal combination in native American languages. *IJAL* 64(2). 83–104.

Lakämper, Renate & Dieter Wunderlich. 1998. Person marking in Quechua – A constraint-based minimalist analysis. *Lingua* 105. 113–148.

Nevins, Andrew. 2012. Haplological Dissimilation at Distinct Stages of Exponence. In Jochen Trommer (ed.), *The Morphology and Phonology of Exponence – The State-of-the-Art*, Oxford: Oxford University Press.

Woolford, Ellen. 2010. Active Stative Agreement in Choctaw and Lakota. In Gabriel de Ávila Othero & Sérgio de Moura Menuzzi (eds.), *Optimality Theoretic Syntax*, vol. 8:4 Special Issue of Revista Virtual de Estudos da Linguagem, 6–46. Revista Virtual de Estudos da Linguagem.

Wunderlich, Dieter. 2006. Towards a structural typology of verb classes. In Dieter Wunderlich (ed.), *Advances in the theory of the lexicon*, 57–166. Berlin: Mouton de Gruyter.

Small worlds inflectional morphology: a fragment for French conjugation

Gilles Boyé
Université Bordeaux-Montaigne & CLLE-ERSSàB
gilles.boye@u-bordeaux3.fr

Inflectional morphology descriptions usually adopt a top-down approach 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). These systems rely on the availability of a complete knowledge of the inflected lexicon for their creation.

On an opposite side of the spectrum, the unsupervised learning of inflectional morphology proposed by Lee (2014) starts with unstructured list of forms by lexemes and build-up paradigms from scratch using complex algorithms for alignment and clustering.

Small World Inflection Morphology (SWIM) adopts an intermediary stance. It starts with partial inflectional knowledge, a sparsely populated organized lexicon and builds up inflectional paradigms through analogies and network connections. In this paper, we describe a particular implementation of the model for French verb inflection.

1 Data

In this study, we consider use the forms listed in BDLex[1] (de Calmès & Pérennou, 1998) as the basis for our Gold standard of French verb inflection (6,561 verbs, 328,103 inflected forms), these forms were associated with frequencies based on Lexique3 (New et al., 2001) information for lemmas and forms in both books and films. The form list was arranged into inflectional paradigms according to the forms inflectional information creating a 51-cell paradigm where 1,690 overabundant forms were stored as cellmates.

From this lexicon, we randomly select incremental samples of 15,000 forms using the frequencies for the probability distribution to constitute training data sets representing plausible partial knowledge of the lexicon.

2 Analysis

Small world inflectional morphology aims to answer the Paradigm Cell Filling Problem (Ackerman et al., 2009) from a sparsely populated lexicon.

The analysis works in three steps. First, we extract analogies from the available lexicon then we populate paradigms by deploying all the analogies and finally we extract the best paradigm from the generated network.

2.1 Extracting analogies

Building on the Minimal Generalization Learner (Albright, 2002; Albright & Hayes, 2003) and proposals by (Bonami & Boyé, 2014; Bonami & Luís, 2014), we calculate minimally general analogies

[1] With some minor revisions on the phonological transcriptions to ensure better coherence.

between every pair of forms. For example, one of the analogies between indicative present 3sg and 3pl is the adjunction of an s on 3sg forms ending in a front unrounded vowel (e.g. fini⟶finis).[2] For every pair of forms, inflectional or derived, these analogies have partially overlapping scopes. We use existing forms to make classes of analogies and evaluate their local competition. For instance, forms like ʁəni ('to disavow') and fini ('to finish') fit only the context of the analogies in (1) and absolutely no other. This defines a class of forms. Inside this class, we count for each analogy how many times they provide the corresponding output form in our training lexicon and we calculate the result of the competition in the class as the distribution of pairs among the analogies characterizing the class in (2). Note, in this case, that the third analogy is never used for forms in this class even though its context targets them.

(1) a. Ø⟶Ø/ X – (ʁəni⟶ʁəni, fini⟶*fini)

b. Ø⟶s / X[iɛea] – (ʁəni⟶*ʁənis, fini⟶finis)

c. Ø⟶t / X[ptbdfsvzmnrɛəɒaɔɛ̃œ̃ə̃ɑ̃][jrwɥiyɛeøœəauɔoɛ̃œ̃ə̃ɑ̃] – (ʁəni⟶*ʁənit, fini⟶*finit)

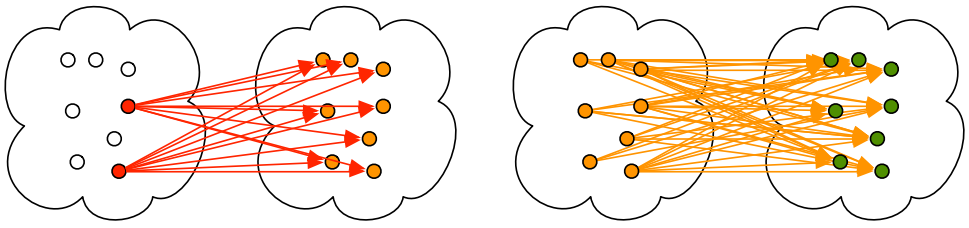
(2) {Ø⟶Ø, Ø⟶s, Ø⟶t} ⟶(20.69%, 79.31%, 0%)

2.2 Populating cells

The generation of paradigms for each lexical entry in the training dataset is based on a two-round mechanism.

In the first round, for a given lexeme, we use every known form to propose candidates based on extracted analogy-classes. Starting from a particular form in a cell, we generate candidates in every cell[3] according to its analogy class distribution. If a form corresponds to an unregistered analogy class, the rules outputs are considered equiprobable. If a form does not correspond to the context of any rule, no candidate is generated. As the generated candidates in target cells receive votes originating from different forms and cells, we standardize the local score of every form in the target cells. Because our training lexicon includes only a limited number of forms and therefore cells per lexeme (red dots in Fig. 1), only the analogy classes (red arrows) pertaining to the known cells play a role in this first round.

On the second round, we apply the same strategy starting from the candidates created in the first round. All cells (orange dots) usually contain forms from the first round and we use this information to propose candidate forms for every cell (green dots) in the paradigm again. In a cell, the candidates proposed by each form are weighted according to the form score from the first round. At this stage all the knowledge about analogies between all cells (orange arrows) is mobilized.



First round Second round

Figure 1: The two-round mechanism

[2] In our notation : Ø⟶s / X[iɛɛa] —

[3] The original cell itself is included as to capture the systematic overabundance.

2.3 Harvesting paradigms

At the end of the second round, all relevant analogies have been used and we have obtained a massively overabundant result. From this anarchic configuration, we extract complete paradigms with the inflectional forms of the verbs using the following morphological hypothesis:

(3) paradigms are structures were all members are analogically related

A candidate paradigm is a set of form candidates where all the forms in all the cells support each other with mutual analogies. In Fig. 2, among the forms in the various cells (green dots), the blue ones form a clique (completely interconnected network in blue) where all forms have co-opted the others.

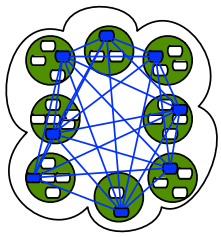


Figure 2: Extracting a candidate paradigm

3 Results

As seen in Table 1, starting from a probabilistic sample of 15,000 inflected forms (4.6%), in a typical run, SWIM produces about 60,000 predicted forms to reach about 175,000 predicted forms from the initial 75,000 (22.9%) with the last sample. Most of the errors seem related to French phonological opacities with yod sequences such as ɲ+j⟶ɲ) or classification errors on ambiguous data in the early stages.

At the core level, SWIM relies solely on binary relations between forms but at the paradigm prediction level, it makes use of multiterm relations through its clique retrieval mechanism. The resulting predictions are based on the known lexicon rather than on preexisting complete knowledge of inflectional classes (see for example Stump & Finkel, 2013). The essential morphological assumption underlying this is that all forms in the paradigm of a given lexeme should be related by predictable oriented analogical relations allowing to harvest paradigms from small worlds of form-cells.

Known forms	New forms	Precision	Recall	F-Score
15,000	62816	88.8%	46.6%	61.1%
30,000	110861	92.3%	61.0%	73.5%
45,000	133245	93.5%	66.6%	77.8%
60,000	147135	94.5%	71.1%	81.2%
75,000	174739	97.5%	86.3%	91.5%

Table 1: SWIM result for French

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 Scholarship Online.

Albright, Adam. 2002. *The identification of bases in morphological paradigms*: UCLA dissertation.

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in english past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Anderson, Stephen R. 1992. *A-Morphous Morphology* (Cambridge Studies in Linguistics 62). Cambridge University Press.

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

Bonami, Olivier & Ana R. Luís. 2014. Sur la morphologie implicative dans la conjugaison du portugais : une étude quantitative. *Mémoires de la Société de Linguistique de Paris* 22. 111–151.

Brown, Dunstan & Andrew Hippisley. 2012. *Network morphology: A defaults-based theory of word structure*, vol. 133 Cambridge Studies in Linguistics. Cambridge University Press.

de Calmès, Martine & Guy Pérennou. 1998. Bdlex : a lexicon for spoken and written french. In *1st international conference on language resources and evaluation*, 1129–1136. Grenade: ELRA.

Corbett, Greville G & Norman M Fraser. 1993. Network morphology: a datr account of russian nominal inflection. *Journal of Linguistics* 29(01). 113–142.

Kilani-Schoch, Marianne & Wolfgang U. Dressler. 2005. *Morphologie naturelle et flexion du verbe français*. Tübingen: Gunter Narr Verlag.

Lee, Jackson L. 2014. Automatic morphological alignment and clustering. Tech. Rep. TR-2014-07 Department of Computer Science, University of Chicago.

New, Boris, C. Pallier, L. Ferrand & R. Matos. 2001. Une base de données lexicales du français contemporain sur internet: Lexique. *L'Année Psychologique* 101. 447–462.

Stump, Gregory & Raphael A Finkel. 2013. *Morphological typology: From word to paradigm*, vol. 138. Cambridge University Press.

Stump, Gregory T. 2001. *Inflectional morphology*. Cambridge University Press.

Reduplication in an implemented HPSG of Hausa

Berthold Crysmann
CNRS, Laboratoire de linguistique formelle, Paris

In this talk, I shall discuss the treatment of reduplication, and most notably total reduplication in an implemented grammar of Hausa. I shall compare several patterns observed in the language and conclude, on the basis of segmental and suprasegmental properties alike, that partial and total reduplication operate on entirely different structures: phonological representations of limited size (max: CVC) in the case of partial reduplication, and morphological representations, such as stems in the case of total reduplication. Finally, I shall show that partial reduplication can readily be captured by the morphophonological means offered by the underlying formalism, yet argue that total reduplication is best modelled on the basis of binary rules. I shall propose a formalisation that keeps all of the linguistic generalisation within the grammar proper, yet ensures compositionality and reversibility by means of token-rewriting.

1 Patterns of reduplication in Hausa

1.1 Partial reduplication

Alongside standard affixation, the grammar of Hausa makes good use of partial reduplication for inflectional (and derivational) purposes: e.g. many of the language's plural formation patterns witness partial reduplication at least in some sub-patterns.

Noun class 1 constitutes one of the most productive plural patterns: as shown in (1), formation of the plural involves reduplication of the last consonant of the base and vocalisation with a fixed *-ō-X-ī* pattern. Regardless of the tonal make-up of the base, an all-H melody is assigned.

(1) a. gyàlè (M.SG) — gyalōlī (PL) 'shawl'
b. bindigà̀ (F.SG) — bindigōgī (PL) 'gun' (Newman, 2000, 432)

A segmentally more complex case is provided by a subpattern of plural class 8: plural formation involves affixation of *-akī* plus insertion of a reduplicative CaC syllable before the stem-final consonant that copies that consonant onto both C slots. As far as tone is concerned, we find again a holistic pattern (H-L-H-H) that is assigned to the derived form, regardless of the tones of the base.

(2) ɓērā (M.SG) — ɓēràrrakī (PL) 'mouse'

A slightly different pattern, at least in tonal respects, is found with pluractionals: the most regular pattern here involves prefixation of a CVC syllable where both consonantal slots copy the first consonant of the base. Some pluractionals formed this way have an alternate form that copies the second base consonant onto the coda of the reduplicant (cf. Newman, 2000, 425).

(3) a. darnàcē — daddarnàcē 'fence in'
b. kāwō — kakkāwō 'bring' (Newman, 2000, 424)

As for tone, the entire form receives whatever would be the regular tone pattern in that grade with respect to the syllable count: since the Hausa grade system (Parsons, 1960) only distinguishes tone patterns up to trisyllabic words, quadrisyllabic pluractionals just spread the left-most tone onto the reduplicant.

1.2 Total reduplication

Amongst the productive reduplicative patterns of the language, total reduplication can count as the more restricted option: while partial reduplication is attested with both of the open class categories of the language, i.e. with nouns and verbs, total reduplication is confined to non-verbal categories in general. Nevertheless, outside the verbal domain, it is attested somewhat productively with both nouns and adjectives.

While total reduplication may also figure in certain lexical classes in the singular (e.g. ideophonic adjectives, adverbs, and nouns), it is clearly more productive in the encoding of notions of plurality, relating to individuals, events, and properties.

1.2.1 Noun class 12

This class, while limited in lexical scope, still constitutes a morphological, rather than lexical pattern. It is typically used for plural formation of recent loans (Newman, 2000).

(4) a. bȍyi (SG) — bȍyi bȍyi (PL) 'house boy'
b. nâs (SG) — nâs nâs (PL) 'nurse' (Newman, 2000, 457)

Despite their more recent introduction, these nouns systematically get subjected to the productive inflectional rules of the language, phonotactics permitting: thus, standard possessive and previous reference marking do apply to reduplicated plurals as well.

(5) jōjì jōjì-n-mù 'our judges'

So, whatever copying of surface segments may apply here, it is clear that it ignores "outer" inflectional markings. Similarly, tonal specification of the reduplicant (left) also faithfully reflects the tones of the base (right), modulo tonal changes effected by subsequent inflectional morphology. The same holds true of the metrical make-up (cf. (4)).

1.2.2 Frequentatives: Noun class 13

Compared to the previous class, frequentatives constitute a way more productive pattern. As to their form, Newman (2000, p. 196) describes them as pseudo-plural deverbal nouns with a reduplicated structure. Consider the examples in (6): base and reduplicant with frequentatives are characterised by short final /e/ and a L^+ H each. Segmental material, including vowel length is identical between base

and reduplicant. Note that there is also no restriction on syllable count, i.e., we find both disyllabic and trisyllabic frequentatives.

(6) a. jēfà — jèfe jèfe
'throw' (Newman, 2000, 197)
b. dafà — dàfe dàfe
'prepare/cook' (Newman, 2000, 198)
c. tàmbayà — tàmbàye tàmbàye
'ask' (Newman, 2000, 196)

Besides encoding plurality of individuals (a fairly small class), frequentatives are much more commonly used to denote repeated action of dynamic and verbal nouns. In this use they are comparable, to some extent, to pluractionals, which are formed using partial reduplication on verbs.

(7) mātā sunà dad-dàfà / dàfe-dàfe-n
women 3PL.CONT RED-cook RED-cooking-LNK
àbinci
food
'the women are cooking up food a lot'
(Newman, 2000, 198)

Since frequentatives are deverbal and may denote activities, they can often take object complements. Being formally nouns, they are marked with the linker *-n* in this case (see 7b). In case their complement is pronominalised, it is realised as a bound low tone affix from the possessive set. Thus, the base of frequentatives may undergo the same additional inflectional morphology we have observed with class 12 plural nouns.

Two properties clearly set frequentatives apart from noun class 12: first, they are far more productive, and second, the base for reduplication is not a root, but itself a derived form, making lexicalisation of these forms a suboptimal option.

1.2.3 Augmentatives: Noun class 14

The third class I am going to study in this paper concerns plural formation of augmentative adjectives. Newman (2000, p. 25) describes this class as "good-sized". We shall focus on class B augmentatives (Parsons, 1963), the most productive subclass according to Newman (2000, p. 75). As witnessed by the data in (8), the (singular) base itself is a partially reduplicated form, featuring a reduplicated final vowel and reduplication of the final consonant. Tonally, they follow a fixed pattern, i.e. H^+ L.

(8) a. ribɗēdè (M.SG) — ribɗā-rìbɗà (PL)
'huge and bulky'
b. zungurērè (M.SG) — zungurā-zùngùrà (PL)
'long, tall' (Newman, 2000, 76)

In the plural, however, we find a total reduplication pattern, based on the singular form without the final partial reduplication: while the metrical structure is preserved, the plural substitutes a final /ā/ in both the base and the reduplicant, and assigns an all-L melody to the base, and an all-H melody to the reduplicant.

1.3 Comparing total and partial reduplication

Partial and total reduplication have some distinctive properties that clearly set them apart: first, most obviously, with partial reduplication the size of the reduplicant is more strictly bounded, never involving more than at most three segments of the base. In addition to quantity, quality is also much more finely constrained in the case of partial reduplication: many partially reduplicative processes in Hausa target the root consonant only, using a fixed vocalisation pattern.

As for metrical information, these two morphological processes are also clearly distinct: while vowel length on a reduplicant always corresponds to some fixed templatic pattern in the case of partial reduplication, showing no alternation depending on the base, with total reduplication, it is determined essentially by the base, giving rise to reduplicants with different length specifications (cf. *jōjì jōjì* (long-short) vs. *sìkêt sìkêt* (short-short)).

Tonally, total reduplication also differs quite markedly from all other derivational and inflectional processes of the language: in fact, it constitutes the only process where entire melodies get replicated on the reduplicant. Equally special is the status of tone with class b augmentative adjectives: although tone is not copied, the assignment of two spreading tones (H^+ L^+) with different polarity is unattested in Hausa outside of total reduplication. Partial reduplication, by contrast, is tonally much more well-integrated into the system: either the reduplicative pattern is part of a formation that holistically assigns a tonal melody to the entire form (e.g. most noun plurals), or the reduplicant just receives its tonal specification by spreading of the adjacent tone of the base, as illustrated by the pluractionals in (3).

Finally, while partial reduplication may be peripheral or non-peripheral (e.g. *ɓērā — ɓēràrrakī* 'mouse/mice'), total reduplication is always peripheral.

Taking stock, it is clear that these two reduplicative processes are clearly governed by different principles: while partial reduplication is best regarded as affixation of a fixed metrical template where the identity of individual segmental slots is determined relative to individual segments of the base, total reduplication establishes identity between reduplicant and base not on the basis of circumscribed phonological entities, but rather on the basis of a minimal morphological word.

2 Implementation

The present work is part of an ongoing grammar implementation effort of the Hausa language developed within the framework of Head-driven Phrase Structure Grammar (Pollard & Sag, 1994). The grammar is developed using the DELPH-IN processing systems, such as the LKB (Copestake, 2002), Pet (Callmeier, 2000), and ace (Crysmann & Packard, 2012).

The grammar already has interesting syntactic coverage, offering a treatment of complementation, modification and coordination, as well wh-extraction and relativisation, including both gap and resumptive strategies. Most importantly, the grammar covers the rich inventory of Hausa plural classes, as well as the verbal grade system (Parsons, 1960), including morphosyntactic alternation according to complement type, for verbs, nouns, adjectives and prepositions. Furthermore, the grammar has been designed from the ground up to support a systematic treatment of tonal and metrical information, providing the possibility for assigning holistic tone melodies in the course of morphological derivations (Crysmann, 2009).

2.1 Partial reduplication

Owing to its very local and circumscribed phonological effect, the treatment of partial reduplication is readily supported by the morphophonological machinery, which has been developed as a variant of string unification (Calder, 1989). These orthographemic rules (see 12 for an example) specify prefixal or suffixal substitution patterns where the LHS describes a match and the RHS a substitution. Character classes (prefixed by !) function as variables whose match is restricted to a particular definable class of segments, thus they simultaneously allow for abstracting over classes of segments and binding of a matching segment to a variable (LHS). Reference to these character classes on the RHS actually copies the concrete instantiation to the output. Gemination and partial reduplication are straightforwardly expressed by reusing a bound character variable multiple times on the RHS of a rule. In case of reduplication of multiple and distinct base segments, we need to provide multiple variables for binding. Thus, this approach works very well for reduplication of circumscribed amounts of material, but will certainly not scale up to reduplication involving strings of either greater or variable length.

Since the grammar employs a systematic distinction between segmental and supra-segmental phonology already (Crysmann, 2009), assignment of *definite* tonal and metrical representations, as is the case with partial reduplication, is not of great concern here.

2.2 Reversibility, compositionality and total reduplication

A central assumptions behind HPSG and MRS is the commitment to monotonicity and rule-by-rule compositionality. This perspective is part of an overall perspective on grammars as declarative knowledge sources that are independent of the direction of processing: i.e. the same grammar is used for parsing and generation.

2.2.1 Baseline approach (parsing only)

Possibly one of the most straightforward approaches to total reduplication is in terms of a binary (morpho)syntactic rule that combines two like lexical entries and assigns the combination the required (plural or pluractional) semantics. While this is fine on the string level, double invocation of the same lexical item (or a form derived from it) in the same derivation tree will inevitably insert that entry's semantics into the derivation, giving us, in the case of *nâs nâs* 'nurses' two *nurse* predicates instead of one. Thus, in order to get the appropriate sentence semantics, it is necessary for any such rule to ignore part of the semantic contribution from the daughters, in other words, we would need that rule to be non-compositional. This can of course be done by way of refining semantic composition on the reduplication binary rule, such as to merely project from the base, ignoring the reduplicant.

As for parsing, one apparently can get away with non-compositionality, however, once we want to generate, more strict assumptions about compositionality apply: both generators (LKB and ace) crucially rely on the fact that every semantic predicate can only be expressed once. While it is possible to inject additional instances of a lexical entry into the chart, using essentially the same device as is used to insert chart elements for semantically empty lexical entries (e.g. complementisers), this approach turns out to be severely limited, since it requires a separate triggering rule for every single lexical entry. While this is feasible for highly unproductive classes (like noun class 12), it becomes utterly unmanageable with the more productive frequentatives and augmentatives.

2.2.2 A reversible approach

As a solution to the problem just mentioned, I shall propose to use a single generic reduplicative lexical entry instead and let the grammar ensure segmental and suprasegmental identity. If we do postulate such a generic entry, we may as well assign it empty semantics, i.e. regard it as an abstract free morph, which will enable us to return to compositional semantics. Furthermore, if this generic reduplicant is not restricted to any particular predicate, we can trigger easily, in generation, on a very general property of input semantics, viz. plurality.

(9)
$$\begin{bmatrix} \textsc{stem} & \langle \textsc{\_\_redup\_\_} \rangle \\ \textsc{cont} & \begin{bmatrix} \textsc{rels} & \langle\,\rangle \\ \textsc{hcons} & \langle\,\rangle \end{bmatrix} \end{bmatrix}$$

Once we use a generic entry, however, we will not have any item-specific stem form anymore that could be use in lexical look-up during parse chart initialisation. Fortunately, both run-time systems for DELPH-IN grammars, i.e. Pet (Callmeier et al., 2004) and Ace (Crysmann & Packard, 2012), feature a token chart with chart mapping to facilitate inter alia the treatment of unknown words, allowing for the possibility to inject alternative hypothesis before parsing proper (Adolphs et al., 2008). Thus, we insert an additional reduplicant hypothesis for every non-final token chart element.The token rewrite rule saves the surface string information in a designated feature (MORPH.–REDUP.–STEM), while replacing the string proper with that of the generic entry (__REDUP__). Suprasegmental information that may have been extracted from the surface string and represented in tone and length features will be carried along as well. As a result, the parser will try to construct analyses with either the native entry, or the hypothesised reduplicant. All that token rewriting does is feed potential reduplicants to lexical look-up, and preserve information about surface form.

Licensing of total reduplication is done entirely within the grammar: morphological rules on the base instantiate the relevant constraints for the reduplicant using the feature MORPH.–REDUP, and a binary rule that combines base and reduplicant and imposes any constraints regarding shape that the base may have for its "copy": in particular, we equate stem form and impose whatever suprasegmental information the base holds in MORPH.–SUPRA directly onto the relevant representation of the reduplicant.

(10)
$$\begin{bmatrix} \textsc{dtrs} \left\langle \begin{bmatrix} \textsc{supra} & \begin{bmatrix} \textsc{tone} & \boxed{t} \\ \textsc{len} & \boxed{l} \end{bmatrix} \\ \textsc{morph} & \begin{bmatrix} -\textsc{redup} & \begin{bmatrix} -\textsc{stem} & \boxed{s} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \textsc{morph} & \begin{bmatrix} -\textsc{redup} & \begin{bmatrix} -\textsc{stem} & \boxed{s} \\ -\textsc{supra} & \begin{bmatrix} \textsc{tone} & \boxed{t} \\ \textsc{len} & \boxed{l} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Relevant likeness constraints are imposed by the inflectional rules operating on the base, which also take care of any changes in surface form: in the case of noun plural class 12, this is the identity function, but in the case of frequentatives and augmentatives, this rule of course effects non-trivial morphophonological changes with respect to the root. For Hausa, the level of gender/number

inflection appears to be the right place to enforce constraints on likeness in the paradigms that require total reduplication. Following Bonami & Webelhuth (2013), I shall regard total reduplication as a case of a local morphologically periphrastic construction: inflectional rules on the base perform local inflection, as well as memorise any form-oriented properties regarding the reduplicant in a selectional feature (here: MORPH.−REDUP). For both morphological rule examples below, identity requirements include not only token identity of segmental form (STEM), but also complete sharing of length and tone information. However, while in noun class 12, all information is carried over directly from the lexical root, the inflectional rule for the plural of class B augmentatives (noun class 14) performs adjustments on all three levels: clipping of partial reduplication on the root (e.g. *zungurere* ↦ *zungura*), adjusting the metrical structure (LEN), and assigning differing spreading tones to the base (SUPRA.TONE) and the reduplicant (MORPH.−REDUP.SUPRA.TONE).

(11)
$$\begin{bmatrix}
\textsc{stem} & \langle \boxed{s} \rangle \\
\textsc{supra} & \begin{bmatrix} \textsc{tone} & \boxed{t} \\ \textsc{len} & \boxed{l} \end{bmatrix} \\
\textsc{synsem} & \left[\textsc{l.cont.hook.index} \left[\textsc{num} \quad \mathit{pl}\right]\right] \\
\textsc{morph} & \begin{bmatrix} \textsc{mclass} & \boxed{c}\ \mathit{12} \\ -\textsc{redup} & \begin{bmatrix} -\textsc{stem} & \boxed{s} \\ -\textsc{supra} & \begin{bmatrix} \textsc{tone} & \boxed{t} \\ \textsc{len} & \boxed{l} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\textsc{dtr} & \begin{bmatrix} \textsc{stem} & \langle \boxed{s} \rangle \\ \textsc{morph} & \left[\textsc{mclass} \quad \boxed{c}\right] \end{bmatrix}
\end{bmatrix}$$

(12)
```
%suffix (!ce!ce !ca)
```
$$\begin{bmatrix}
\textsc{stem} & \langle \boxed{s} \rangle \\
\textsc{supra} & \begin{bmatrix} \textsc{tone} & \mathit{l^*\text{-}list} \\ \textsc{len} & \boxed{l} \end{bmatrix} \\
\textsc{synsem} & \left[\textsc{l.cont.hook.index} \left[\textsc{num} \quad \mathit{pl}\right]\right] \\
\textsc{morph} & \begin{bmatrix} \textsc{mclass} & \boxed{c}\ \mathit{14} \\ -\textsc{redup} & \begin{bmatrix} -\textsc{stem} & \boxed{s} \\ -\textsc{supra} & \begin{bmatrix} \textsc{tone} & \mathit{h^*\text{-}list} \\ \textsc{len} & \boxed{l} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\textsc{dtr} & \begin{bmatrix} \textsc{supra} & \begin{bmatrix} \textsc{tone} & \mathit{h^*\text{-}l\text{-}list} \\ \textsc{len} & \langle \mathit{long} \mid \boxed{l} \rangle \end{bmatrix} \\ \textsc{morph} & \left[\textsc{mclass} \quad \boxed{c}\right] \end{bmatrix}
\end{bmatrix}$$

Information regarding the reduplicant is carried up along the cascade of lexical rule, to be matched with the properties of the (generic) reduplicant. As should be clear, all linguistic information pertaining to reduplication is contained in the grammar, a major design decision. The only task performed by token mapping is normalisation to a specific canonical form for lexical look-up. True surface of the reduplicant is always preserved and forms the basis for identity requirements.

In generation, triggering of reduplicants becomes trivial, since there is only a single entry contingent on plurality. Given that all linguistic information is already in the grammar, and hence, the derivations, it is, of course, trivial to replace the generic placeholder phonology with the true phonology of the reduplicant. To this end, we rely on the same token mapping technology as we do in parsing.

Keeping parsing and generation parallel down to the single rule or lexical entry is not only preferable from a linguistic or grammar development point of view: it also bears the further benefit of reversible treebanks: as long as derivations are truly reversible, disambiguated parse treebanks can be used to automatically annotate generation treebanks (Velldal & Oepen, 2005).

3 Conclusion

I have discussed patterns of reduplication in Hausa and established — based on segmental, metrical, and tonal evidence — that partial reduplication observes very circumscribed phonological conditions, whereas total reduplication does not, and is therefore best understood as a process involving morphological entities. In the context of an implemented grammar of Hausa, I have argued that the theoretical difference between the two reduplication processes is replicated at the level of implementation: while partial reduplication is straightforwardly integrated with standard inflectional rules of the grammar, total reduplication presents challenges to compositionality and reversibility. I have shown how a minimal adjustment in terms of token mapping to a generic reduplicative entry can solve both problems, while keeping all linguistic constraints readily within the grammar.

References

Adolphs, Peter, Stephan Oepen, Ulrich Callmeier, Berthold Crysmann, Dan Flickinger & Bernd Kiefer. 2008. Some fine points of hybrid natural language parsing. In Nicoletta Calzolari et al. (ed.), *Proc. LREC'08*, Marrakech, Morocco: ELRA.

Bonami, Olivier & Gert Webelhuth. 2013. The phrase-structural diversity of periphrasis: a lexicalist account. In Marina Chumakina & Greville G. Corbett (eds.), *Periphrasis: The role of syntax and morphology in paradigms*, 141–167. Oxford: Oxford University Press.

Calder, Jonathan. 1989. Paradigmatic morphology. In *Proc. eacl*, 58–65. Manchester: ACL.

Callmeier, Ulrich. 2000. PET — a platform for experimentation with efficient HPSG processing techniques. *Journal of Natural Language Engineering* 6(1). 99–108.

Callmeier, Ulrich, Andreas Eisele, Ulrich Schäfer & Melanie Siegel. 2004. The DeepThought core architecture framework. In Maria Teresa et al. Lino (ed.), *Proc. LREC 2004*, 1205–1208. ELRA Lisbon, Portugal: ELRA.

Copestake, Ann. 2002. *Implementing typed feature structure grammars*. Stanford: CSLI Publications.

Crysmann, Berthold. 2009. Autosegmental representations in an HPSG for Hausa. In *Proc. GEAF 2009*, 28–36. ACL.

Crysmann, Berthold & Woodley Packard. 2012. Towards efficient HPSG generation for German, a non-configurational language. In *Proc. COLING 2012*, 695–710. Mumbai, India.

Newman, Paul. 2000. *The Hausa language. an encyclopedic reference grammar*. New Haven, CT: Yale University Press.

Parsons, Fred. 1963. The operation of gender in Hausa: Stabilizer, dependent nominals and qualifiers. *African Language Studies* 4. 166–207.

Parsons, Fred W. 1960. The verbal system in Hausa. *Afrika und Übersee* 44. 1–36.

Pollard, Carl & Ivan Sag. 1994. *Head–Driven Phrase Structure Grammar*. Stanford: CSLI.

Velldal, Erik & Stephan Oepen. 2005. Maximum entropy models for realization ranking. In *Proc. MT-Summit (X), Phuket, Thailand*, .

Learning from the Computational Modeling of Plains Cree Verbs

Atticus G. Harrigan[1], Lene Antonsen[2], Antti Arppe[1], Dustin Bowers[1], Katie Schmirler[1], Trond Trosterud[2], Arok Wolvengrey[3]

This paper describes a computational model for Plains Cree (ISO-692: crk, Algonquian, western Canada), a polysynthetic, morphologically complex language, using the finite-state transducer (FST) framework. Specifically, this paper presents modeling Plains Cree verbal morphology, as prior work has focused on nouns in this language (Snoek et al. 2014). We will discuss simple affixation with prefixes and suffixes, and illustrate our approach to modeling circumfixation. Furthermore, we will present how we implemented reduplicative structures. Finally, we evaluate the accuracy of our model through the analysis of a small corpus, which allows us to also examine and learn how the aforementioned individual morphological processes, as implemented computationally, interact with each other, possibly in unexpected but informative ways.

Plains Cree is generally considered an endangered language - in many, if not most, communities where Cree is spoken, children are learning English as a first language, and encounter Cree only in the language classroom. Like many North American languages, Plains Cree lacked a written tradition until European contact. As a result, Plains Cree corpora and texts are few in number and small in size, though still substantial for an endangered language. What do exist are several quite comprehensive grammatical descriptions covering Plains Cree morphology (Wolfart, 1973; Okimâsis, 2004; Wolvengrey, 2011). With these descriptions, we are able to model general morphological phenomena in Plains Cree, though without extensive corpora we must rely only on those forms that are well documented and deemed important enough for inclusion by the aforementioned linguists. A key challenge is that while individual morphological structures or processes are typically quite well described in how they work in general, the details of all their possible combinations, as well as constraints on paradigms (e.g. verbs which make semantically or pragmatically sense only with singular or plural subjects/actors) are not explicitly addressed - this would be where large corpora, or alternatively, extensive consultation of fluent native speakers of Plains Cree, would be worthwhile, but often not practically feasible.

As a polysynthetic language (Wolvengrey, 2011), Plains Cree exhibits substantial morphological complexity and a variety of morphological processes, most notably in verbs.. Plains Cree verbs are marked for both actor (subject) and goal (object) agreement as well as tense and aspect (Wolvengrey, 2011). Argument agreement is generally marked through a system of circumfixes, wherein a verbal stem is surrounded by a syntactically and semantically related pair of prefix and suffix, such as *ni- ... -nân* for the (animate) first person plural exclusive actor of an intransitive verb, in *ni-nipâ-nân* 'we (but not you) sleep', or *ki-...-wâw* for both the animate second person singular actor and an animate third person singular goal of a transitive verb, in *ki-wâpam-âw* 'you see him' (Wolvengrey, 2011). Each verb can occur in two orders, which are distinguished by which affixes they may take. The independent order takes the person prefixes *ni-* '1st person', *ki-* '2nd person', and *ø-* '3rd person' (singular or plural, depending on the corresponding suffix element), while verbs in the conjunct order may only be prefixed with the conjunct order prefixes *ê-* or *kâ-*, with person indicated by suffixes alone (Wolvengrey, 2011). Other circumfixes, such as *wîci-...-m* (meaning "with me") add adjunctive meanings to the verb. Further, Plains Cree verbs can include a class of prefixes known as *preverbs*. Preverbs occur before verbal stems (but after prefixal person and tense marking), and are stackable, with each new preverb modifying the

[1] University of Alberta
[2] UiT Norges Arktiske Universitet
[3] First Nations University of Canada

concatenative meaning of the verbal complex (cf. Wolvengrey, 2015), see *-nohtê-* 'want' in (1), and in combination with *-pê-* 'come' in (2). While preverbs are often present in Plains Cree constructions, they are not mandatory. Tense markings, often referred to as grammatical preverbs (*ki-* for past, *kâ-* for future definite, and *wî-* for future intentional), precede the non-grammatical preverbs, while the iterative and continuative reduplicative morphemes are closest to the verb stem (Bakker, 2006). Nominal incorporation occurs in Plains Cree, though our model does not yet incorporate this feature.

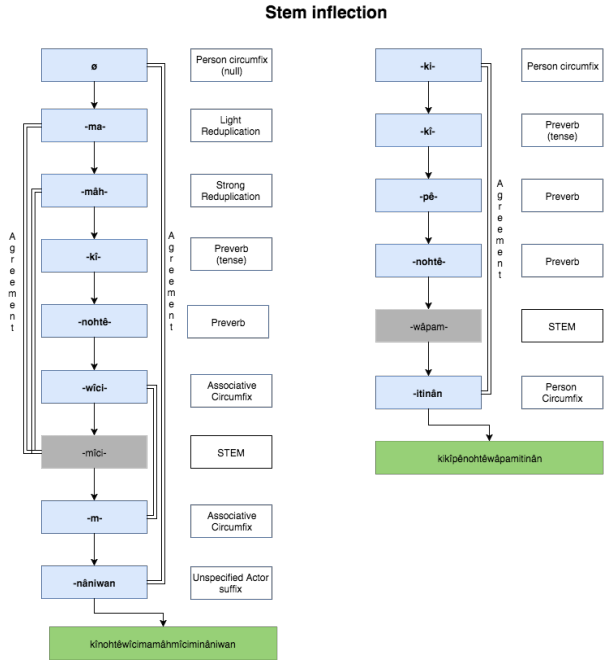
(1) kînohtêwîcimâmahmîciminâniwan
kî-nohtê-wîci-mâ-mah-mîci-m-i-nâniwan
PST-want-ASSOC.-REDUP-eat.NI-ASSOC-EPENTH-UNSPEC.ACTOR
'Someone always really wanted to eat it with me'

(2) kikîpênohtêwâpamitinâwâw
ki-kî-pê-nohtê-wâpam-it-inâwâw
2-PST-come-want-see.NA-INV-2PL.OBJ.1SG.SUBJ
'I wanted to come and see you.'

We modelled the Plains Cree verb with a Finite-State Transducer (FST) (cf. Beesley & Karttunen 2003), using the (1) *lexc* formalism for general morpheme concatenation, (2) the flag diacritic notation in *lexc* for implementing feature agreement in circumfixes, as well as constraining allowable morpheme sets for the various orders, i.e. independent and conjunct, and (3) the *twolc* formalism for general morphophonological regular alternations triggered by the immediate ortho-phonemic context. First is the assignment of *order* which restricts which morphemes can be attached to a verb (e.g. the grammatical preverb *ê-* never occurs in the independent form, nor the person prefixes *ni-* and *ki-* with conjunct forms), and is followed by the assignment of grammatical person. Grammatical person is assigned through flag diacritics, wherein person morphemes (such as *ki-* in (2)) are affixed to the verb stem along with specifying a flag diacritic that requires, or allows for, the addition of one or more specific suffixes (e.g. *-itinâwâw* in (2)) later in the morpheme concatenation process. Next is the assignment of non-grammatical preverbs, which can be currently stacked one after the other multiple times without restriction. Circumfixal elements other than grammatical person, i.e. the associative *-wîci-...-m-*, are dealt with next and make use of flag diacritics as before. Before moving on to post-verbal stem elements, reduplication is addressed by assigning a special character followed by *-â-* (in weak reduplication) or *-ah-* (in strong reduplication) directly before the verb stem. This special character is then changed to match the first character of the verb stem via a set of *twolc* rules. The conjugation class of the verb stem then determines which set of suffixes can be attached. Here, if a previously set flag diacritic indicates that a circumfix is present, suffixal elements of circumfixes are added. After affixal concatenation, morpho-phonological processes specified with the *twolc* formalism are applied over the entire verb. In (1), for example, we see the epenthesis of *-i-*, as Cree does not normally allow /mn/ consonant clusters. In cases where two vowels occur together (as in *ni-atoskan*), *twolc* morphophonological rules are also responsible for epenthesis of *-t-*. Figure 2 offers a visual representation of the FST's verbal analysis and generation process.

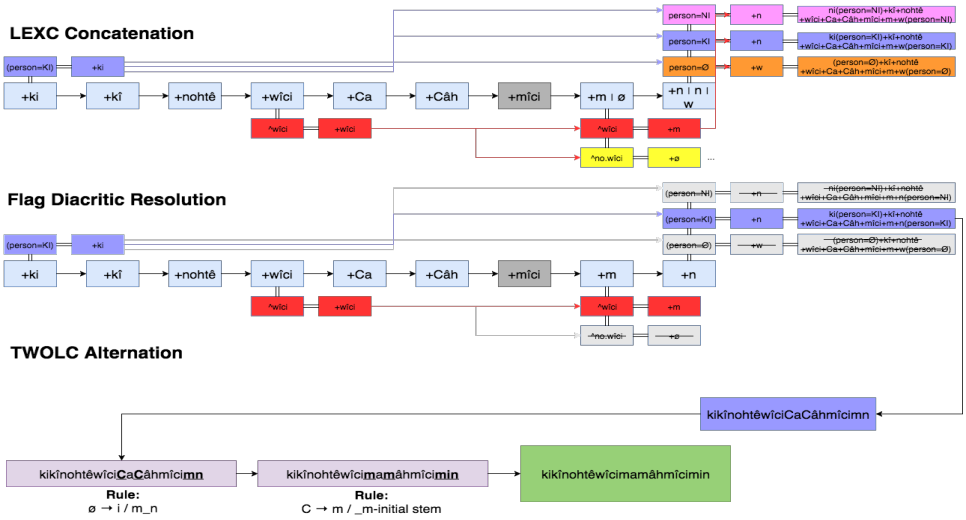
The FST affixations are not always equivalent to the morphological breakdown seen in example (1) or (2): indeed, we see that our system treats *-itinân* as a single suffix representing both inverse marking and argument agreement. This is done mostly for the purposes of parsimony.

Figure 1. Morphological structure for examples (1) and (2)



Because the inverse morpheme indicates a pairing antithetical to the Plains Cree animacy hierarchy of 2>1>3>3-obviative>3-further-obviative (i.e. a subject acts on an object higher on the hierarchy), the inverse morpheme *-iti-* must be included which, in turn, requires a particular set of person suffixal elements. Because we can enumerate the handful of cases where this will occur, and because *-iti-* occurs only with these cases, we treat *-iti-* as part of the suffix element of the subject/object circumfix.

Figure 2. FST generation for *kikînohtêwîcimamâhmîcimin* "you wanted to eat it with me"



Further, in words like *mowêw* 'see someone (animate)', we can in principle see three distinct morphemes: *mow-, -ê- and -w*. In this case, *-ê-* is considered to act as a 'theme' sign (Wolvengrey, 2011), as with *-iti-* seen in (2). Here we treat *-êw* as a single non-decomposed chunk for the same reasons previously described.

As a very preliminary test of the coverage and accuracy of our model, we analyzed a subset corpus of our roughly 100,000 word corpus, amounting to 1327 word form types, with our FST. These word form types were manually analyzed and these verified analyses were compared with FST analyses. Of the 1327 word forms, 951 (72%) were analyzed as properly written Plains Cree words, with the rest being primarily misspelled or contracted Cree words or their fragments (317: 23.8%), English words and names (52: 3.9%), and Arabic or Roman numerals (44: 3.3%). Among the 951 Cree words, 511 were verb form types (representing 617 verb tokens), and of these 405 (79.2%) received at least one correct analysis (representing 500 verb tokens). In those verbs that did receive an analysis, some issues still remained. Of the unanalyzed forms, 119 verb types were stems missing from our lexicon, either due to dialectical differences or simply under documentation in our reference material. A further 77 types were the results of similarly underdocumented/underimplemented morphology. These are issues we are continually working on as more data is observed, collected, and conglomerated from various sources.

This analysis does not aim to model a structure of the mind. Although recent literature has tended towards an amorphous approach to morphological analysis (e.g. Anderson, 1992; Crysmann & Bonami, 2015), this papers presents a model that allows for parsimonious and systematic computational implementation, a task that we have found best reached via morphous morphology. Although the current FST is being implemented in language learning materials and a smart dictionary, there still exists much room for improvement. As next stages in our computational modeling work, we intend to explore and implement further morphological processes such as noun incorporation and nominalization to increase coverage, as well as study the results of the corpus analysis to learn aspects of our model that should be restricted in their productivity or generality, in order to reduce unlikely or incorrect analyses.

References

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

Bakker, Peter. 2006. Algonquian Verb Structure: Plains Cree. In Studies in the Verbal Morphology of Languages of the Americas, eds. G. Rowick and E. Carlin, 3-27. Netherlands Graduate School of Linguistics, Utrecht: LOT.

Crysmann, Berthold and Bonami, Olivier. 2015. Variable morphotactics in Information-based Morphology. Journal of Linguistics, available on CJO2015. doi:10.1017/S0022226715000018.

Okimāsis, Jean. L. 2004. *Cree, Language of the Plains*. Regina: Canadian Plains Research Center.

Conor Snoek, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen, and Trond Trosterud. 2014. Modeling the noun morphology of plains cree. In Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages, 34–42, Baltimore, Maryland, USA, June. Association for Computational Linguistics.

Wolfart H. C. 1973. Plains Cree: A Grammatical Study. Transactions of the American Philosophical Society, New Series, Vol. 63, No. 5: 1-90.

Wolvengrey, Arok. 2011.Semantic and pragmatic functions in Plains Cree syntax. Utrecht Amsterdam: LOT Universiteit van Amsterdam Host.

Wolvengrey, Arok. 2015. Preverb Combinations, Co-occurrences and Sequences: Preliminary Findings from a Preliminary Plains Cree Corpus, 2nd Prairie Workshop on Language and Linguistics (PWoLL), March 14, 2015, Winnipeg.

Enriching the Démonette morpho-semantic network: computational and linguistic issues

Nabil Hathout
Université de Toulouse & CLLE

Fiammetta Namer
Université de Lorraine & ATILF

Démonette is a network designed for the description of French derivational morphology. It features an original architecture that enables its use as a formal framework for the description of morphological analyses, as a repository for existing lexica, as a source of examples for morphological research and of experimental material for psycholinguistics, and as a showcase for the dissemination of morphological research projects. Adding new entries and incorporating new resources generates new information that emerge from the combination of the new data with the descriptions already present in Démonette.

In this paper, we present the integration into Démonette of an existing morphological resource, namely Verbaction. Verbaction is a lexicon of action nouns designed for natural language processing (NLP) and information retrieval (IR). We focus in the paper on the computational and linguistic issues raised by this integration which puts the presented work at the articulation between computational morphology, descriptive and theoretical morphology.

1 Démonette

Démonette (Hathout & Namer, 2014b) is a general resource designed for word formation (WF) description of French. It is eventually intended to partially fill the lack of broad-coverage morphological resources of French as they exist for other languages (e.g. DerivBase for German (Zeller et al., 2013) or CELEX (Baayen et al., 1995) for English, German and Dutch). Démonette has an original structure of directed graph. As illustrated in Table 1 by the relation between *développement* 'development' and *développeur* 'developer', entries are derivational relations between pairs of words (F1). Démonette contains descending relations between derivatives and their bases such as *développement*:*développer* 'develop$_V$', ascending relations between bases and their derivatives such as *développer*:*développement*, and indirect relations between words of the same derivational family such as *développement*:*développeur* (F5), connected in the derivational network by a path that goes by *développer*. By design, Démonette is highly redundant in order to make information handling easier and to facilitate adding new entries and descriptions. It also features a wide range of information about the words and the relations: category (F1), derivational process (F2), derivational stem, etc. An original feature of Démonette is the morpho-semantic typing of the words and the derivational relations (F2). The lexemes that participate in the relations are assigned to morpho-semantic types such as predicate (@), action noun (@ACT), agentive masculine noun (@AGM), agentive feminine noun (@AGF), etc. The semantic description of the relations also include concrete (F3) and abstract definitions (F4) of the meaning of the first word with respect to the meaning of the second. In short, Démonette is by design a system that focuses on the double paradigmatic nature of derivational morphology; it accounts for both form-oriented and meaning-oriented perspectives (for an overview of derivational paradigms, see (Štekauer, 2014)). Démonette is an open resource which can be fed by descriptions from existing derivational bases. However, this information is not "dissolved" in the database, which records the origin of all the information it contains (F5). In the currently distributed version of Démonette (1.1), words and categories originate from TLFnome;[1] in Table 1, all the information except for the morphological relation are calculated within Démonette.

F1	**Word1**, **Cat1**, **Word2**, **Cat2**	développement	Ncms	développeur	Ncms
F2	**Op1**, **Type1**, **Op2**, **Type2**	ment, suf	@ACT	eur, suf	@AGM
F3	**Concrete definition**	action whose (masculin agent OR instrument) is the développeur			
F4	**Abstract definition**	action whose (masculin agent OR instrument) is @AGM			
F5	**Origin**, **Orient**, **Complexity**	morphonette		indirect	simple

Table 1: Main information recorded in a Démonette entry.

[1] TLFnome is a lexicon created from the TLF dictionary word list. See www.cnrtl.fr/lexiques/morphalou

Version 1.1 of the network[2] was built from two existing resources: the analyses of the derived words of TLFnome by the morphological parser DériF (Namer, 2013) and Morphonette (Hathout, 2011), a derivational lexicon automatically built from TLFnome using the Proxinette morphological distance and formal analogy (Gosme & Lepage, 2009). This version contains words formed by 7 suffixations (*-age*, *-ment*, *-ion*, *-eur*, *-euse*, *-rice*, *-if*) and the corresponding verb bases, when available. It gathers 77,323 relations that describe 53,382 word pairs.

2 Verbaction

Verbaction (Hathout & Tanguy, 2002) is one of the first freely distributed derivational lexicons for French. Intended for NLP, this lexicon contains 9,393 Noun-Verb pairs such that (*i*) the Noun is morphologically related to the Verb in synchrony or historically and (*ii*) the Noun can be used to express the action denoted by the Verb. For example, the presence of the pair *fabrication*:*fabriquer* 'manufacturing:manufacture' is motivated by statements like *la fabrication des voitures se fait dans de grandes usines* 'car manufacturing is done in large plants,' and *la société fabrique des voitures* 'the company manufactures cars.' This resource can therefore be used to identify nominal and verbal expression of variants of the same information. Verbaction has been fully manually checked.

Verbaction involves a great variety in derivational processes including a large number of suffixes (*-ade*, *-age*, *-aison*, *-ance*, *-ée*, *-ence*, *-ette*, *-ie*, *-ment*, *-ion*, *-ure*, etc.), various types of conversion, and a great heterogeneity in morphological orientation since it includes pairs where the Noun derives from the Verb (*balisage*:*baliser* 'markup:tag'), others where the Verb is derived from the Noun (*collection*:*collectionner* 'collection:collect') and still others where the relation is indirect (*salarisation*:*salarier* 'salarization:pay a salary', which are connected through *salaire* 'salary'). On the other hand, Verbaction is very consistent on the semantic level since all Nouns can denote action and are related to their corresponding Verbs in the same way.

3 Enhancing Démonette with Verbaction

Incorporating an NLP resource into a network that records linguistic descriptions requires several adaptations because their structures reflect their different purposes and do not match perfectly. The development of a Verbaction-to-Démonette converter has to solve two kinds of problems.

Individual cases. Besides *-age*, *-ment* and *-ion* suffixations, already dealt with in the version 1.1 of Démonette and therefore easily processed, the conversion program has to analyze carefully many particular situations. We first excluded some "wrong" Noun-Verb pairs included in Verbaction such as *pub*:*publier* 'ad:publish' where the Noun does not denote the action of the Verb. Almost all of the Noun-Verb pairs of Verbaction were integrated into Démonette. 30 pairs of the original lexicon (0.31%) were excluded in the conversion, bringing the recall to 99.68%. These include *insurrectionnaire*:*insurrectionner* 'insurrectionist:rise up' (the first is not an action noun but refers to an agent) or *faute*:*falloir* 'fault:have to' (the noun and the verb are not connected in synchrony). Similarly, the accuracy of the converter is greater than 99% thanks to the implemented methodology: the conversion program supports exceptions and distinguishes the Verbaction couples according to their morphological relationship. The program detects unfrequent WF rules such as the *-ing* suffix in *zapping*, or the backformation process (Becker, 1994) used to form the verb *hydroplaner* 'hydroplane' on the base noun *hydroplanage* 'hydroplaning', cf. Namer (2012). It identifies the complexity and orientation (only when relevant; Ex 2, Table 3) of Noun-Verb morphological relations (where the Noun is Word1 and the Verb is Word2). As Table 2 shows, the *complex*/*simple* value of complexity combined with the *ascending*/*descending*/*indirect* value of orientation characterizes the morphological proximity between the two words: the relation is simple in case of a direct parenthood, or when Noun and Verb have an immediate common ascendant; it is complex otherwise.

The variability of the information to encode. Démonette's architecture allows us to leave some of the fields empty depending on the nature of the morphological relation that connects the Noun and the Verb (Table3). Often, a value is provided for all the fields in the target entry (Ex 1). For other pairs, one or

[2] Available at http://redac.univ-tlse2.fr/lexiques/demonette.html

Verb	(action) Noun	Complexity	Orientation
vasectomiser ‘vasectomize’	vasectomie ‘vasectomy’	simple	ascending
syndiquer ‘unionize’	syndicalisation ‘unionization’	complex	indirect
charlatan er ‘charlatan’	charlatanerie ‘charlatanism’	simple	indirect
militer ‘militate’	militantisme ‘militancy’	complex	descending

Table 2: Complexity and orientation combinations in Verbaction

several fields are left blank: (a) the orientation (Ori) is undeterminate in many conversion relations (Ex 2, cf. Tribout (2010)); (b) the Noun definition (Ex 3), when it cannot be spontaneously interpreted with regard to the Verb meaning (cf. Hathout & Namer (2014b)); (c) the complexity (Cx) value is *lexical* when the Noun and the Verb belong to the same morphological family, but lack mutual motivation (Ex 4); (d) the same holds when the Noun is formally and semantically related to the Verb (Ex 5), but the formal relation involves an exceptional variation.

Ex	Noun	Verb	Op_N	Op_V	Cx	Ori	$Type_N$	$Type_V$	Abstract def
1	cathétérisme ‘catheterism’	cathétériser ‘catheterize’	-isme	-iser	simple	ind	@ACT	@	Action of @
	baignade ‘bath’	baigner ‘bath’	-ade	–	simple	desc	@ACT	@	Action of @
	vasectomie ‘vasectomy’	vasectomiser ‘vasectomize’	–	-iser	simple	asc	@ACT	@	Action of @
2	arnaque ‘scam’	arnaquer ‘con’	conv	–	simple	–	@ACT	@	Action of @
3	barbarerie ‘barbarianism’	barbariser ‘barbarianize’	-erie	-iser	simple	ind	@ACT	@	–
4	oblation	offrir ‘offer’	–	–	lexical	–	@ACT	@	–
5	parole ‘speech’	parler ‘speak’	–	–	lexical	–	@ACT	@	–

Table 3: Noun-Verb entries in Démonette (excerpt)

4 Linguistic consequences

Besides these conversion issues, the incorporation of Verbaction into Démonette raises various linguistic questions. They mainly concern the definition of the Noun with respect to the Verb (cf. last column of Table 3). Hathout & Namer (2014b) suggested that a Démonette definition is meaningful only when speakers are able to naturally interpret the corresponding relation. In this case, the relation is semantically motivated. Verbaction data allow us to sketch some of the conditions which favour or prevent the assignment of definitions to simple and complex indirect relations.

- The presence of a synonymy relation in the morphological subfamily containing both Noun and Verb is a favourable condition. For instance, the members of the *sautillonner*:*sautillage* ‘hop:hopping’ pair have *sautiller* ‘hop_V’ as common ascendant and can be cross-defined easily: the synonymy *sautiller*/*sautillonner* entails a semantic equivalence of the relations *sautillonner*:*sautillage* and *sautiller*:*sautillage*. The Verb is defined as ‘Perform *sautillage*’ and the Noun as ‘Action of *sautillonner*/*sautiller*.’ Note that *sautillonner* and *sautiller* are synonymous because the former is marked twice by diminutive suffixes and the latter once (cf. Hathout & Namer (2014a) on over-marking in WF).

- On the other hand, factors that make the definition assignment undesirable involve frequency, polysemy and paradigmatic organization (comparable to Bochner’s (1993) cumulative patterns, see also Strnadová (2015)) criteria. For instance, with the *hanter*:*hantise* pair, the difficulty that occurs when one tries to define *hanter* ‘haunt’ with respect to *hantise* ‘dread’ (and vice-versa) arises from the fact that the stative interpretation of $hantise_N$ is far more frequent than the action interpretation; as a con-

sequence, the semantic type of *hanter* (action) diverges from that of *hantise* (state), preventing the mutual definition of Verb and Noun.

5 Conclusion

In this paper, we show that Démonette can host existing resources, and thus increase their visibility and profitability. One of Démonette's goals is to provide scholars with a framework for a sustainable access to lexica and datasets developed by the morphologists as part of their research. The linguistic issues raised by the integration of Verbaction lead to other questions: which information has to be supplied to Démonette, in order to detect polysemous words and synonymy relations? Should the definition of denominal verbs formed by highly polysemous WF processes reflect the ambiguity of their constructed meaning (cf. Aronoff (1980)), or should we avoid under-specification (cf. Rainer (2014))? These are some of the questions related to the interface between computation and linguistic description addressed in the paper.

References

Aronoff, Mark. 1980. Contextuals. *Language* 56(4). 744–758.

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. The CELEX lexical database (release 2). CD-ROM. Linguistic Data Consortium, Philadelphia, PA.

Becker, Thomas. 1994. Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1993*, 1–25. Dordrecht: Kluwer Academic Publishers.

Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin & New-York: Mouton de Gruyter.

Gosme, Julien & Yves Lepage. 2009. A first study of the complete enumeration of all analogies contained in a text. In *4th Language & Technology Conference (LTC 2009)*, 401–405. Poznań, Poland.

Hathout, Nabil. 2011. Morphonette: a paradigm-based morphological network. *Lingue e linguaggio* 2011(2). 243–262.

Hathout, Nabil & Fiammetta Namer. 2014a. Discrepancy between form and meaning in Word Formation: the case of over- and under-marking in French. In Franz Rainer, Wolfgang U. Dressler, Francesco Gardani & Hans Christian Luschützky (eds.), *Morphology and meaning*, 177–190. Amsterdam: John Benjamins.

Hathout, Nabil & Fiammetta Namer. 2014b. Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology* 11(5). 125–168.

Hathout, Nabil & Ludovic Tanguy. 2002. Webaffix : Finding and validating morphological links on the WWW. In *Proceedings of the 3rd international Conference on Language Resources and Evaluation*, 1799–1804. Las Palmas de Gran Canaria: ELRA.

Namer, Fiammetta. 2012. Nominalisation et composition en français: d'où viennent les verbes composés ? *Lexique* 20. 173–205.

Namer, Fiammetta. 2013. A rule-based morphosemantic analyzer for French for a fine-grained semantic annotation of texts. In Cerstin Mahlow & Michael Piotrowski (eds.), *SFCM 2013* CCIS 380, 93–115. Heidelberg: Springer.

Rainer, Franz. 2014. Polysemy in derivation. In Rochelle Lieber & Pavel Štekauer (eds.), *The Oxford hanbook of derivational morphology*, 338–353. Oxford: Oxford University Press.

Strnadová, Jana. 2015. Multiple derivation in French denominal adjectives. In *Proceedings of the 8th Décembrettes*, 327–346. Bordeaux. Carnets de grammaire, 22.

Tribout, Delphine. 2010. *Les conversions de nom à verbe et de verbe à nom en français*: Université Paris 7. Phd thesis.

Zeller, Britta D, Jan Snajder & Sebastian Padó. 2013. DErivBase: Inducing and evaluating a derivational morphology resource for german. In *Proceedings of the 51th annual meeting of the Association for Computational Linguistics (ACL)*, 1201–1211. Sofia, Bulgaria.

Štekauer, Pavel. 2014. Derivational paradigms. In Rochelle Lieber & Pavel Štekauer (eds.), *The Oxford hanbook of derivational morphology*, 354–369. Oxford: Oxford University Press.

Modeling the Grammaticalization Path of Functional Morphemes: A Game-Theoretic Analysis

Roland Mühlenbernd[1], Dankmar Enke[2], Igor Yanovich[1,3]

[1]*Department of Linguistics, University of Tübingen, Germany*
[2]*Institute of German Philology, University of Munich, Germany*
[3]*Department of Philosophy, Carnegie Mellon University, PA, United States*

1 Introduction

It is a well-known typological observation that languages without a distinct progressive (PROG) morphology realize the communicative function of the PROG through the imperfective (IMP) aspect (if morphologically instantiated). This primarily motivates to treat the PROG as a subdomain of the IMP (cf. Comrie 1976). Consider the following example from Russian:

(1) a. Olga **pisa-la** pis'ma kogda pojavilsja Vadim
Olga.NOM write-IMPF.PST.F letter.ACC.PL when appear.PERF.PST.M Vadim.NOM
'Olga *was writing* letters when Vadim appeared.'

b. Olga **pisa-la** pis'mo materi po voskresenjam
Olga.NOM write-IMPF.PST.F letter.ACC.SG mother.DAT on Saturday.DAT.PL
'Olga *used to write* a letter to her mother on Sundays.'

c. Olga **zhi-la** v Moskv-e
Olga.NOM live-IMPF.PST.F in V-LOC
'Olga *lived* in Moskow.'

The imperfective form *pisala* 'wrote' (1a) licences a PROG interpretation, while the same form in (1b) refers to a habitual/generic (HAB/GEN) situation. The lexical stative verb *zhila* 'live' in (1c) occurs in the IMP form. In contrast to languages such as Russian, there are languages which have both the PROG and the IMP aspects. Here, the IMP often does not licence a PROG reading (Isačenko 1962), such as in English.

In the grammaticalization literature there is a crosslinguistically robust generalization in the diachrony of such markers: functional elements restricted to PROG reading semantically generalize to licence IMP readings such as the HAB/GEN or the stative. This generalization has been attested according to data from, e.g., Turkish (Göksel & Kerslake 2005: 331):

(2) a. saat ikide çaliş-**iyor-du**-m
At two o'clock work-PROG-PST.COP-1SG
'A two o'clock I *was working*.'

b. genekkikle iki saat çaliş-**ir-di**-m
Usually for two hours work-IMPF-PST.COP-1SG
'I *would* usually *work* for two hours.'

The verb form with *-(I)yor* in (2a) refers to an ongoing eventuality, while the inflected verb with *-ir* in (2b) refers to a HAB reading. Recently, the PROG *-(I)yor* has begun to licence a wider range of readings, notably in everyday language. (3a) shows that *-(I)yor* occurs with a

(3) a. sen Ömer'i benden daha iyi tan-**iyor-du**-n
you Omer me better than know-PROG-PST.COP.2.SG
'You *knew (were knowing)* Ömer better than me.'

b. O zamanlarda mehmet çok sigara iç-**iyor-du**
At the time Mehmet.NOM lot cigarette smoke-IMPF-PST.COP.3.SG
'At that time, Mehmet *used to smoke (was smoking)* a lot.'

stative verb 'know' and is also interchangeably used with the Aorist form (IMP aspect) with a HAB/generic reading (3b) (Göksel & Kerslake 2005: 331).

These data indicate that the Turkish progressive is expanding to semantically overlap with the domain of the IMP Aorist morphology, thus instantiating the PROG-to-IMP shift.
The PROG-to-IMP historical cycle starts with the language having only one broad IMP form covering all imperfective meanings, cf. (a) below. Then an optional PROG form is innovated, (b); it becomes obligatory for PROG meanings, (c); and at the last stage (d), it generalizes and takes the semantic place of the old broad-IMP form. Note that (a) and (d) are identical except for their formal exponents: the cycle ends where it started (Table 1).

The four states (a-d) can be intuitively regarded as distinct strategies for communicating *phenomenal* and *structural* sub-meanings (Goldsmith & Woisetschlaeger 1982) within the IMP domain. In systems with two forms, namely emergent-PROG and categorical-PROG, the choice of form helps the hearer to correctly identify the speakers intended sub-meaning. The zero-PROG and generalized-PROG strategies use a single form while relying on the hearers understanding of contextual cues for successful communication. Importantly, PROG induces a cycle through (a-d), but habitual HAB, though also being more specific than the broad imperfective IMP, does not eventually generalize to IMP (Deo 2015). In other words, there is no (d)-type stage for HAB, and therefore no HAB-to-IMP cycle.

Deo (2015) sets up an evolutionary game-theoretic model for studying the dynamics of these systems. She defines a simple game representing individual interactions of agents with different grammars, in the tradition of game-theoretic models of communication (cf. Benz, Jäger, and Rooij (2005) for an overview.) Deo then lifts this game to an evolutionary game that represents the dynamics of grammar retention and change over long temporal horizons. Both the basic and the evolutionary games are defined by Deo on the space of four strategies corresponding to (a-d) above. In this talk, we investigate systems with PROG, IMP and HAB at a more granular level. We consider populations of agents that build their own grammar through reinforcement learning on the basis of what they hear other agents say. In particular, we investigate (i) which strategies actually arise, and what assumptions would rule out the ones that are not empirically observed in human languages, as well as (ii) how to induce the PROG-to-IMP, but crucially not the HAB-to-IMP cycle.

We show that under simple assumptions, exactly the right grammars and the correct diachronic sequences emerge. In particular, regarding (ii) we show that Deo was correct in conjecturing that being more exposed to PROG-type meanings in childhood induces the PROG-to-IMP, and rules out the HAB-to-IMP, development (Deo 2015: 22). We thus effectively provide microfoundations for Deos macro model of the PROG-to-IMP cycle.

	label	type	sample languages
(a)	X_{imp}	zero-PROG	Arabic, Russian
(b)	$(Y_{prog})X_{imp}$	emergent-PROG	Dutch, Italian
(c)	Y_{prog}, X_{imp}	categorical-PROG	English, Swahili
(d)	Y_{imp}	generalized-PROG	Turkish, Tigre

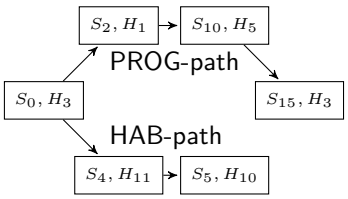
Table 1: The PROG-to-IMP historical cycle with representative languages.

	C_p t_p	C_p t_s	C_s t_p	C_s t_s
S_0	f_{im}	f_{im}	f_{im}	f_{im}
S_1	f_{im}	f_{im}	f_{im}	f_{pr}
S_2	f_{im}	f_{im}	f_{pr}	f_{im}
S_3	f_{im}	f_{im}	f_{pr}	f_{pr}
S_4	f_{im}	f_{pr}	f_{im}	f_{im}
S_5	f_{im}	f_{pr}	f_{im}	f_{pr}
S_6	f_{im}	f_{pr}	f_{pr}	f_{im}
S_7	f_{im}	f_{pr}	f_{pr}	f_{pr}
S_8	f_{pr}	f_{im}	f_{im}	f_{im}
S_9	f_{pr}	f_{im}	f_{im}	f_{pr}
S_{10}	f_{pr}	f_{im}	f_{pr}	f_{im}
S_{11}	f_{pr}	f_{im}	f_{pr}	f_{pr}
S_{12}	f_{pr}	f_{pr}	f_{im}	f_{im}
S_{13}	f_{pr}	f_{pr}	f_{im}	f_{pr}
S_{14}	f_{pr}	f_{pr}	f_{pr}	f_{im}
S_{15}	f_{pr}	f_{pr}	f_{pr}	f_{pr}

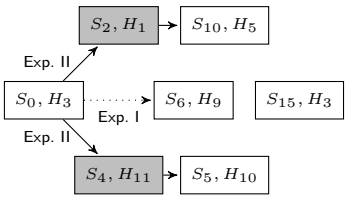
(a) Speaker strategies

	C_p f_{pr}	C_p f_{im}	C_s f_{pr}	C_s f_{im}
H_0	t_p	t_p	t_p	t_p
H_1	t_p	t_p	t_p	t_s
H_2	t_p	t_p	t_s	t_p
H_3	t_p	t_p	t_s	t_s
H_4	t_p	t_s	t_p	t_p
H_5	t_p	t_s	t_p	t_s
H_6	t_p	t_s	t_s	t_p
H_7	t_p	t_s	t_s	t_s
H_8	t_s	t_p	t_p	t_p
H_9	t_s	t_p	t_p	t_s
H_{10}	t_s	t_p	t_s	t_p
H_{11}	t_s	t_p	t_s	t_s
H_{12}	t_s	t_s	t_p	t_p
H_{13}	t_s	t_s	t_p	t_s
H_{14}	t_s	t_s	t_s	t_p
H_{15}	t_s	t_s	t_s	t_s

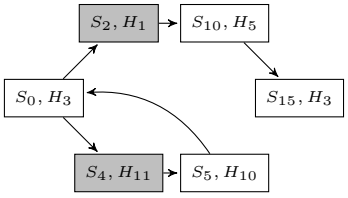
(b) Hearer strategies



(c) Expected paths



(d) Experiment I and II



(e) Experiment III

Figure 1: All possible speaker and hearer strategies (Fig. 1a,b) that are theoretically possible for the Basic Imperfective Game. Of those, Deo (2015) considered only 4 speaker and 3 hearer strategies (highlighted in gray in the tables), which constitute the 'progressive ≫ imperfective' cycling path (*PROG-path*). The expected diachronic development includes next to the *PROG-path* its alternative deadlock *HAB-path* (Fig. 1c). Our studies reveal: in unconditioned Experiment I agents switch to an unexpected strategy pair $\langle S_6, H_9 \rangle$, whereas in Experiment II with reduced context cues by 20% agents stabilize on strategy pair $\langle S_{10}, H_5 \rangle$ (categorical-PROG) or strategy pair $\langle S_5, H_{10} \rangle$ (categorical-GEN) (Fig. 1d). Experiment III – with two additional assumptions of i) increasing costs for a two-form system, and ii) childhood input asymmetry – produces the expected paths (Fig. 1e). Note: unstable strategy pairs are displayed in gray.

2 The Imperfective Game

The Imperfective Game is modeled as a signaling game (Lewis 1969), a game-theoretic model that depicts the communication situation between a speaker and a hearer. The communicative act is modeled as a decoding/encoding process: with the goal to communicate a state t, a sender sends a form f, and the hearer construes f by choosing an interpretation state t'. Communication is valued as successful, iff $t = t'$. The model of the Imperfective Game abstracts i) from the concrete situation, but only considers the types of state t_s (structural) and t_p (phenomenal), and ii) from the concrete form, but only considers the types of form f_{pr} (PROG marking) and f_{im} (IMP marking). Furthermore, since contextual cues play an important role, the model includes two contexts C_s (rather licensing the structural state) and C_p (rather licensing the phenomenal state).

The communicative behavior of players of the game can be modeled as a *strategy*. A speaker strategy defines which form to choose for each combination of context type and state type. Appropriately, a hearer strategy defines which state to construe for each combination of context type and form type. For the IMP Game, there are 16 logically possible speaker and hearer strategies, as depicted in Figure 1(a) and 1(b). The expected steps of PROG-to-IMP cycling path and the HAB-to-IMP dead end path are depicted in Figure 1(c) by referencing to the appropriate pair of speaker and hearer strategy.

To test hypotheses that are possible candidates for explaining the particular sequence of the PROG-to-IMP cycling path, we embedded the Imperfective Game in a computational model for simulation experiments. Our model is implemented as a population of agents that interact repeatedly and pairwisely by playing the Imperfective Game. Furthermore, agents i) update

their communicative behavior by an update rule called Roth-Erev reinforcement learning (Roth & Erev 1995), and ii) are alive for a particular number of simulation steps: 'old' agents are posthumously replaced by 'young' agents.

With our first line of simulation runs (Experiment I) we tested the basic model. It turned out that the population always ended up in a linguistically odd 'surprisal' strategy $\langle S_6, H_9 \rangle$. This course resulted from the rather unrealistic property that agents have always access to the contextual cues C_p or C_s. By reducing the frequency of cues to 80% (Experiment II), the population equiprobably ends up in either the categorical-PROG system $\langle S_{10}, H_5 \rangle$ or the categorical-HAB system $\langle S_5, H_{10} \rangle$ (see Figure 1(d)).

Since both categorical systems are optimal in a game-theoretic sense, agents will never leave this state in the current model. To incentive agents to shift to a one-form system, we added an increasing cost value for two-form systems. Furthermore, we added the assumption that 'young' agents are confronted with phenomenal states more often than older agents. The results of the simulation runs (Experiment III) are depicted in Figure 1(e): the PROG-to-IMP cycling path was perfectly reconstructed and the HAB-to-IMP deadlocking path eventually looses the new form f_{pr} and shifts back to the initial state.

3 Conclusion

We used experiments with reinforcement learning agents playing the Imperfective Game to investigate whether the empirically observed grammar changes involving the imperfective IMP, progressive PROG and habitual HAB would emerge in this setting. With relatively simple assumptions, we achieved the emergence of both the PROG-cycle and the deadlocking HAB-trajectory. The assumptions that we sequentially added to the basic model and their consequences were: (1) the emergence of linguistically odd 'surprisal' strategy $\langle S6, H9 \rangle$ was prevented by sometimes withdrawing the context cue from the hearer; (2) the switch from a perfectly communicatively efficient two-form grammar calls for a cost function; and (3) when agents were mostly presented with phenomenal statements in the childhood, the empirically observed PROG-to-IMP cycle emerged, while the unattested hypothetical HAB-to-IMP cycle was ruled out vindicating a conjecture by Deo (2015).

4 References

Benz, A., G. Jäger & R. van Rooij (2005). *Game Theory and Pragmatics*. New York: Palgrave Macmillan.

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.

Deo, A. (2007). *Tense and Aspect in Indo-Aryan Languages: Variation and Diachrony*. PhD dissertation. Stanford University.

Deo, A. (2015). The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics and Pragmatics* 8, Article 14.

Göksel, A. & C. Kerslake (2005). *Turkish: A Comprehensive Grammar*. London: Routledge.

Goldsmith, J. & E. Woisetschlaeger (1982). The logic of the English progressive. *Linguistic Inquiry* 13(1), 79-89.

Isačenko, A. (1962). *Die russische Sprache der Gegenwart. Teil I: Formenlehre*. Halle: Niemeyer.

Lewis, D. (1969). *Convention*. Cambridge, MA: Harvard University Press.

Lindfors, A.-L. (2003). *Tense and Aspect in Swahili*. Uppsala University, unpublished ms.

Roth, A. & I. Erev (1995). Learning in extensive-form games: Exerimental data and simple dynamic models in the intermediate term. *Games and Economic Behaviour* 8, 164-212.

Measuring the acquisition of morphology in big longitudinal corpora

Sabine Stoll, Jekaterina Mažara, and Balthasar Bickel

University of Zurich

Introduction Morphology is one of the biggest challenges for the language learning child since morphology includes substantial parts of grammatical information and at the same time can occlude the identity of lexical roots. Since language learning critically depends on the successful and rapid identification and memorization of constant and re-usable units, any variation in the form of these units makes the acquisition process harder. As a result, the more complex a morphology is, the harder it renders the extraction of re-usable units. Two dimensions of morphological complexity are particularly important for this: the number of elements in a form (its degree of synthesis) and the variability in the expression of these elements (complex exponence patterns, allomorphy, etc.). In order to understand the different challenges that languages pose for the language learning child and how children cope with them, we need to be able to quantitatively measure the cross-linguistic differences in morphological complexity, the distribution of complexities that a child encounters in her linguistic environment, and the development of morphological complexity during acquisition.

There is a third dimension of morphological complexity relevant for acquisition and this is the predictability of a specific form based on another form in the paradigm (Ackerman & Malouf 2013). This type of complexity captures the intuition that speakers must be able to generalize beyond direct experience with individual forms because it is unlikely that they will encounter all possible forms of all verbal stems, especially in polysynthetic languages with thousands of verb forms. It has been suggested that individual forms in a paradigm allow the speaker to make predictions about new forms (Finkel & Stump 2007, Stump & Finkel 2013, Ackerman & Malouf 2013). However, to make this measure applicable for cross-linguistic language acquisition research we need to extract potential predictability from corpus information, ideally both for languages with inflectional classes and for languages without. Thus we need to develop methods that allow us to estimate the forms a child has previously encountered and learned to predict her production of new forms based on frequency measures.

Here we discuss some possible ways to resolve this, but our main concern here is the quantitative exploration of the first two dimensions of complexity, corresponding to what Ackerman & Malouf (2013) call 'enumerative complexity'. We first assess quantitatively the structure of the input of children learning languages with radically different morphologies: English, an analytic language, and Chintang, a polysynthetic language. We then explore the verbal development of children learning Chintang.

Data Chintang is an endangered Sino-Tibetan language spoken in a small village in Eastern Nepal. The morphology is polysynthetic but has no inflecional classes. A verb stem combines with a large number of other morphemes that are subject to an adjacency (but not an ordering) requirement and range from grammatical affixes to other lexical verb stems. The expression of these combinations shows considerable variation in exponence patterns (simple, cumulative, and distributed exponence) and co-dependence of exponents (e.g. the form of third person singular

agent agreement marking depends on the nature of the patient co-coded in the same verb form). Categories expressed include various notions of person, number, tense, aspect, polarity and many derivational and lexical meanings (Bickel et al. 2007, Stoll et al. 2012, Stoll & Bickel 2013, Schikowski 2013, Stoll et al. 2015, Schikowski et al. 2015).

Our analysis of Chintang is based on a longitudinal corpus of 6 children recorded monthly for four hours over a period of 18 months. For present purposes, we use the glossed part of the corpus, focusing on pure Chintang words. This includes a total of 315,816 orthographic words. For the analysis of child-surrounding speech (input), we focused on speakers older than 12 years, which amounts to 148,935 words, spoken by 119 different speakers.

For English, we use the Manchester acquisition corpus available on CHILDES (Theakston et al. 2001). In order to match the corpora in length, we cut down the English corpus to the same number of verbs as encountered in the Chintang corpus (by taking a random subsample). This leaves us with 96,279 verb form tokens per corpus.

Differences in the input In the Chintang corpus, we have recorded 147 verb morphemes (grammatical affixes and verb stems that select for another verb stem) and these form 4,745 unique combinations among themselves. In terms of verb forms, this amounts to 1,840 different forms that a single stem can occur in. By contrast, the English verbal paradigm consists of a handful of affixes, which cannot be combined among themselves, severely limiting the range of forms.

The differences in the distributions of the verb forms are extreme in the two languages (Figure 1). Chintang children hear 16.6 times more verb forms than English children (Chintang children hear 23,888 different types, whereas English children hear merely 1,449 different forms). Frequency distributions in the two languages also differ strongly. In English, only 37% of the forms are hapax legomena, whereas in Chintang, 66% of the forms occur only once.

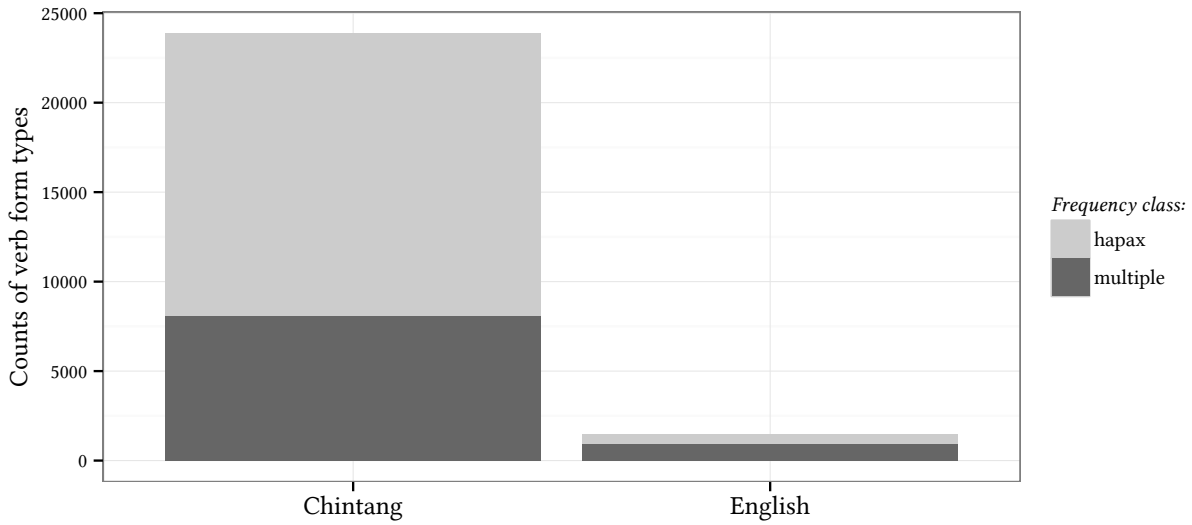


Figure 1: Counts of verb form types in normalized Chintang and English input corpora (N = 96,279 verb forms). A verb form type is the unique combination of a specific lexical verb stem and any morphemes that requires this stem.

As a result, the input that Chintang children receive strongly differs from the input of English children. We show this by analyzing the deviation of these two distributions from that of an ideal Zipfian curve (Zipf 1932, 1935). As shown in Figure 2, English has a steeper-than-predicted slope,

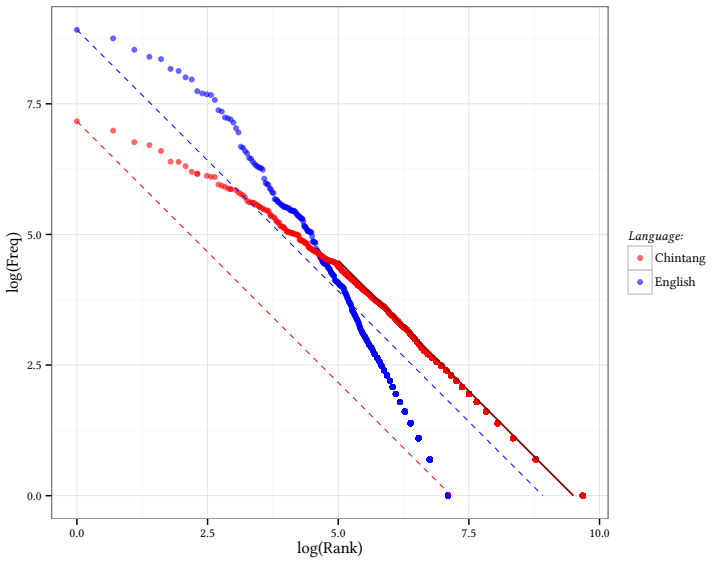


Figure 2: Frequency of verb form types ordered by frequency ranks on a logarithmic scale. The dashed lines represent the expected distribution under Zipf's Law with a slope $s = -1$. The solid red line shows the expected distribution under Zipf's Law for Chintang word forms starting with log(Rank)=5.

which implies a smaller inventory of types with higher rates of repetition and facilitated learnability. Chintang has a flatter-than-predicted slope, which implies a more diverse inventory of types and lower rates of repetition.

The effect of these differences in frequency distribution for learning, however, might not be as big as it seems. One of the key task for children is to segment stems. Since there are no inflectional classes, the child could in principle generalize the individual verb forms to other stems as soon as she has encountered a single full form and is able to segment stems from grammatical markers. This would suggest that even though the enumerative complexity of the Chintang verb paradigm is very big, the challenge for the learner reduces mostly to learning the different verb forms and their grammatical meanings.

Morphological development To measure morphological learning in Chintang children, we used two different methods. First, we used growth rates to measure the actual number of new forms encountered over time. Chintang children add several hundred new forms to their verb inventory every month. A growth rate analysis suggests that children show a strong increase in their inventory of verb form types between the ages of 2 and 4. The peak of increase is located between the age of 3 and 3;6. The growth rate of verb forms is due to a joint effort of lexicon and grammar learning, with grammar taking the lead from age 2;6 onward (see Figure 3).

To explore this difference further, we also measured children's command of the two domains – lexical stems and affix combinations – with Shannon entropy (Moscoso del Prado Martín et al. 2004) and compared these measurements to those in the input that children receive (following Stoll et al. 2012). Results confirm the growth analysis and show that, also with regard to proficiency and productivity, children's development is initially mainly driven by lexical learning and later by grammatical learning (affix/V2combinations), before both are integrated to reach adult performance levels.

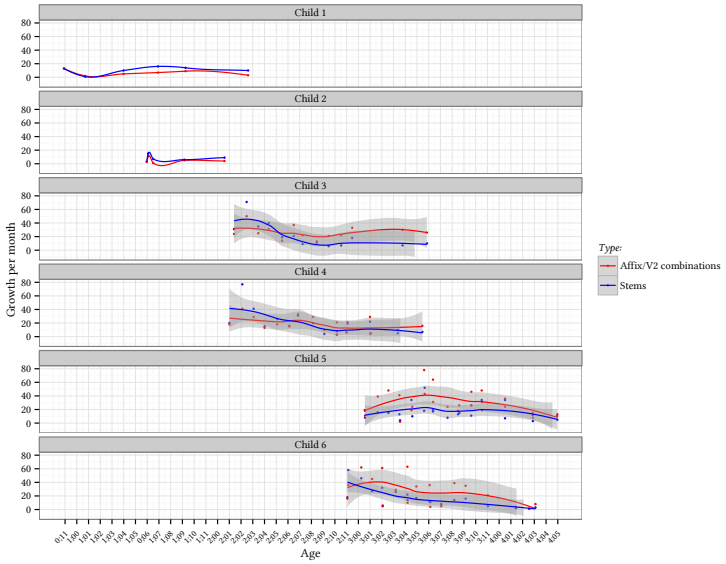


Figure 3: Counts of new regular verb stems vs. new unique affix/V2 combinations in the recorded productions per month, with LOESS regression lines estimating the overall trends and 95% confidence intervals (where appropriate).

References

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

Bickel, B., G. Banjade, M. Gaenszle, E. Lieven, N. Paudyal, I. P. Rai, M. Rai, N. K. Rai & S. Stoll. 2007. Free prefix ordering in Chintang. *Language* 83. 44–73.

Finkel, R. & G. Stump. 2007. Principal parts and morphological typology. *Morphology* 17(1). 39–75.

Moscoso del Prado Martín, F., A. Kostic & R. H. Baayen. 2004. Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94. 1–18.

Schikowski, R. 2013. Object-conditioned differential marking in Chintang and Nepali. PhD Dissertation, University of Zürich, [http://opac.nebis.ch/ediss/20131807.pdf].

Schikowski, R., N. P. Paudyal & B. Bickel. 2015. Flexible valency in Chintang. In B. Comrie & A. Malchukov (eds.), *Valency classes in the world's languages: a comparative handbook*, 669 – 707. Berlin: De Gruyter Mouton.

Stoll, S. & B. Bickel. 2013. The acquisition of ergative case in Chintang. In E. L. Bavin & S. Stoll (eds.), *The acquisition of ergativity*, 183–208. Amsterdam: Benjamins.

Stoll, S., B. Bickel, E. Lieven, G. Banjade, T. N. Bhatta, M. Gaenszle, N. P. Paudyal, J. Pettigrew, I. P. Rai, M. Rai & N. K. Rai. 2012. Nouns and verbs in Chintang: children's usage and surrounding adult speech. *Journal of Child Language* 39. 284–321.

Stoll, S., T. Zakharko, S. Moran, R. Schikowski & B. Bickel. 2015. Syntactic mixing across generations in an environment of community-wide bilingualism. *Frontiers in Psychology* 6. 82.

Stump, G. & R. A. Finkel. 2013. *Morphological typology: From word to paradigm.* Cambridge: Cambridge University Press.

Theakston, A. L., E. V. M. Lieven, J. M. Pine & C. F. Rowland. 2001. The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language* 28. 127–152.

Zipf, G. K. 1932. *Selected studies of the principle of relative frequency in language.* Harvard University Press.

Zipf, G. K. 1935. *The psycho-biology of language. An introduction to dynamic philology.* M.I.T. Press.

Ida Raffaelli

Lexicalization patterns in colour naming: a cross-linguistic perspective

The workshop aims to provide a forum about "lexicalization patterns" in colour naming, mostly in Indo-European (IE) languages but also in non-IE languages. The term *lexicalization pattern* comprises word-formation patterns as well as other grammatical (e.g. syntactic) patterns used in colour naming. The workshop will provide new evidence regarding lexicalization patterns of colour terms from a cross-linguistic perspective. It will also point to some regular lexicalization patterns shared by different languages within the IE language family.

The workshop is based on results of the EOSS project (Evolution of Semantic Systems, 2011-2014, Max Planck Institute for Psycholinguistics, Nijmegen) on how meanings vary over space and change over time. Data from 50 IE languages were collected and some preliminary results were provided in Majid, Jordan & Dunn (2015). One of the project tasks was a colour elicitation task that consisted of 84 Munsell colour chips arranged in a single fixed random order (materials developed by Majid & Levinson 2007). Colour chips differed with respect to brightness, hue and saturation. Twenty participants were asked to name the chips that were presented to them individually and randomly. Approximately 1,680 full responses were collected in total for each IE language. These results point to some regularities and specificities of lexicalization patterns in colour naming in IE languages. For example, data from some Slavic languages (such as Croatian, Czech and Polish) show that compounding is more pervasive than derivation. Compounding allows for a more fine-grained naming of the colour spectrum and allows for greater creativity in colour naming than derivation does.

Preliminary research results led us to formulate the following topics and questions.

1) Colours have been investigated from many perspectives, and there is a constant debate between proponents of a relativist view and those of a universalist view. The relativist view considers languages as influential in the partition of the colour space, while the universalist view regards colour space as organized around universal focal colours (Rieger, Kay & Ketharpal 2007). The approach of the workshop is a relativist one, since its main goal is to investigate how language structures can provide answers regarding the differentiation and partition of the colour space. The focus is on investigating productive word-formation patterns in colour naming. According to our preliminary research, derivation and compounding are not equally productive in colour naming. Therefore, one of the important questions that will have to be answered is what types of meanings are conveyed via derivation in different languages, and what types of meanings are conveyed via compounding (some examples are given below). If compounding is more pervasive in colour naming, what patterns can be identified in different languages?

2) Different languages use different affixes to convey different types of meanings related to colours: these are mostly suffixes but occasionally also prefixes, such as Latin *sub-* in *subalbus* 'whitish'. As the data show, suffixes convey two main types of meanings in colour naming. In Croatian, for example, the suffix *-ast* means 'N-like', such as in *naranč-ast* 'orange-like', whereas the suffix *-kast* has an approximative meaning, such as in *zelen-kast* 'greenish'. In Croatian, the suffix *-ast* regularly realizes the metonymic shift OBJECT COLOUR FOR COLOUR. The same pattern of semantic differentiation between two types of suffixes can be identified in Polish as well. Polish has the suffix *-owa*/*-owy* to realize the metonymic shift OBJECT COLOUR FOR COLOUR. The Polish suffix is more productive than the Croatian one. It can be added to nouns referring to more types of entities (*łososiowy* 'salmon-like') than in Croatian where it is restricted to fruits or vegetables. In Polish, the suffix *-awa*/*-awy*, as in *zielonkawy* 'greenish', has an approximative meaning. Czech has a suffix *-ov* that corresponds to the Croatian *-ast*, realizing the aforementioned metonymic shift. However, Czech has no corresponding suffix for the approximative meaning and therefore has to resort to other grammatical devices like prepositional phrases (e.g. *do modra* 'bluish'). We aim to provide insights into what types of suffixes are used and what meanings they convey in colour naming. Are these suffixes polysemous? Do languages

have a tendency to be categorized as more of an "approximation"-type or more of a "metonymy"-type?

3) Results of previous research (Conklin 1973, Wierzbicka 2005, Malt & Majid, 2013) show that languages regularly and frequently use terms for familiar and well-known entities to name colours. The third important topic of the workshop therefore concerns the patterns used to lexically realize the metonymic shift OBJECT COLOUR FOR COLOUR in IE and non-IE languages. Preliminary research carried out within the EOSS project points to several lexicalization patterns. The first pattern is the [Nentity + suffix] pattern, as in the above-mentioned example of Croatian *naranč-ast* 'orange-like'. Nouns for familiar entities (very often fruits and vegetables) are frequently used as stems in the formation of colour terms. The second pattern is the [N +Ngen] pattern. For example, in Croatian the noun *boja* 'colour' is often the head of a phrase, such as in *boja breskve* 'peach colour' or *boja višnje* 'cherry colour'. Thus the phrase *suknja boje breskve* (lit. 'skirt colour cherry') would be used to name a colour of a skirt. In such a construction, the dependent element is a noun in the genitive case, usually referring to a fruit or a vegetable. In French, metonymy is realized by a single word, as in (*jupe*) *saumon* 'salmon (skirt)', by the [N+N] construction as in (*jupe*) *couleur saumon* 'lit. skirt colour salmon', or by the [Adj+N] construction, e.g. (*jupe*) *bleu ciel* 'sky-blue (skirt)', in which the adjective is a colour term and is the head of the construction. Croatian and French data show that different languages have different ways of syntactically realizing the metonymic shift of OBJECT COLOUR FOR COLOUR. What are the differences in lexicalization patterns that languages use to express this metonymy? Do cultural differences play a role? What other types of constructions can be identified? Can regular lexicalization patterns be defined that realize the OBJECT COLOUR FOR COLOUR metonymy in IE languages?

4) The final topic considers types of modifiers that are used in compounding. Modifiers differ with respect to part-of-speech and semantic content. They are mostly adverbs, such as Croatian *tamno* 'dark' and *svijetlo* 'light', or can be adjectives that form compounds, such as Croatian *zeleno-plava* and French *vert bleu*, both meaning 'green-blue'. Adverbs that are used as modifiers mostly refer to hue, and sometimes to brightness as well, e.g. the Croatian adverbs *zagasito* 'dull' and *jarko, žarko* 'bright'. Adjectives used as modifiers form compound colour terms that could be characterized as headless compounds, e.g. *zeleno-plava* 'blue-green'. The components of this type of compound may switch position. According to the Croatian EOSS data, *zeleno-plava* and *plavo-zelena* are used to name the same colour shade. In this context, it could be interesting to investigate whether headless compounds in colour naming can be identified in other languages as well.

We invite the submission of papers dealing with the proposed topics. Papers may have both synchronic and diachronic orientation. We strongly encourage corpus-based research and linguistic analysis stemming from the usage-based approach.

References

Conklin, Harold C. 1973. Color categorization. *American Anthropologist* 75. 931–942.

Majid, Asifa, Fiona Jordan & Michael Dunn. 2015. Semantic systems in closely related languages. *Language Sciences* 49(1). 1–18.

Majid, Asifa & Stephen C. Levinson. 2007. The language of vision I: Colour. In Asifa Majid (ed.), *Field manual*, vol. 10. Max Planck Institute for Psycholinguistics, Nijmegen, 22–25. http://fieldmanuals.mpi.nl/volumes/2007/language-of-vision-colour/.

Malt, Barbara C. & Asifa Majid. 2013. How thoughts is mapped into words. *WIREs Cognitive Science* 4(6). 583–597.

Rieger, Terry, Paul Kay, Naveen Ketharpal. 2007. Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America* 104(4). 1436–1441.

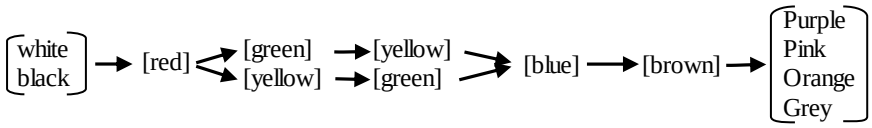
Wierzbicka, Anna. 2005. There are no ''color universals'' but there are universals of visual semantics. *Anthropological Linguistics* 47(2). 217–244.

The lexicalization of colour in verbal derivation: The case of Modern Greek and English

Chariton Charitonidis
University of Cologne
charitonidis@uni-koeln.de

This lecture deals with the fundamental role of semantic constraints in word formation. The object of investigation are the colour verbs in Modern Greek (MG) and English (EN). The analysis is based on the studies by Berlin & Kay (1969), Kay & McDaniel (1978), and Kay & Maffi (1999) on Basic Colour Terms (BCTS).
In (1) the main findings of Berlin & Kay (1969) are given. As can be seen, seven stages of BCT systems are defined, depending on the number of BCTs that a particular language has. Languages with eight to eleven BCTs are assigned to stage VII.

(1)



[Berlin & Kay 1969: 4]

Possible/established verbs in MG and EN address a primary categorial core (Stages I-V), cf. MG *kokinízo* 'make/become red' (*kókino* 'red') and EN *to redden* 'make/become red'. Possible verbs/neologisms are adjacent to this primary core, cf. MG *movízo* 'make/become purple' (*mov* 'purple') and EN *to pinken* 'make/become pink' (Stages VI-VII). In the case of MG, very fresh neologisms are far away from a cognitive core (i.e. beyond Stage VII), cf. *ladhízo* 'make/become olive-green' (*ladhí* 'olive-green'), whereas not possible/forced verbs are cognitively dissociated, cf. !*trapezízo* or !*potirízo,* whose bases *trapézi* 'table' and *potíri* 'glass', respectively, refer to objects without a characteristic colour, etc.
In the lecture I will argue that the formalization of the MG and EN colour verbs is feasible. All we need is a causative/inchoative structure with special semantic conditions on the end-state ([+Loc]) argument, see (2) and (3), respectively.

(2) **MG colour verbs**

$$\underline{[+\text{dynamic } ([_{\text{volitional-i}}\ ], [_{j}\ ])]};\ [+\text{dynamic } ([_{i}\ ], [+\text{dynamic, } +\text{IEPS } ([_{j}\ ], [+\text{Loc}_{\substack{[+\text{BCT}][+\text{PRIMARY}]\\ [+\text{BCT}][-\text{PRIMARY}]\\ [-\text{BCT}][+\text{CHAR}]}}\ ])]), <\text{base}>]$$

(3) **EN colour verbs**

$$\underline{[+\text{dynamic } ([_{\text{volitional-i}}\ ], [_{j}\ ])]};\ [+\text{dynamic } ([_{i}\ ], [+\text{dynamic, } +\text{IEPS } ([_{j}\ ], [+\text{Loc}_{[+\text{BCT}]}\ ])]), <\text{base}>]$$

In (2) and (3) 'IEPS' [Inferable Eventual Position or State] refers to a sequence of places or states (progression). The indices *i* and *j* co-index the agent and patient arguments, respectively. The base argument is co-indexed with the [+Loc] argument. In the inchoative variant the underlined part of the structure drops (Lieber 2004).

In the semantic skeleton for the MG colour verbs in (2), three disjoint conditions are defined, i.e. [+BCT] [+PRIMARY], [+BCT][–PRIMARY], and [–BCT][+CHAR]. [±PRIMARY] refers to the distinction between primary and non-primary BCTS in Kay & McDaniel (1978), i.e. colour terms up to the emergence of blue and colour terms after the emergence of blue, respectively (cf. (1)). [+CHAR] stands for 'characteristic colour' and refers to colour adjectives derived from the names of objects/substances having a characteristic colour, cf. *asimí* 'silver', etc. The options [+BCT][–PRIMARY] and [–BCT][+CHAR], will additionally drop the causative (=underlined) part of the semantic skeleton.

In the semantic skeleton for the EN colour verbs in (3), a single condition is defined, i.e. [+BCT]. This suggests that non-BCTs tend to yield no verbs in EN, cf. **to beige*, **to maroon*, etc.
According to an extensive internet research, I will finally show that systematic gaps linked to the causative/inchoative paradigm of *-ízo* derivatives are due to the semantics of the base rather than deficiencies in the base's morphophonological character. The lower the evolutional stage that a base can be fitted to, the more likely an *-ízo* derivative will be produced.

REFERENCES

Berlin, Brent & Paul Kay. 1969. *Basic color terms. Their universality and evolution.* Berkeley, CA: University of Los Angeles Press.

Kay, Paul & Chad K. McDaniel. 1978. The linguistic significance of the meanings of basic color terms. *Language* 54(3). 610-646.

Kay, Paul & Luisa Maffi. 1999. Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist* 101(4). 743-760.

Lieber, Rochelle. 2004. *Morphology and lexical semantics.* Cambridge, UK: Cambridge University Press.

Lexicalization patterns in colour naming: the case of Modern Hindī

Andrea Drocco and Orsola Risato
Ca' Foscari University, Venice
andrea.drocco@unive.it & orsolarisato@gmail.com

Since the 1960s, especially starting from Berlin & Kay's study (1969), much attention has been paid to the categorization and development of colour terms cross-linguistically. As a consequence, thanks to the works of various scholars, we possess numerous investigations on colour terms of IE and non-IE languages. However it is rather surprising that no significant research has been conducted to investigate Indo-Aryan, but especially Hindī colours terms (that is Modern Standard Hindī). Starting from those that are considered Hindī basic colour terms, as comes to light from the few studies found in literature (cf. Caskey-Sirmons & Hickerson 1977: 362, 363; see also Berlin & Kay's about basic colour terms of Urdū, the official language of Pakistan mutually intelligible with Hindī), the aim of this presentation is to draw attention to a topic that has remained so far relatively unexplored: the various "lexicalization patterns" that are used in Modern Hindī to define colours and in particular to convey different shades of meaning of basic colours. The analysis, grounded in a corpus-based research and on the most authoritative grammars (cf. Gurū 1978: 73; McGregor 1979) and dictionaries (cf. Śyāmasundardās (ed.) 1965-75; McGregor (ed.) 1993) of this language, is divided in two parts. In the first part, we will focus on the "metonymy"-type of Hindī colours lexicalization patterns, whereas in the second one the "approximation"-type mechanisms will be examined. As for the first part, we will have the opportunity to show that some Hindī colour terms are actually the result of derivation by means of the suffix *-ī*. The function of this suffix is exactly to realize the metonymic shift OBJECT COLOUR FOR COLOUR, as the suffix lead to the creation of a colour term starting from the noun of the object that is characterised by the very same colour [thus: N_{entity} + suffix]. As we will see, this lexicalization mechanism related to colour naming is not restricted, however, to fruits and vegetable (as in *baiṃganī* = "purple" from *baiṃgan* = "eggplant" + *-ī*, or *gulābī* = "pink" from *gulāb* = "rose" + *-ī*), because it is also adopted with other kind of entities (e.g. *āsmānī* = "azure, sky-blue" from *āsmān* = "the sky"). Another Hindī strategy to realize the metonymic shift OBJECT COLOUR FOR COLOUR is by way of the polysemous suffix *–vālā*. Even if this function of *–vālā* is apparently not mentioned in any Hindī grammar/dictionary, perhaps because used only in a more colloquial register of this language, it is possible to find, in our corpus, some few examples (e.g. Hindī *ret vālā* "sand (colour)"). In the second part of the analysis, data will be presented to show the two Hindī strategies to convey the "approximation"-type meaning as regards colour terms: by means of 1) the suffix *–sā* as in *lāl-sā* "reddish, red-like" from *lāl* = "red", 2) the reduplication of the colour term, meant as the juxtaposition of the same repeated adjective concerned colours as in *nilī-nilī ṭopī* (F) = "bluish cap" from *nilā* = "blue". According to some scholars (see, among others, Abbi: 1980: 95-98; Montaut 2009: 32, 34-37), reduplication pattern is a typical feature of non-Indo-Aryan languages of India and therefore Modern Hindī probably maintains this characteristic resulting from linguistic contamination. The analysis, however, will not be meant to be a diachronic one. The goal, in fact, will be to identify peculiar mechanisms that may be significant in a cross linguistic analysis of colour lexicalization and to evaluate their productivity so that to understand how Modern Hindī can be categorized as regards lexicalization patterns in colour naming.

References

Abbi, Anvita. 1980. *Semantic grammar of Hindi. A study in reduplication*. New Delhi: Bahri Publications.

Berlin, Brent & Kay Paul. 1969. *Basic color terms: their universality and evolution*. Berkeley: University of California Press.

Caskey-Sirmons Leigh A. & Hickerson Nancy P. 1977. Semantic shift and bilingualism: variation in the color terms of five languages. *Anthropological Linguistics* 19(8). 358-367.

Gurū, Kāmtāprasād. 1978. *Hindī vyākaraṇ*. Vārāṇasī: Nāgarīpracāriṇī Sabhā. (12th edn.)
McGregor, Ronald Stuart. 1979. *Hindī outline of Hindi grammar, with exercises.* Oxford: Oxford University Press. (2nd edn.)
McGregor, Ronald Stuart (ed.) 1993. *The Oxford Hindi-English Dictionary*. Oxford: Oxford University Press.
Montaut, Annie. 2009. Reduplication and echo-words in Hindi. In Singh, Rajendra (ed.), *Annual Review of South Asian Languages and Linguistics*, 21-91. Berlin: Mouton de Gruyter.
Śyāmasundardās (ed.). 1965-75. *Hindī śabd-sāgar, arthāt hindī bhāṣā kā ek bṛhat koś*. Banaras: Nāgarīpracāriṇī sabhā. (2nd edn.)

The productivity of German color compounds

Stefan Engelberg
Institut für Deutsche Sprache, Mannheim / University of Mannheim
engelberg@ids-mannheim.de

Productivity is a concept that is usually applied to derivational processes in morphology. A number of measures have been developed in order to quantify the productivity of affixes (cf., e.g., Baayen 2009). I will transfer the quantitative concepts of productivity into the domain of compounds, assuming that the productivity of simplex words as heads of compositional processes can be measured on the basis of large corpora in a way similar, yet not identical to derivational morphology. I will further show how productivity values reveal idiosyncratic differences between basic color words and between different semantic patterns that can be observed in the formation of color compounds. I will focus on the ten simplex words *blau* 'blue', *braun* 'brown', *gelb* 'yellow', *grau* 'grey', *grün* 'green', *orange* 'orange', *rot* 'red', *schwarz* 'black', *violett* 'purple', *weiß* 'white' and on three types of phenomena:

(i) In German corpora with a size of around 5.000.000.000 running words, basic color words like *rot* 'red', *blau* 'blue', and *schwarz* 'black' give rise to more than 1000 different compound lexemes each. However, applying different measures of productivity to the data reveals clear differences in productivity between basic color words. For example, independently of the particular productivity measure, *weiß* 'white' is clearly less productive in morphological composition than *schwarz* 'black'. I will use examples like these to discuss the different productivity measures. In contrast to productivity in derivational morphology, the head of a complex word in composition is a free morpheme and can occur by itself. If we take the frequency of the basic color word as a simplex into account, it can be shown that *orange* 'orange' and *violett* 'purple', which have rather low frequencies as simplex words, are comparatively productive in compound formation.

(ii) The formation of color compounds in German is dominated by three different semantic patterns: color compounds that involve a comparison (e.g., *saharagelb* 'sahara-yellow', *löwenzahngelb* 'dandelion-yellow'), color-color-compounds (e.g., *grünlichgelb* 'greenish-yellow', *rotblau* 'red-blue'), and (de)intensifying compounds (*blassgelb* 'pale-yellow', *schrillgelb* 'shrill-yellow'). Listing the compounds of a particular basic color word according to their token frequency reveals that compounds of the three semantic types are differently distributed over the frequency spectrum. That is to say, each of these morpho-semantic patterns corresponds to a particular productivity pattern.

(iii) Employing ten basic color words for the formation of binary color-color compounds yields 100 compounds. The token frequencies between these compounds differ between zero and more than five thousand. While high frequencies for *rotbraun* 'redbrown' in contrast to *violettorange* 'purple-orange' might be explained on semantic grounds or by recurrence to world knowledge, it is less easy to explain the different frequencies of a compound compared to its reversal (*blaugelb* 'blue-yellow' vs. *gelbblau* 'yellow-blue'). Despite their presumed semantic similarity they almost systematically differ strongly in frequency; e.g., *blaugelb* occurs five times as often as *gelbblau*.

I will try to explain the observed frequency distributions on semantic grounds. However, I will argue that some conspicuous frequency distributions of color compounds cannot be explained semantically but must recur to general principles of language processing, in particular to principles that force words in particular lexical domains to follow Zipfian and similar distributions.

References

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In: A. Lüdeling & M. Kytö (eds.). *Corpus Linguistics. An International Handbook*. Volume 2. Berlin, New York: Mouton de Gruyter. 899-919.

"A little greenish blue": morphological vs syntactic strategies in colour naming in French and Occitan

Benjamin Fagard and Anetta Kopecka
Lattice (CNRS, ENS & Université Paris 3 Sorbonne Nouvelle)
Dynamique Du Langage (CNRS & Université Lumière Lyon 2)
benjamin.fagard@ens.fr & anetta.kopecka@univ-lyon2.fr

In this paper, we address the use of derivational suffixes in colour naming in French and Occitan. While the use of primary (non-derived) terms is common in these languages, both have secondary colour terms derived from primary terms by way of suffixation, such as, for instance, the suffixes *-astre in* Occitan (*verdastre* "greenish") and *-âtre* in French (*verdâtre* "greenish"), which carry an approximative meaning comparable to *-ish* in *greenish*.

However, it is well-known that French, compared to other Romance languages, has a tendency to be very analytic, a trait that has led some authors to classify it as "more grammaticalized" on the whole than other Romance languages (cf. Lamiroy 2011, De Mulder & Lamiroy 2012). The question that arises then is will speakers of French and Occitan differ in morphological strategies used to express meaning in the semantic domain of colour?

We compare the colour descriptions provided by speakers of French (N=20) and speakers of Occitan (N=20) (EoSS data, Majid et al. 2015 – Colour task, Majid & Levinson 2007), classifying all terms in primary (no affixation, simple term at least synchronically) and secondary (one or more affixes) colour terms. The results show striking differences between the two Romance languages in morphosyntactic strategies used to name colours. Speakers of Occitan employ extensively secondary colour terms using 10 derivational suffixes, such as, for instance, *–às* (e.g. *marronàs*, *jaunàs*, *verdàs*), and *–enc* (e.g. *jaunenc*, *rosenc*, *verdenc*). In contrast, speakers of French tend to rely more often on syntactic strategies using adverbials such as *assez* "quite", *très* "very", *plutôt* "rather", to nuance primary colour terms. The examples below illustrate quite well these contrasting strategies. We see the suffixes *–enc* and *–astre* in Occitan, both having an approximative meaning (examples 1 and 2).

(Occitan)
(1) *jaune verd**enc***
yellow green-ish
"a greenish yellow" (EoSS_Oc_14_colournaming_66)

(2) *blau verd, blau**astre***
blue green blue-ish
"a blue green, bluish" (EoSS_Oc_07_colournaming_10)

In French, we frequently find adverbial modifiers such as *assez* "quite" (example 3) or *très* "very" (example 4, with reduplication).

(French)
(3) *bleu **assez** pâle*
blue quite light
"a rather light blue" (EoSS_Fr_01_colournaming_15)

(4) *bleu **très très** pâle*
blue very very pale
"a very very pale blue" (EoSS_Fr_20_colournaming_41)

By examining the use of linguistic strategies in French and Occitan to name various colours, with an approach combining qualitative and quantitative analyses, we provide a systematic account of the differences between the two languages with regards to the use of derivational vs syntactic processes in reference to particular colours and assess the shades of meaning conveyed by these processes.

References

De Mulder, Walter & Béatrice Lamiroy. 2012. Gradualness of grammaticalization in Romance. The position of French, Spanish and Italian. In Kristin Davidse, Tine Breban, Lieselotte Brems & Tanja Mortelmans (eds.), *Grammaticalization and language change: New reflections*, 199-226. Amsterdam & Philadelphia: John Benjamins.

Lamiroy, Béatrice. 2011. Degrés de grammaticalisation à travers les langues de même famille. *Bulletin de la Société de Linguistique de Paris* 19. 167-192.

Majid, Asifa & Stephen C. Levinson. 2007. The language of vision I: Colour. In Asifa Majid (ed.), *Field manual*, vol. 10, 22-25. Nijmegen: Max Planck Institute for Psycholinguistics. http://fieldmanuals.mpi.nl/volumes/2007/language-of-vision-colour/. [accessed October 2015]

Majid, Asifa, Fiona Jordan & Michael Dunn. 2010. *Evolution of Semantic Systems: Procedures Manual*. Nijmegen: Max Planck Institute for Psycholinguistics.

Morphological processes in the lexicalization of colour terms in Basque: labels and functions

Iraide Ibarretxe-Antuñano
Universidad de Zaragoza
iraide@unizar.es

This paper explores the lexicalization patterns and word-formation strategies that Basque speakers use to name colours in this language. Data come from 20 native speakers of Basque and were elicited using the colour elicitation task as part of the EoSS project at the Max Planck Institute for Psycholinguistics (Majid et al. 2015). After a brief introduction about the colour system in Basque (basic colours in Basque and the use of loan words (Larzabal 1975, Perurena 1992, Preciado 1992, 2009, Txillardegi 1975), this paper analyses five main morphological devices for the description of colour terms: (i) the role of compounding. Basque tends to create noun+noun compounds where the two nouns correspond to two different colours, e.g. *urdin berdea* 'blue green-abs', (ii) the role of the metonymy OBJECT COLOUR FOR COLOUR as in *azal kolorea* (flesh colour.abs) 'the colour of flesh', (iii) the role of adjectival modifiers such as *argi* 'light' and *ilun* 'dark' to describe different colour hues, e.g. *urdin argia* 'light blue', (iv) the use of reduplication as in *urdin argi-argia* 'light-light blue', and (v) the role of diminutive suffixes to describe an "approximated meaning", e.g. the suffix *-ska/-xka* '-ish' as in *horixka* 'yellowish'.

Basque speakers extensively use these morphological devices in the colour elicitation task (full response only), but it is argued that these strategies respond to different goals. Strategies (i) and (ii) are used for "colour description", i.e., those cases when the speaker clearly identifies the colour chip and finds the appropriate label to name it. Here, the speaker uses a single morph (*gorri* 'red'), loan words (*arroxa* 'pink'), compounds (*berde urdin* 'green blue'), and metonymy-based labels (*haragi kolorea* 'the colour of flesh' to name the colour. Strategy (iii) is used for "hue modulation", i.e., to modulate the hue colour, once the colour chip has been clearly identified (*hori argia* 'light yellow'). Strategies (iv) and (v) are used for "identification with colour prototype". That is, the speaker either identifies the colour chip as the best example of the focal colour (e.g., *gorri-gorria* 'red-red'), or considers the colour chip just as an approximation, barely close to the focal colour (*gorrixka* 'redish'). Strategy (iv) can also be used to describe different degrees of hue colour (*hori argi-argia* 'very light yellow') which are equivalent to the use of quantifiers such as *oso* 'very' (*horia oso argia* 'yellow, very light').

References

Larzabal, Piarres. 1975. Euskaldunak eta koloreak. *Euskera* XX. 401-405.

Perurena, Patziku. 1992. *Koloreak euskal usarioan*. Donostia: Erein.

Majid, Asifa, Fiona Jordan & Michael Dunn. 2015. Semantic systems in closely related languages. *Language Sciences* 49(1). 1-18.

Preciado, José Mª. 1992. *Kolorea herri euskaldun batean*. Leioa, E.H.U.: Argitarapen Zerbitzua.

Preciado, Txema. 2009. *Euskarak erakutsi koloreak ikusten*. Gasteiz: Eusko Jaurlaritza.

Txillardegi. 1975. La structuraction du champ semantique de la couleur en Basque, *Euskera* XX. 221-238.

Welsh colour terms - from compounding to loan translation

Katarzyna Jędrzejewska-Pyszczak
Adam Mickiewicz University, Poznań, Poland
jkat@wa.amu.edu.pl

This paper investigates different formation patterns within the Welsh colour spectrum and their productivity. Apart from the traditional set of single-word entities for basic colour terms such as *coch* 'red', *gwyn* 'white' or *du* 'black' (cf. Lazar-Meyn 1991), diachronically one comes across a variety of compound terms for colours in Welsh (as in *cochddu* 'brownish' , *llwydfelyn* 'yellowish–grey' or *gwyngochliw* 'of reddish-white colour') (cf. Zimmer 2000, Stanulewicz 2003) and colour terms with adjectival endings suffixed to them (*glas* 'blue' vs *glas**ol***). Thus, on the one hand, the present study focuses on the structure of compounded expressions, mutation rules at play and the possible reordering of elements within given compounds. An attempt is made at addressing the question whether the order of elements within a compound affects its meaning and usage (as in *gwynlas* vs *glaswyn* where *glas* 'blue' and *gwyn* 'white'). On the other hand, however, the paper hopes to test the hypothesis that in the Welsh language it is the distinction between the literal and figurative meaning that can shape the structure of a compounded colour term and,- in the case of uncompounded colour terms - explain the choice of suffixed adjectival endings. Moreover, special attention is paid to the linguistic make-up of relatively recent colour terms in Welsh (such as *pinc* 'pink', *oren* 'orange', *wmber crai* 'raw umber') (Williams 1973, Evans 1981).

REFERENCES

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The syntax of Welsh*. Cambridge: Cambridge University Press.

Evans, H. Meurig. 1981. *Y geiriadur Cymraeg cyfoes: The dictionary of modern Welsh*. Llandybie: Hughes a'i Fab.

Evans, H. Meurig (et al.). 2003. *Y geiriadur mawr*: *The complete Welsh-English English-Welsh dictionary*. Llandybie: Gwasg Dinefwr.

Fynes-Clinton, O.H. 1913. *The Welsh vocabulary of the Bangor district*. London: Oxford University Press.

Geiriadur Prifysgol Cymru [The University of Wales dictionary]. 2014. http://www.welsh-dictionary.ac.uk/. (August-October 2015)

Lazar-Meyn, Heidi Ann. 1991. The colour systems of modern Celtic languages: Effects of language contact. In Per Sture Ureland and George Broderick (eds.), *Language contact in the British Isles: Proceedings of the eighth international symposium on language contact in Europe, Douglas, Isle of Man, 1988*, 227-242. Berlin: De Gruyter.

Lazar-Meyn, Heidi Ann. 2004. Colour naming: 'Grue' in the Celtic Languages of the British Isles. *Psychological Science* 15/4. 288.

Stanulewicz, Danuta. 2003. Słownictwo barw w języku walijskim [Colour vocabulary in the Welsh language]'. *Linguistics across culture*.

Thomas, Peter Wynn. 1996. *Gramadeg y Gymraeg* [The grammar of the Welsh language]. Caerdydd: Gwasg Prifysgol Cymru.

Williams, Jac L.(ed.). 1973. *Geiriadur termau* [Dictionary of terms]. Cardiff: University of Wales Press.

Winward, Fiona. 2002. *Colour terms in early Welsh literature*. Cambridge: University of Cambridge dissertation.

Zimmer, Stefan. 2000. *Studies in Welsh word formation*. Dublin: Dublin Institute for Advanced Studies.

The Syntax of Light and the Morphology of Luminosity: Danish Visual Semantics Revisited

Carsten Levisen
Department. of Culture and Communication,
Roskilde University, Denmark
calev@ruc.dk

Danish visual semantics is, like other European languages, hue-based in its orientation, but with as focus on "brightness" embodied in its grammar, lexicon and discourse. There are three modificational grammatical devices in Danish based on *lys* "light": *lys-*, *lyse-*, and *lys*, for instance: *lyselilla*, *lys lilla*, *lyslilla*, all of which can roughly translated as "bright purple" in English. What do these modifiers mean? And why are there three such devices in Danish?

The discussion of *lyse- lys-* and *lys* in linguistic semantics, is framed within a general framework of "luminosity", a keyword in contemporary Danish cultural anthropology. Drawing on insights from linguistic anthropology (Bille), and visual semantic studies (Wierzbicka 2006, 2008, 2014, Goddard & Wierzbicka 2013; Aragón, 2013, in press; Levisen, Sippola & Aragón, in press) and historical semantic studies of brightness (Biggam 2012, Jones 2013), the paper seeks to revitalize the non-hue based visual agenda in colour studies. With an in-depth case study of Danish, and with contrastive reference to Swedish, German and English, the study explores how *rød* 'red', *blå* 'blue' and *grøn* 'green' appear to provide special cases within the Danish visual system of "brightness".

The paper takes a two-pronged approach, using EoSS data (the Danish database), and an interpretative analysis guided by the natural semantic metalanguage (NSM) approach to visual semantic analysis. Based on the EoSS results, the paper argues that terms such as "bright","light", "luminous" need to be deconstructed, given their intranslatability and instability across even closely related languages. Also, the supposed dualistic relationship between "brightness" and "darkness" in language is also critically examined based on the data. The study confirms Sapir's old dictum that "all grammars leak", showing the details of how the Danish morphosyntax of "brightness" both operates as a systems, and defies the system.

The paper concludes that Danish is indeed a "brightness"-focused language, and that the different morphosyntactic options *lys*, *lys-*, and *lyse-* provide Danish speakers with an intricate semantic-conceptual grid of meanings, all of which are configured around the concept of *lys* 'light, brightness'. Further, it is concluded that modern Danish *lyserød*, despite its surface form, is not (or no longer) a brightness-based concept. By contrast, *lyseblå* and *lysegrøn*, have developed into "hyper-brightness" terms, with sunlit natural scenarios as semantic prototypes.

Color naming in a three-dimensional linguistic system: The case of Polish Sign Language (PJM)

Sylwia Łozińska and Paweł Rutkowski
Section for Sign Linguistics, University of Warsaw
s.lozinska@uw.edu.pl & p.rutkowski@uw.edu.pl

Recent typological studies on sign languages of the Deaf have made important contributions to our understanding of color naming strategies in human communication. As discussed in de Vos & Pfau (2014), sign languages of small village communities (whose emergence is usually related to a high incidence of deafness) differ from those used in urban areas by having a limited lexical set in the domain of color. Adamorobe Sign Language (Nyst 2007), Ban Khor Sign Language (Nonaka 2004), Providence Island Sign Language (Woodward 1989) have only three color names (BLACK, WHITE and RED), whereas the color vocabulary of Inuit Sign Language (Schuit 2014) is limited to two signs: RED and BLACK. Other colors are referred to by pointing at an appropriate object. The situation looks very different in most urban sign languages, whose color vocabularies generally reflect those of spoken languages that surround them.

The aim of this paper is to discuss and exemplify color naming strategies observed in Polish Sign Language (*polski język migowy*, PJM) and to present detailed data on the frequency of use of color-related vocabulary in the PJM corpus. PJM is a natural visual-spatial language used by the Deaf community in Poland. It emerged around 1817, with the foundation of the first school for the deaf in Poland. The present study derives from a large-scale research project aimed at documenting PJM through the compilation of an extensive and representative corpus of annotated video material showing Deaf people using their language in a variety of different contexts. Recording sessions always involve two signers and a Deaf moderator. The procedure of data collection is based on a list of tasks to be performed by the two informants. Typically, the signers are asked to react to certain visual stimuli, e.g. by describing a scene, naming an object, (re-)telling a story, or explaining something to their partner. The elicitation materials include pictures, videos, graphs, comic strips etc. The raw material obtained in the recording sessions is further tokenized, lemmatized, annotated, glossed and translated using the iLex software.

For the purposes of the present paper, we have analyzed signed utterances produced by 44 corpus informants. We have scanned this sub-corpus for color-related terminology, calculating the number of occurrences of each lexical item. This inspection of corpus data showed that the eleven basic color categories listed by Berlin & Kay (1969) (white, black, red, green, yellow, blue, brown, purple, pink, orange, grey) have all been lexicalized in PJM. In the sample we have analyzed, they were used 972 times. Other color-related terms found in the data were limited to LILAC, SILVER and GOLDEN (24 occurrences in total). Color signs were often accompanied by the signs DARK and LIGHT (forming phrases such as RED DARK, RED LIGHT). The most frequent color terms are: WHITE (195 tokens), RED (164 tokens) and BLACK (131 tokens). These findings support the universal patterns described by Berlin & Kay (1969). Our paper discusses the corpus frequency data in detail and provides examples of how PJM color terms are used.

Having compiled the list of all color terms attested in the corpus of present-day PJM, we compared it to the earliest known linguistic description of the lexicon of PJM (Hollak & Jagodziński 1879). Although Hollak & Jagodziński's remarks concerning articulatory nuances are often imprecise, our examination shows that only four PJM color signs (BLACK, BLUE, YELLOW, GOLDEN) have not changed since 1879. Three more signs (WHITE, RED, GREEN) have been modified with respect to one articulatory feature but are still clearly related to their 19th-century predecessors. The remaining color signs are later additions.

REFERENCES:

Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
Hollak, Józef & Teofil Jagodziński. 1879. *Słownik mimiczny dla głuchoniemych i osób z nimi styczność mających*. Warszawa: IGO.
Nonaka, Angela M. 2004. The forgotten endangered languages: Lessons on the importance of remembering from Thailand's Ban Khor Sign Language. *Language in Society* 33. 737-767.
Nyst, Victoria. 2007. *A descriptive analysis of Adamorobe Sign Language (Ghana)*. Utrecht: LOT.
Woodward, James. 1989. Basic color term lexicalization across sign languages. *Sign Language Studies* 63. 145-152.
Schuit, Joke. 2014. *Signs of the Arctic: Typological aspects of Inuit Sign Language*. Amsterdam: University of Amsterdam dissertation.
Vos, Connie de & Roland Pfau. 2014. Sign language typology: The contribution of rural sign languages. *Annual Review of Linguistics* 1. 265-288.

Word-formation and meaning of colour terms in Croatian, Czech and Polish

Ida Raffaelli, Jan Chromý, Anetta Kopecka
Department of Linguistics, Faculty of Humanities and Social Sciences, University of Zagreb
Institute of the Czech Language and Theory of Communication, Faculty of Arts, Charles University in Prague
Dynamique Du Langage (CNRS & Université Lumière Lyon 2)
ida.raffaelli@ffzg.hr & jan.chromy@gmail.com & anetta.kopecka@univ-lyon2.fr

How closely related languages construe meaning in the semantic domain of colour? Slavic languages have a relatively rich inventory of derivational affixes that allow to derive adjectives (as well as nouns) from nominals. Besides, they also use compounding as a word-formation process to construe complex adjectives, including adjectives for colours. Do speakers of Slavic languages rely on these word-formation processes to the same extent or do they vary in the way they use them to describe similar stimuli of colours? To investigate this question we explore data from Croatian, Czech and Polish that we collected using the visual material designed at the Max Planck Institute for a cross-linguistic project Evolution of Semantic Systems (Majid & Levinson (2007); Majid, Jordan & Dunn (2015)). In particular, we examine which types of meanings of colour terms are conveyed via derivation (suffixation) or compounding, as the two most pervasive word-formation processes, in the three Slavic languages.

Derivation is utilized in the three languages but there are cross-linguistic differences in the use of derivational suffixes in the colour domain. In Croatian, for example, the suffix *-ast* means 'N-like', such as in *naranč-ast* 'orange-like', whereas the suffix *-kast* has an approximative meaning, such as in *zelen-kast* 'greenish'. In Croatian, the suffix *-ast* is regularly used to form colour terms from nouns referring to objects, i.e. the 'N-like' meaning. The same pattern of semantic differentiation between two types of suffixes can be identified in Polish as well. Polish has the suffix *-ow* to realize colour terms via nouns referring to objects, i.e. the 'N-like meaning, and its use is even more productive than the use of the Croatian suffix. In Polish, it can be added to nouns referring to different types of entities including fruits, vegetables and animate species (e.g. *łososiowy* 'salmon-like'), whereas in Croatian the use of the suffix *-ast* is restricted to fruits and vegetables. Moreover, Polish has the suffix *-aw*/*-aw* (e.g. *zielonk-aw-y* 'greenish'), which has an approximative meaning. Czech has the suffix *-ov* that corresponds to both the Croatian *-ast* and the Polish *-ow* suffixes, realizing the 'N-like' meaning. As in Polish, Czech *-ov* suffix can be added to nouns which refer to various things with a typical colour (e.g. *petrolej-ov-ý* 'kerosene-like'). However, Czech has no corresponding suffix for the approximative meaning and therefore has to resort to other grammatical devices like prepositional phrases to convey similar meaning (e.g. *do modra* 'bluish').

For the formation of compounds in colour naming different types of modifiers are used. Modifiers differ with respect to part-of-speech and semantic content. They are mostly adverbs, such as Croatian *tamno* 'dark' and *svijetlo* 'light' (*světle* in Czech) or can be adjectives that form compounds, such as Croatian *zeleno-plava* 'blue-green' Czech *oranžovohnědá* 'orange-brown' or Polish *morsko-niebieski* 'marine-blue'. Adverbs that are used as modifiers mostly refer to hue, and sometimes to brightness as well (e.g. Cro. *zagasito* 'dull' and *jarko, žarko* 'bright'). Adjectives used as modifiers form compound colour terms that could be characterized as headless compounds (e.g. Cro. *zeleno-plava* 'blue-green') and that may switch position. According to the Croatian EOSS data, *zeleno-plava* 'blue-green' and *plavo-zelena* 'green-blue'are used to name the same colour shade.

Preliminary analyses show that derivation and compounding are not used in the given three languages with similar regularity and productivity. While the two word-formation processes are available in the three languages, Polish relies more often on derivational suffixes to form meaning in this domain of expression whereas Croatian and Czech rely more often on compounding strategies (e.g. *zelenohnědá* 'green-brown' is preferred to *olivová* 'olive-like' as a name for the same colour in Czech).

REFERENCES :

Majid, Asifa, Fiona Jordan & Michael Dunn. 2015. Semantic systems in closely related languages. Language Sciences 49(1). 1–18.

Majid, Asifa & Stephen C. Levinson. 2007. The language of vision I: Colour. In Asifa Majid (ed.), Field manual, vol. 10. Max Planck Institute for Psycholinguistics, Nijmegen, 22–25. http://fieldmanuals.mpi.nl/volumes/2007/language-of-vision-colour/.

Lexicalization Patterns in Color Naming in Korean

Seongha Rhee
Hankuk University of Foreign Studies
srhee@hufs.ac.kr)

Korean has arguably one of the richest inventories of color terms across languages. Lexicographers list a large number of terms expressing colors across nominal, adjectival and adverbial categories. For instance, Park (1989) lists 362 adjectives that describe colors and hues, and Nam (1992) lists 41 nouns that denote color names.
In the nominal category of 41 terms, in addition to seven native terms for five focal colors (i.e. black, white, yellow, red, and grue (green/blue)), the two dominant naming strategies are (i) compounding involving an object name and a noun *pich* 'light' (e.g. *kum-pich* 'gold-color') and (ii) borrowing from foreign languages (e.g. *tahong* 'crimson red' from Chinese and *pola* 'purple' from Mongolian).
A far more intriguing state of affairs is observed with respect to adjectival and adverbial categories. In lexicalizing colors and hues in these categories, diverse strategies are recruited. Sound symbolism is the most determinative factor, one of the widespread constraints of vowel harmony in word formation in general, often invoking the notion of 'positive' vowels (a, o, æ, and diphthongs involving them) and 'negative' vowels (ə, e, u, and diphthongs involving them). As a general principle, positive vowels are associated with describing small, delicate, and bright objects, whereas negative vowels, with big, crude, and dark objects. Sound symbolism also makes the tense vs. non-tense distinction among consonants, the former representing stronger force (Koo 2007). Sound symbolisms operative in color naming may be summarized in brief as follows:

(a) positive vowels for bright hue vs. negative vowels for darker hue
e.g. norat- 'bright yellow' vs. nurət- 'dark yellow'
(b) tense consonants for stronger hue vs. non-tense consonant for weaker hue
e.g. ppalgat- 'red' vs. balgat 'redish'
e.g. gəm- 'black' vs. kkəm- 'dark black'
(c) prefixation of intensifiers sæ-, sæt-, si- and sæt- for darker, stronger hue
e.g. parat- 'grue (blue/green)' vs. sætparat- 'strongly grue'
(d) phonetic extension to invoke the sense of extended surface or subdued hue by way of suffixing -ʉkkeha-, ʉreha-, ʉsrʉmha-, -ʉkkʉrʉmha-, -ʉmureha-, -ʉkkumureha-, etc.
e.g. norat- 'bright yellow' vs. norʉkkeha- 'slightly yellowish'
(e) reduplicative suffixation to invoke the sense of extended surface, uneven distribution, spotty hue, and evaluation of pleasurability by way of suffixing -ʉsuksukha-, -ʉjapjapha-/-ʉjəpjəpha-, -ʉjokjokha-/-ʉjukjukha-/-ʉjəkjəkha-, -ʉchikchikha-/-ʉchukchukha-/-ʉchʉkchʉkha-, -ʉchungchungha-, -ʉdedeha-/-ʉdædæha-, -ʉthwethweha-/-ʉthwithwiha-, -ʉdengdengha-/ʉdingdingha-/-ʉdungdungha-/ʉthungthungha-, etc.
It is noteworthy that some of the suffixes are sound-onomatopoeic, suggesting that extended color distribution is similarly conceived of as sound ringing out in the air.
e.g. gəm- 'black' vs. gəmujukjukha- 'unevenly, unpleasingly black'
(f) discontinuous reduplicative suffixation for similar effect in (e) by suffixing -ʉt-ʉtha-, -sung-sungha-, -jək-jəkha-, -sil-silha-, etc.
e.g. kkam- 'strongly black' vs. kkamutkkamutha- 'spotty black here and there'

An investigation into the color naming patterns in Korean reveals an iconic relationship between the perceived state of affairs in the world and the linguistic forms in intricate application of elaborate sound symbolism, which also suggests visual-auditory synesthesia in color perception. Since the sound symbolism has such a robust representation in the minds of the native speakers, when new color terms are encountered, they can correctly conceive of the colors and hues denoted by the novel expressions without trouble. Color term formation is a productive derivational process which borders on, and sometimes blends into, other related perceptual domains such as auditory perception, distance perception, opacity perception, value judgment, etc. Equipped with such a rich derivation system, Korean has a paradigm of extraordinarily fine-grained color terms that defy faithful translation into other languages. The lexicalization pattern of color terms in Korean is a cross-linguistically unique idiosyncratic feature in terms of its productivity and far-reaching effects across morphological categories.

References

Koo, Hyun Jung. 2007. Force dynamics as a variational factor: A case in Korean. *LACUS Forum* 33. 201-210.
Nam, Youngsin. 1992. *Wulimal pwunlyu sacen* [Korean lexicon by semantic categories]. 4 vols. Seoul: Hangang Publishing.
Park, Yongsu. 1989. *Wulimal kalay sacen* [Korean lexicon by semantic categories]. Seoul: Hankil Publishing.

Peculiarities of the Ossetic colour system

Arseniy Vydrin and Oleg Belayev
Institute for Linguistic Studies, Russian Academy of Sciences, St. Petersburg, Russia
Moscow State University, Moscow, Russia
senjacom@gmail.com & obelyaev@gmail.com

In the paper[1], we will examine the peculiarities of the colour terms in Ossetic, the affixes used with the colour terms and the main patterns to describe colour in modern standard Ossetic. The study is based on the data from Ossetic National Corpus (ONC, http://corpus.ossetic-studies.org/), Ossetic Oral Corpus (http://ossetic-studies.org/en/texts) and our field data collected in North Ossetic, Russia.

From the morphological point of view Ossetic distinguishes basic colours, derivational colours and compounds. A basic colour term consists of one word without affixes which expresses only colour. There are 6 basic colours in Ossetic: *šaw* 'black', *urš* 'white', *šərx* 'red', *bur* 'yellow', *c'ɜx* 'green, blue, grey', *morɜ* 'brown'.

There are different models to form derivational colour terms. The general one is the suffix *-xwəž* 'colour' used with an object. The most frequent colour terms formed after this pattern are *kɜrdɜgxwəž* (lit. grass colour) 'green, blue' (close to *c'ɜx*) and *ɜrvxwəž* (lit. sky colour) 'blue'. The suffix *-xwəž* grammaticalized from the word *xwəž* 'colour', 'appearance, look'. It can be used with objects of different semantic types: body (*b^{w}ar-xwəž* (body-colour) close to pink and violet, EOSS's chips 5, 33, 40, 51, 77); articles of domestic utility: *černilɜ-xwəž* (ink-colour) close to violet and blue, EOSS,s chips 3, 77, 79, 80, 81; names of flowers: *wardi-xwəž* (rose-colour) etc. More abstract is the use of the object 'earth' in this pattern: *žɜxx-xwəž* (earth-colour) close to brown and green-blue (EOSS's chips 36, 44, 55). The suffix can be employed with "colour attributes", e.g. *ird* 'bright', 'shining' — *irdxwəž* close to pink (EOSS's chip 38). *-xwəž* can also be attached to abstract nouns to form attributes; the meaning of the suffix in this case is 'appearance, look'; the derivate sometimes undergoes a metonymic shift, e.g. *rənčən-xwəž* (sick-colour) '(someone) with an unhealthy appearance', *xɜrž-xwəž* (good-colour) 'fat, well-fed'.

The general grammatical difference between the basic colours and the derivates in *-xwəž* is the inability of the derivational colours to employ most of the morphological markers used with the basic colours.

Colour compounds are basically derived by the combination of a basic colour as a second component and the first component which can be an object (*ɜrv-c'ɜx* (sky-green.blue) 'blue', EOSS's chips 2 and 10), a colour modifier *ird-c'ɜx* (bright-green.blue) 'light blue', chips 18, 78, 81) or another basic colour (*bur-bən-sərx* (yellow-SUF-red) 'orange'). There are also some other patters to form compound colour terms in Ossetic.

Ossetic possesses at least five different affixes which are mainly used to mark different degrees of saturation and brightness. There are two markers for intensive degree and three diminutive suffixes.

Intensive affixes:

a) The prefix *šɜnt-* is used with colour terms to convey high saturation of the colour or 'absolute' colour without any admixture, e.g. *šɜnt-sərx* 'very red', *šɜnt-urš* 'absolutely white'. According to our corpora data, it can be used only with the basic colours except *morɜ* 'brown'. *šɜnt-* is also employed with other adjectives, cf.: *šɜnt-xuš k'aliw* 'absolutely dry limb'.

b) The suffix *-id* attached to the second component of the reduplicated basic colour term expresses the same meaning as the prefix *šɜnt-*, e.g. *šərx-šərx-id* 'very red', *šaw-šaw-id* 'very/absolutely black'. It can not be used with any kind of compound colours. The only basic colour which is not used in this pattern in *morɜ* 'brown'.

Diminutive suffixes:

[1] This research is carried out with the financial support of the fellowship of the President of Russian Federation (MK-1920.2014.6)

a) The suffix *bən-* which can be used only with basic colours: *urš-bən* ‘whitish’, ‘grey’. It can also be employed with colour modifiers: *tar-bən* (dark-SUF) ‘darkish’. The suffix can be used with other qualitative attributes (conveying ‘the weaker quality’) and nouns (forming an attribute): *wəmɜl* ‘wet’, ‘damp’ — *wəmɜl-bən* ‘dampish’; *xwər-bən* (gravel-SUF) ‘stony’.
b) The suffix *-gomaw* is used with the basic colours as well as with the compounds with *-xwəž*. e.g. *šaw-gomaw* ‘blackish’, *ɜrvxwəž-gomaw* ‘bluish’. The suffix can be employed with other qualitative attributes: *lɜmɜɣ-gomaw* (weak-DIM) ‘weakish’.
c) The suffix *-gond* is used with basic colour, e.g. *šərx-gond* ‘reddish’ (lit. ‘close to red’). The suffix *-gond* is derived from the past stem of the verb *kɜnən* ‘to do’. It can be used with nouns to mark the nouns of action (*xwəm-gɜnd* field-do.PST ‘ploughed field’) or the similarity with the object (*xəž-gond* net-do.PST ‘lace’ (something similar to a net).

In the paper, we will examine the Ossetic colour system in the framework of the typology of colour terms, highlighting its peculiar features.

Phaedra Royle and Karsten Steinhauer

Morphology and the brain: Online studies of morphological processing.

Neuropsycholinguistic models of morphology range from 'full-decomposition' (e.g., Marantz, 2013) to 'no morphology' (e.g., Rumelhart & McClelland, 1987), with many intermediate models incorporating lexicalized and rule-based processing (see Royle et al, 2012). In some dual models, morphological (de)composition is only an optional process in word production and comprehension. While neuroscientific work often lacks in fine-grained description of linguistic structure, linguistic theory is often exempt of information on the time-course of language processing. Our workshop will directly address models of mental representation of morphology and bring together the domains of morphology and neuroscience. We aim to develop an informed discussion that integrates knowledge from these two research domains in a scientifically rich and fruitful way. By doing this we expect to push the boundaries of scientific research on morphology while illustrating how the neurosciences and linguistics can speak to each other.

This workshop brings together a highly international group of established scientists and emerging scholars from various backgrounds and countries. We will focus on online processing of morphology at the word and sentence level using event related brain potentials (ERP), eye-tracking, magnetoencephalography (MEG), and near-infrared spectography (NIRS), combined with more traditional behavioural measures. Morphological processes involved in word recognition and production as well as those involved in sentence processing will be our main topics of interest. Themes discussed include morphological parsing of affixes and compounding in real time, effects of multilingualism, verb class, and morphological coherence in gender.

José Alemán Bañón, Jason Rothman & David Miller

The role of markedness in the native and nonnative processing of agreement: An ERP study

We use ERP to examine the unique contribution of morphological markedness to the processing of number and gender agreement in Spanish by native speakers and adult English-speaking learners.

Sendy Caffarra

Influence of gender-to-ending consistency on monolingual and bilingual agreement processing.

We study whether and when gender-to-ending correspondences impact agreement computation during sentence comprehension and whether effects vary by language profile.

Davide Crepaldi, Federica Degno, Aureliu Lavric & Kathy Rastle.

Is there morphology in the parafovea?

We present data on sentence reading, in Italian and English, investigating morphological structure processing in parafoveal words, minimally at a morpho-orthographic level.

Laura Gwilliams & Alec Marantz.

Taking Morphology Seriously: MEG Studies of Morphological Representations.

MEG experiments on Arabic and English show that linguistic representations are necessarily and crucially exploited in language use, and that speakers decompose apparently simple words into roots and affixes as a necessary step in word recognition.

Gunnar Jacob, Rithwik Mutyala, Isabell Wartenburger & Harald Clahsen.

A tale of blood and light: Using Near-Infrared Spectroscopy to investigate the production of inflected forms in native and non-native speakers.

This NIRS study investigates the production of regular vs. irregular English past-tense forms in native speakers and highly proficient second-language learners of English.

Anna Jessen, Elizabeth Fleischhauer & Harald Clahsen.

ERP-data from a silent-production plus delayed-vocalisation paradigm of regular vs irregular German past participles in adults and children.

Data on adults and children suggest differential processing in production of regular and irregular past participles, supporting dual mechanism models and developmental changes between the ages of 8 and 13 years.

Gary Libben, Charles Davis & Sid Segalowitz.

Compound words, semantic transparency, and metaphors for morphological representations in the brain.

Experiments on the processing of semantically transparent and opaque real and novel compounds indicate very early semantic effects in brain reactions showing that compound constituents may have distinct representations in the brain.

Joanna Morris.

Evidence for a dual-route model of complex word processing.

We present ERP evidence for a dual-route model of morphological processing that includes both orthographic and semantic representations in the early stages of derived word recognition

Phaedra Royle, Kate Coughlen & Karsten Steinhauer.

On the existence of morphology through the lifespan: ERP and lexical decision experiments.

We present evidence for the independence of morphological, semantic and orthographic representations in neuroimaging studies of French natives and second language learners.

Linnaea Stockall & Christina Manouilidou.

MEG and EEG investigations of morpho-syntactic and morpho-semantic wellformedness in prefixed words.

EEG and MEG show distinct spatio-temporal dynamics of two grammatical information types : novel derived words that violate lexical category restrictions on affixation (*reflat, *kareklatís/ 'chairer') and those which violate argument-structure restrictions (*relaugh, orimastís/ 'maturer') in English and Greek.

References

Marantz, Alec. 2013. No escape from morphemes in morphological processing. *Language and Cognitive Processes* 28(7). 905-916.

Royle, Phaedra, John E. Drury, Nicolas Bourguignon & Karsten Steinhauer. 2012. The temporal dynamics of inflected word recognition: A masked ERP priming study of French verbs. *Neuropsychologia,* 50(14). 3542-3553.

Rumelhart, David E. & James L. McClelland. 1986. On learning the past tenses of English verbs. In *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological and biological models* 2, 216 – 71. Cambridge, MA: MIT Press.